

Title: Bowling together: Scientific collaboration networks of demographers at European Population Conferences

Authors:

1. Guy J. Abel*

Asian Demographic Research Institute (ADRI), Shanghai University, Shanghai, China
Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria

2. Raya Muttarak*

School of International Development, University of East Anglia, UK and Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria.

3. Valeria Bordone

Institut für Soziologie, Ludwig-Maximilians-Universität München, Munich, Germany

4. Emilio Zagheni

Department of Sociology, University of Washington, Seattle, USA

*Notes: Guy J. Abel and Raya Muttarak are co-first authors

Corresponding author:

Raya Muttarak

International Institute for Applied Systems Analysis (IIASA)

Schlossplatz 1, A-2361 Laxenburg, Austria

Phone : +43 2236 807 329

Fax: +43 2236 71 313

Email: muttarak@iiasa.ac.at, raya.muttarak@oeaw.ac.at

Bowling together: Scientific collaboration networks of demographers at European Population Conferences

Abstract

Studies of collaborative networks of demographers are relatively scarce. Similar studies in other social sciences provide insight into scholarly trends of both the fields and characteristics of their successful scientists. Exploiting a unique database of metadata for papers presented at six European Population Conferences (EPC) this report explores factors explaining research collaboration among demographers. We find that 1) collaboration among demographers has increased over the past ten years, however, among co-authored papers, collaboration across institutions remains relatively unchanged over the period, 2) papers based on core demographic subfields such as fertility, mortality, migration and data and methods are more likely to involve multiple authors and 3) multiple author teams that are all female are less likely to co-author with colleagues in different institutions. Potential explanations for these results are discussed alongside comparisons with similar studies of collaboration networks in other related social sciences.

Keywords: demography; population studies; scientific collaboration; collaboration networks.

Introduction

Studies on scientific collaboration networks have gained scholarly interest since the past couple of decades. There are only a few studies that specifically investigate collaboration networks of demographers. These are predominantly focused on data from journals and academic meetings based in the United States and without much attention on the contribution of potential explanatory factors for co-authorship patterns. This report aims to fill this void, focusing on collaboration networks of papers presented at past European Population Conferences. It provides new insights into factors related to whether a paper is likely to involve multiple authors and whether co-authors are likely to come from different institutes. We compare our findings to similar research on demographers from other regions as well as results in related social sciences.

The number of co-authors of an academic article has been found to provide an indicator of social capital. It can play a crucial role in individuals' job mobility and academic success (Bäker 2015). Through globalization, changing communication patterns and increasing mobility of scientists, a rise in collaborative research, especially international collaboration has been witnessed over the two past decades (J. 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 law (J. Adams 2012) as well as social science disciplines such as economics (Laband and Tollison 2000; Sutter and Kocher 2004), political science (W. C. Adams et al. 2014; Fisher et al. 1998), and sociology (Hunter and Leahey 2008). Of these, Hunter and Leahey (2008) take the most in-depth look at collaboration patterns. They found no difference between men and women's rates of collaboration and a curvilinear relationship

between year of publication and cross-institution collaboration of sociologists. Further, co-authorship was found to be more common in quantitative research than in theoretical and qualitative research.

Studies looking at demographers and their collaborative networks are limited. Merchant (2015) provided extensive historical analysis of the development of demography between 1925 and 2010. She undertook a text analysis of the abstracts from research articles published in *Demography*, *Population and Development Review*, and *Population Studies* over the period and performed a network analysis of the linkages between journals and authors over time. Other studies on the characteristics of demographic research are based exclusively on articles published in the journal *Demography*. Teachman et al. (1993) analysed changes in demographic subfields using articles from the journal published between 1964 and 1991. 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.

Existing studies on scientific collaboration networks commonly use bibliometric databases of scientific publications (i.e., journal articles, books and book chapters) to identify authors, their publications, affiliations, and co-authors. These methods are useful in analysing collaboration practices such as co-authorship and citation networks, but can include a sample selection bias. Junior academics and doctoral students typically have a lower number of journal publications compared to senior academics. Subgroups of academics 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 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*. However, it is usually at the European Population Conference that European scholars working in the demographic field meet and present their ongoing work.

Exploiting a database of papers presented at the European Population Conferences (EPC), this paper aims to investigate the driving factors behind co-authorship both within and across institutions. The EPC is the largest demographic conference in Europe, with average attendance of approximately 1,000 participants. Organised biannually by the European Association for Population Studies (EAPS) the conferences covers a wide range of topics in population studies and related disciplines. Utilising data on papers presented at conferences allows us to reduce the potential sample selection problem of the bibliographic database and provides a more Eurocentric view of demographers in comparison to the previous studies discussed above.

The remainder of the paper is organised as follows. The data section describes the data used for the analysis and presents descriptive summary of location of authors and collaboration networks. This is followed by two logistic regression analyses on the determinants of collaborations between

demographers. First, we look at factors related to the probability that a paper has multiple authors. Second, we study differences in the probability of multiple-authored papers involving multiple institutions. In the final sections we discuss the findings and limitations before we conclude.

Data

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 with assistance from the EAPS. The data are maintained by Pampa 4.1, hosted by Princeton University. For each accepted paper (either for oral or poster presentations) information was provided on the author's name and institutional affiliation, 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.

A number of data cleaning steps were undertaken. We identified and harmonised inconsistencies related to a) 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 b) 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). In cases where the authors have more than one affiliation on a given paper, we selected the first affiliation that the authors indicated in the electronic submission system. As the data were in panel format, institutions of a given author (based on their harmonised name) might have changed over time allowing us to build variables on the number of past institutions for the analysis in the next section of this paper. In 49 cases (out of 9,183) we were unable to identify an author's institution.

For our analysis on collaborative networks we wanted to derive further information on the location of authors' institutions, the authors' gender, and a broad categorisation of the paper to a demographic sub-discipline. To identify author locations we used the Google Maps API via the *ggmap* package in R (Kahle and Wickham 2013) to provide the geo-coordinates of the institution location. In the cases where the Google Maps API was unable to locate the institution, we manually looked up the location using an Internet search. Author's gender was predicted using comparisons with first names provided in the EPC data with gender-name frequencies in the *genderize.io* database (based on a large database of known gender-name combinations from social media profiles). This was undertaken using the *gender* function in the *gender* R package (Mullen et al. 2015). For some non-English names, no probability was available due to a lack of observations in the *genderize.io* database. In these rare cases, we assigned gender using other data sources (again via the *gender* function in the *gender* R package).

The papers in the data were already assigned to a theme by the authors when completing the submission form. The possible options of themes in the submission form changed over time, with a total of 31 theme titles across the six EPCs. Many of the changes in the theme titles were relatively minor. Based on similarities of the research area as well as the frequency distribution of the number of papers (see Table A1 in the appendix) we reclassified the 31 themes of oral presentations into seven topics of research: a) *Ageing, Health and Mortality*, b) *Data and Methods*, c) *Economic and Policy Issues*, d) *Fertility and Family*,

e) *History, Development and Environment*, f) *Migration*, and g) *Life Course*. In most cases this classification was fairly straight forward, following variants of theme names that have consistently appeared in EPCs over the period. In some years the themes straddled two or more of our seven topics, in which case we allocated based on the titles of the session and the papers presented. An eighth topic h) *Poster* is used for our analyses related to all papers presented in a poster session.

The final cleaned data set covered a ten-year period of six EPCs involving 2,751 oral papers and 1,445 posters presented in 251 sessions. Summary counts of the papers, authors and authors-year can be found in the first three columns of Table 1.

[TABLE 1: ABOUT HERE]

Exploratory Analysis

In order to investigate the network of demographers at EPCs we first plotted the geographical distribution of authors by cities where the institutions of their main affiliation are located in the world and in Europe, in the top and bottom panels of Figure 1 respectively. In both plots the size of the point represents the number of researchers having their primary affiliation in the city indicated. The majority of the authors at the EPCs were based in European institutions. Author locations were dispersed across large countries such as Italy and Germany as well as smaller countries like Belgium and the Netherlands. In some nations it was possible to identify a concentration of authors in cities where there are large demographic centres such as Rostock, where there is the Max Planck Institute for Demographic Research (MPIDR), Southampton, home of the Centre for Population Change (CPC), Paris, where the National Institute for Demographic Studies (INED) is located, Barcelona, hosting the Centre for Demographic Studies (CED), Vienna, home of the Vienna Institute of Demography (VID), Budapest, where there is the Hungarian Demographic Research Institute (HDRI), and Stockholm, home of the Stockholm University Demography Unit (SUDA). The largest number of authors presenting at the EPCs originating from non-European institutions came from the United States followed by India, Turkey, and Australia.

[FIGURE 1: ABOUT HERE]

Of the 4,196 papers in our data, 1,319 were single author papers. Of the remaining 2,877 papers, 1,500 had authors from more than one institution. In order to obtain a better understanding of the linkages between institutions represented at the EPCs we used the *circlize* package in R to produce a chord diagram plot shown in Figure 2 (Gu 2016). In order to minimise visual clutter, the plot contains only 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.

[FIGURE 2: ABOUT HERE]

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 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. 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).

However, international collaborations are not uncommon; 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. 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. Non-European institutions in Iran and India tend to have strong collaborative links within their own country. The largest collaboration across institutions located in different countries is between the Free University of Brussels and NIDI. Other large demographic institutes such as MPIDR and INED have cross-institute collaborations with many multiple institutes both within and beyond Germany and France respectively.

Determinants of Co-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, with a focus on the methods and results; our interpretations of the results and their relationship with existing literature on collaborative authorships and demographers are left to the following section. First, we examine the factors relating to collaboration (i.e., papers that have multiple authors). Second, we explore the driving forces behind multiple institutional co-authorship (i.e., papers that involve authors from more than one institution).

As an initial step to investigate the collaboration of demographers at the EPC, we explore which papers involve 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.

The explanatory variables used into our analysis were motivated by including factors highlighted as relevant in past research on collaborative networks in the social sciences, and additional aspects available in the data that we believed might provide additional explanatory power in our conference-specific context. They can be categorised into three areas.

The first set of variables included characteristics of the paper such as the year of the conference in which the paper was presented and the topic of the paper itself based upon our classification described in the previous section. The year of the conference was included to capture the time trend while the topic of the paper can reflect the nature of collaboration in social sciences which is mainly based on methodological foundation (Moody 2004).

The second set of variables includes characteristics of the author: gender, whether they were 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. Little is known about gender differences in collaboration intensity in demography. In other disciplines gender has been found to be a key determinant of collaboration patterns with variations depending on specific discipline and subfield (Bird 2011; Rigg et al. 2012). The number of other papers presented at the same conference can be used as a proxy for productivity. Collaboration is found to be positively associated with productivity (Laband and Tollison 2000). Since institutional evaluations and tenure decisions are often determined by the number of publications, collaboration frequencies have been found to stimulate intellectual collaboration to increase the sum of research produced (Ductor 2015). It is therefore possible that the number of papers presented at the same EPC is positively associated with co-authorship. In other fields, intra-institutional collaboration has been found to occur more frequently than across institutions (Li et al. 2016). In demography a similar finding might occur – particularly in larger institutes, where there is a larger potential pool of collaborators. We use the count of number of colleagues from the same institution at the same conference to study this potential relationship.

The third set of variables are based on author's relations to past and current EPCs. These include the number of affiliations the author has had (counted over the current and past EPCs), the number of past EPCs attended, and a dichotomous variable indicating if the author attended the previous EPC. The number of affiliations the author has had can potentially serve as a proxy for research migration or mobility. Scientific migration has been found to foster innovation, to enable the flow of ideas, and increase productivity in terms of research output and research impact (Gibson and McKenzie 2014). Thus we might expect that authors with higher number of affiliations over the ten-year period will also have higher rate of co-authorship. In fact, previous research finds that the number of collaborators is positively associated with academic age (Gingras et al. 2008). An author number of past affiliations, as well as their previous attendance at past EPCs provide a proxy for each author's research experience and duration of their academic career.

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 a single superior 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.

Figure 3 shows the parameter estimates from the model fitted to all covariates. The parameter uncertainties illustrated using the error bars correspond to 1.96 times the standard error. In sets of categorical covariates we include the baseline category, where the parameter estimate is set to zero. The posterior inclusion probabilities for the covariates are indicated via the background shading of the panel.

There are four covariates with posterior probabilities greater than 0.5: the intercept, 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 for each of these covariates as predictors of our outcome variable – the probability of an individual having co-author(s) on a paper at an EPC. 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.

[FIGURE 3: ABOUT HERE]

The first predictively optimal parameter, the intercept, has a parameter estimate of 0.37. This value corresponds to the average log-odds of an EPC paper involving multiple authors in the baseline categories (at EPC 2006, in the poster session, with a female non-European author who attended the previous EPC) and has a zero for the remaining continuous measures. For this combination, the log-odds correspond to a 0.59 probability of an author-paper combination being part of a multiple authorship paper.

The second predictively optimal parameter is the EPC year. In comparison to the EPC 2006 baseline category all parameters have log-odds values greater than zero indicating larger probabilities of papers in more recent EPCs involving multiple authors. The parameters form a broadly increasing trend, with the highest log-odds values corresponding to the most recent EPC.

The third predictive optimal parameter is related to the topic of the paper. Papers in poster sessions are used as a baseline category allowing comparison to papers on specific topics that were presented as talks. As the poster session comprised of papers across a mixture of topics, it forms a synthetic average of papers across sub-disciplines. The parameters related to papers in topics central to traditional demographic areas such as Ageing, Health and Mortality, Migration, Fertility and Family, and Data and Methods all have log-odds parameter values greater than 0.4 and lower bounds of error bars above zero. An Ageing, Health and Mortality paper has a log-odds of multiple authorship of 0.71 corresponding to an increase in the probability a given author having co-author(s) of 0.67 when all other covariate measures are set to the baseline values or zero.

The fourth predictively optimal parameter is for the number of number of other papers accepted at the same EPC. The log-odds parameter estimate is positive (0.57) indicating that authors with other covariate measures in the baseline category or set to zero see a 0.63 increase in their probability of being involved in a multiple author paper when they have one additional paper at the same EPC.

The remaining parameter estimates are not predictively optimal (i.e., their posterior inclusion probabilities are less than 0.5). Of these, two have parameter estimates with error bars not containing zero. The log-odds parameter value for the measure of past affiliations is -0.22 suggesting that authors with more past affiliations are more likely to be single authors, when all other covariate values are set to zero or the baseline category. The log-odds parameter value for the past number of EPCs is positive (0.12) suggesting an increase in the probability of an author co-authoring papers at a given EPC as the number of past conferences attending increases, and all other parameters remain unchanged.

Determinants of Co-authorship across Institutions

As previously noted, over half of the multiple author papers in our data involved individuals from multiple institutes. In order to investigate the determinants of collaborations between demographers at the EPCs, we filtered the author-paper data 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. We then applied the same model averaging approach to our new data. The resulting parameter estimates from the full model containing all covariate measures are shown in Figure 4.

[FIGURE 4: ABOUT HERE]

There are eight covariates with posterior probabilities greater than 0.5: the intercept, the EPC year, the number of co-authors, the gender composition of the authorship team, the location of the authors, the number of other papers by the author at EPC, the number of colleagues from the same institute at the EPC, and the number of past affiliation of an individual.

The first predictively optimal parameter, the intercept, has a parameter estimate of -1.02. This value corresponds to an average probability of 0.26 of a multiple-author EPC paper involving multiple institutions in the baseline categories (at EPC 2006, in the poster session, with a female non-European author, with only female co-authors, who attended the previous EPC) and zero counts for the remaining continuous measures.

The second predictively optimal parameter is based on the EPC year. In all but 2008, the log-odds values are less than zero indicating smaller probabilities of multiple-author papers in more recent EPCs involving more institution in comparison to the EPC 2006 (the baseline category). The error bars of all parameters

values, except for EPC 2016 are crossing zero suggesting the probability of multi-author papers coming from multiple institutions has remained stable over time.

In the third, sixth and eight predictively optimal parameters for the number of co-authors, number of other papers at the EPC and number of past affiliations are all greater than zero. The parameter values for the number of other papers is considerably smaller than for the other two measures, suggesting it has a small effect on the probability of co-authors. The log-odds parameter value for the number of co-authors (0.66) indicates that the addition of one co-author increases the marginal probability by 0.67 of the paper involving multiple institutes, where all other covariate measures are set to the baseline category or to zero. A similar size effect is found for the number of past affiliations (log odds of 0.64).

The fourth and fifth optimal parameters suggest the log-odds of multiple institutions increase for 1) papers that involve either all males or a mixed gender set of researchers in comparisons to papers that involve only females and 2) papers that involve non-European researchers (when all other parameters are set to zero). The seventh predictively optimal parameter is very close to zero (-0.02) indicating a slight reduction in the log-odds of a paper with authors from multiple institutions if an author has a large number of other colleagues at the EPC.

Discussion

The analysis of the determinants of co-authorship in general and co-authorship across institutions reveal key aspects in the collaboration amongst demographers in the past decade. First, formal research collaboration as measured by multiple authorship in papers presented at the EPC has risen overtime. This trend is similar to other social science disciplines such as economics (Laband and Tollison 2000; Sutter and Kocher 2004), political science (Fisher et al. 1998), and sociology (Hunter and Leahey 2008; Pontille 2003). Previous studies find that co-authored papers more frequently involve empirical studies, quantitative research and/or survey than theoretical, interview and qualitative work (Henriksen 2018; Hunter and Leahey 2008). Being a traditionally quantitative discipline, rising co-authorship in demography is not unexpected as statistical techniques advance and data quality and quantities increase over time. The data suggest co-authored papers are likely to involve authors from different institutions. Likewise, previous studies suggest that, in a social science context, demography has on average a high share of co-authored articles, especially with international co-authorship (Henriksen 2016). We find that in 2006, the share of papers with co-authors from different institutions was 68%, declining slightly to 65% in 2016. Our finding that among co-authored papers, collaboration across institutions remains unchanged in the past ten years possibly reflects the fact that cross-institutional co-authorship in demography has already been at a high level.

The second key aspect that we found in the determinants of co-authorship relates to subfields. Papers in core demographic subfields including a) *Fertility and Family*, b) *Ageing, Health and Mortality*, c) *Migration*, and d) *Data and Methods* are more likely to be co-authored than those in other themes. This is possibly because in demography, research in these fields and especially that related to fertility, has a higher chance of being funded than other themes (Riley and McCarthy 2003). This consequently is likely to result in collaborative work. Amongst co-authored papers, papers in *Data and Methods* strands are less likely be

carried out by researchers from multiple institutions. In sociology and other social science disciplines, empirical papers based on quantitative methods have found to be more likely to involve multiple authors than theoretical papers (Henriksen 2018; Hunter and Leahey 2008; Moody 2004). As demographic research is more dominated by analysis of empirical data than sociology (Burch 2018), research related to data and methods could be considered as ‘theoretical’ in demography and hence result in less collaboration across institutions.

A third notable aspect in the determinants of collaborations in the EPC data relates to the authors’ gender. 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 transnational movements for females more than for males (Abramo et al. 2013). Shauman and Xie (1996) noted that the presence of children and a working husband reduce international mobility of female academics. Correspondingly, the authors with higher number of past affiliations – a proxy for mobility – also have higher probability of collaborating across institutions.

There are a number of limitations in our study that could potentially be addressed in further research efforts. First, although the data on papers presented at conferences cover a wider range of demographers than bibliographic database of published papers, they remain subject to selection bias. We are not able to observe collaboration networks of authors whose papers were not selected for presentation at the EPC. Furthermore, papers submitted to and presented at EPC conferences are predominantly produced by authors based in Europe. Our study therefore does not cover the entire networks of demographers and the findings on collaboration patterns are primarily limited to those of Europe-based demographers. Data from other international conferences such as the International Population Conference would enable a wider study to capture a larger network of demographers. Likewise, our findings on the trends in collaboration networks of demographers is limited by the ten year time-span in our data, and hence we cannot reflect historical development of collaboration in the field of demography. Finally, the categorisation of papers into different research topics is not entirely objective. The different research themes were merged together based on our subjective criteria.

Despite the upward trend in research collaboration in social sciences, less is known about factors driving collaboration in demography especially among European demographers. Using a novel data set that provides a unique window into the network of collaborations among demographers, this study contributes to improve our understanding of the discipline of demography. Ties among researchers and institutions at six European Population Conferences over the course of a decade are evaluated. We identified differences in factors such as research subfields and gender composition in multi-author papers plays a key role in determining the nature of collaboration across institutions.

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
- Adams, W. C., Infeld, D. L., Minnichelli, L. F., & Ruddell, M. W. (2014). Policy Journal Trends and Tensions: JPAM and PSJ. *Policy Studies Journal*, *42*, S118–S137. doi:10.1111/psj.12051
- 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
- Bird, D. K. S. (2011). Do women publish fewer journal articles than men? Sex differences in publication productivity in the social sciences. *British Journal of Sociology of Education*, *32*(6), 921–937.
doi:10.1080/01425692.2011.596387
- Burch, T. K. (2018). Data, Models, Theory and Reality: The Structure of Demographic Knowledge. In *Model-Based Demography* (pp. 21–42). Cham: Springer International Publishing.
doi:10.1007/978-3-319-65433-1_2
- Ductor, L. (2015). Does Co-authorship Lead to Higher Academic Productivity? *Oxford Bulletin of Economics and Statistics*, *77*(3), 385–407. doi:10.1111/obes.12070
- Fisher, B. S., Cobane, C. T., Ven, T. M. V., & Cullen, F. T. (1998). How many authors does it take to publish an article? Trends and patterns in political science. *PS: Political Science and Politics*, *31*(4), 847–856. doi:10.2307/420730
- Gibson, J., & McKenzie, D. (2014). Scientific mobility and knowledge networks in high emigration countries: Evidence from the Pacific. *Research Policy*, *43*(9), 1486–1495.
doi:10.1016/j.respol.2014.04.005

- Gingras, Y., Larivière, V., Macaluso, B., & Robitaille, J.-P. (2008). The effects of aging on researchers' publication and citation patterns. *PLoS ONE*, 3(12). doi:10.1371/journal.pone.0004048
- 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>
- Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2), 455–476. doi:10.1007/s11192-016-1849-x
- Henriksen, D. (2018). What factors are associated with increasing co-authorship in the social sciences? A case study of Danish Economics and Political Science. *Scientometrics*, 1–27. doi:10.1007/s11192-017-2635-0
- Hunter, L., & Leahey, E. (2008). Collaborative Research in Sociology: Trends and Contributing Factors. *The American Sociologist*, 39(4), 290–306. doi:10.1007/s12108-008-9042-1
- 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
- Laband, D. N., & Tollison, R. D. (2000). Intellectual Collaboration. *Journal of Political Economy*, 108(3), 632–662. doi:10.1086/262132

- 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
- Li, L., Catalá-López, F., Alonso-Arroyo, A., Tian, J., Aleixandre-Benavent, R., Pieper, D., et al. (2016). The Global Research Collaboration of Network Meta-Analysis: A Social Network Analysis. *PLOS ONE*, *11*(9), e0163239. doi:10.1371/journal.pone.0163239
- 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
- Moody, J. (2004). The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review*, *69*(2), 213–238. doi:10.1177/000312240406900204
- Mullen, L., Blevins, C., & Schmidt, B. (2015). *Predict Gender from Names Using Historical Data*. <https://github.com/ropensci/gender>
- Pontille, D. (2003). Authorship Practices and Institutional Contexts in Sociology: Elements for a Comparison of the United States and France. *Science, Technology, & Human Values*, *28*(2), 217–243.
- 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>
- Rigg, L. S., McCarragher, S., & Kremenec, A. (2012). Authorship, Collaboration, and Gender: Fifteen Years of Publication Productivity in Selected Geography Journals. *The Professional Geographer*, *64*(4), 491–502. doi:10.1080/00330124.2011.611434

Riley, N. E., & McCarthy, J. (2003). *Demography in the age of the postmodern*. Cambridge: Cambridge University Press.

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

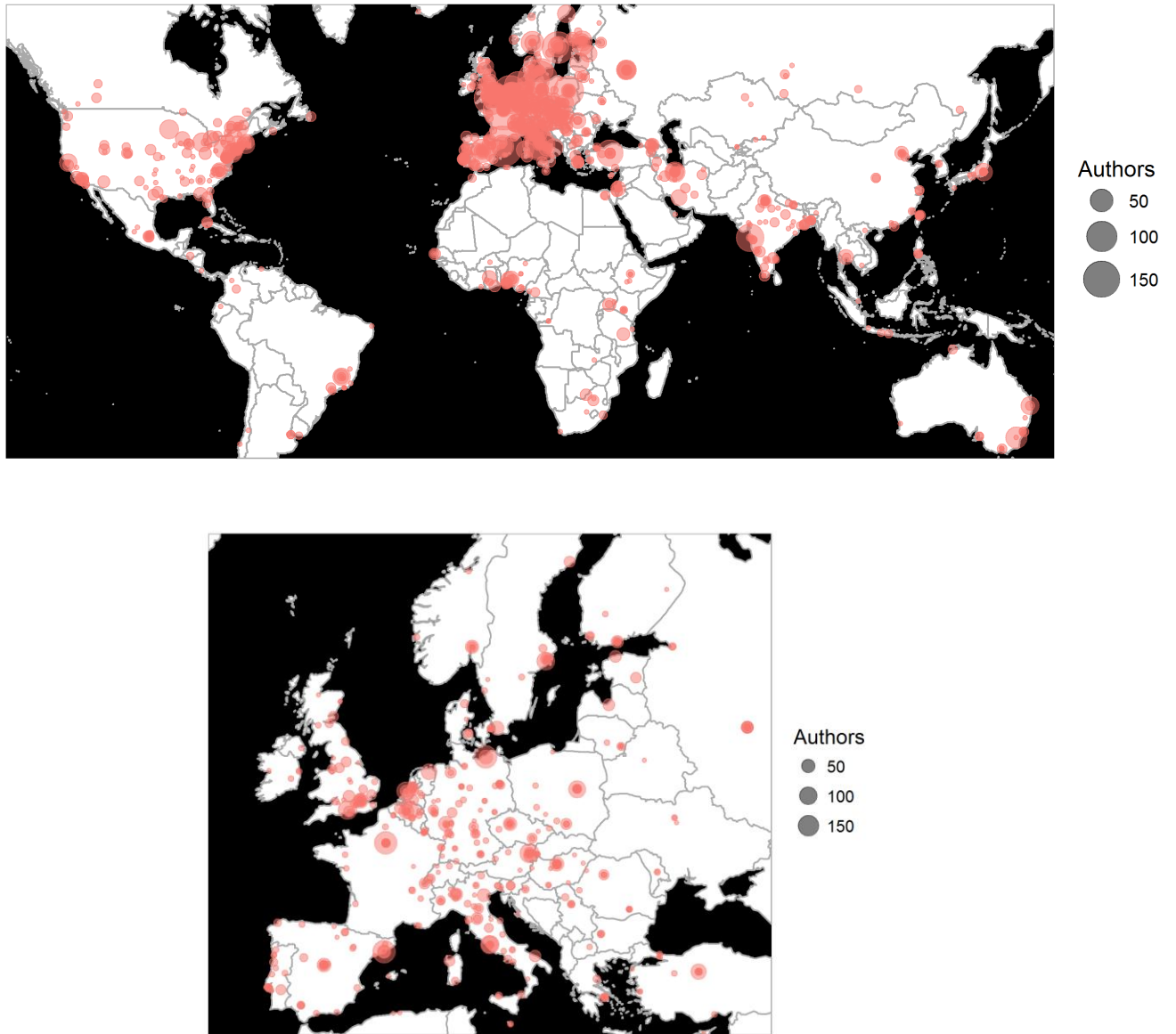
Sutter, M., & Kocher, M. (2004). Patterns of co-authorship among economics departments in the USA. *Applied Economics*, 36(4), 327–333. doi:10.1080/00036840410001674259

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

Acknowledgements

We are grateful to the European Association for Population Studies (EAPS) for granting us access to the data and the supporting staff for PAMPA 5.0 for supplying us the data in an electronic format.

Figure 1: Locations of authors: global and in Europe



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

Figure 3: Logistic regression estimates of the probability of multiple authorship (coefficients and 95% CI)

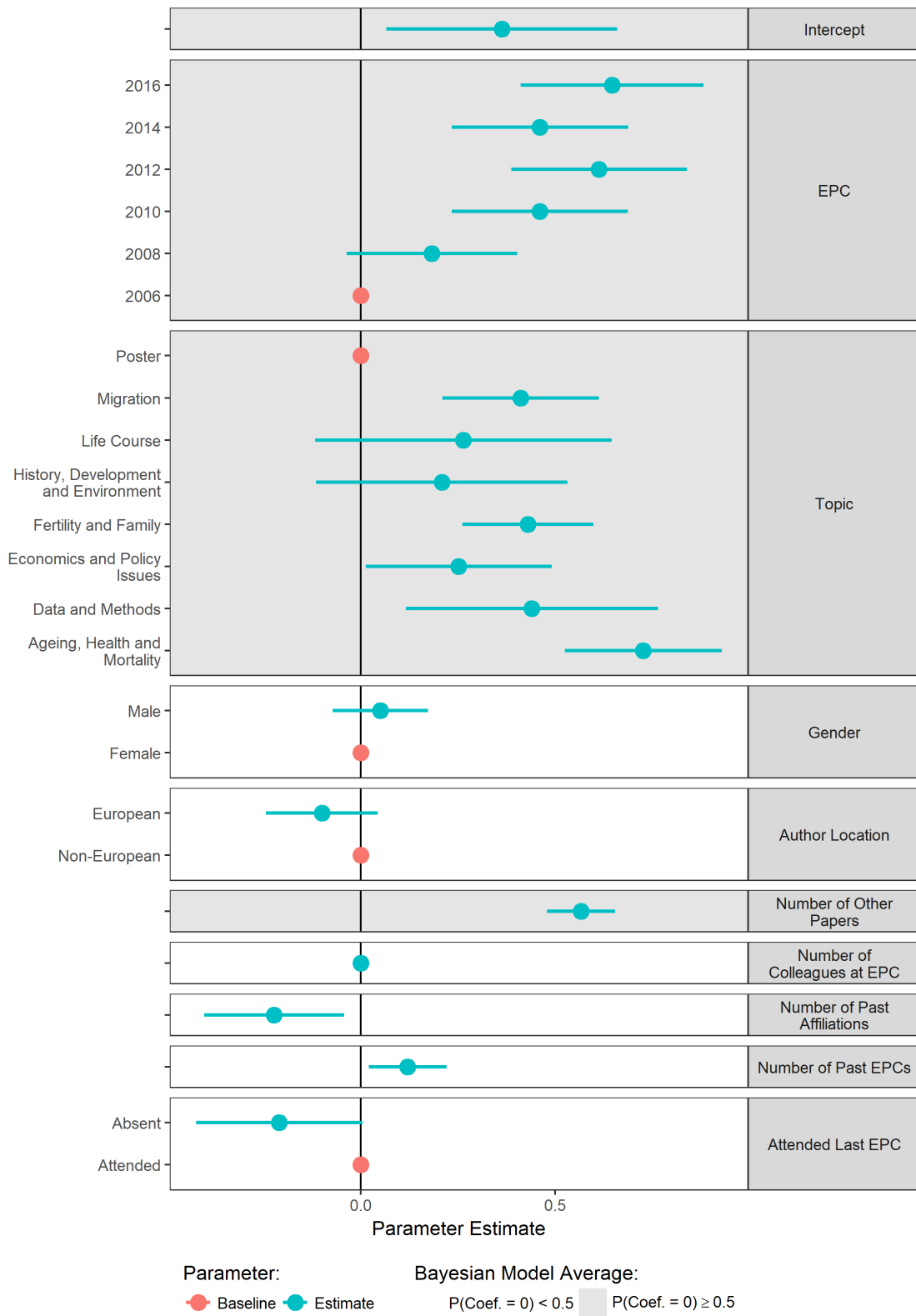


Figure 4: Logistic regression estimates of the probability of co-authorship with authors from other institutions (coefficients and 95% CI)

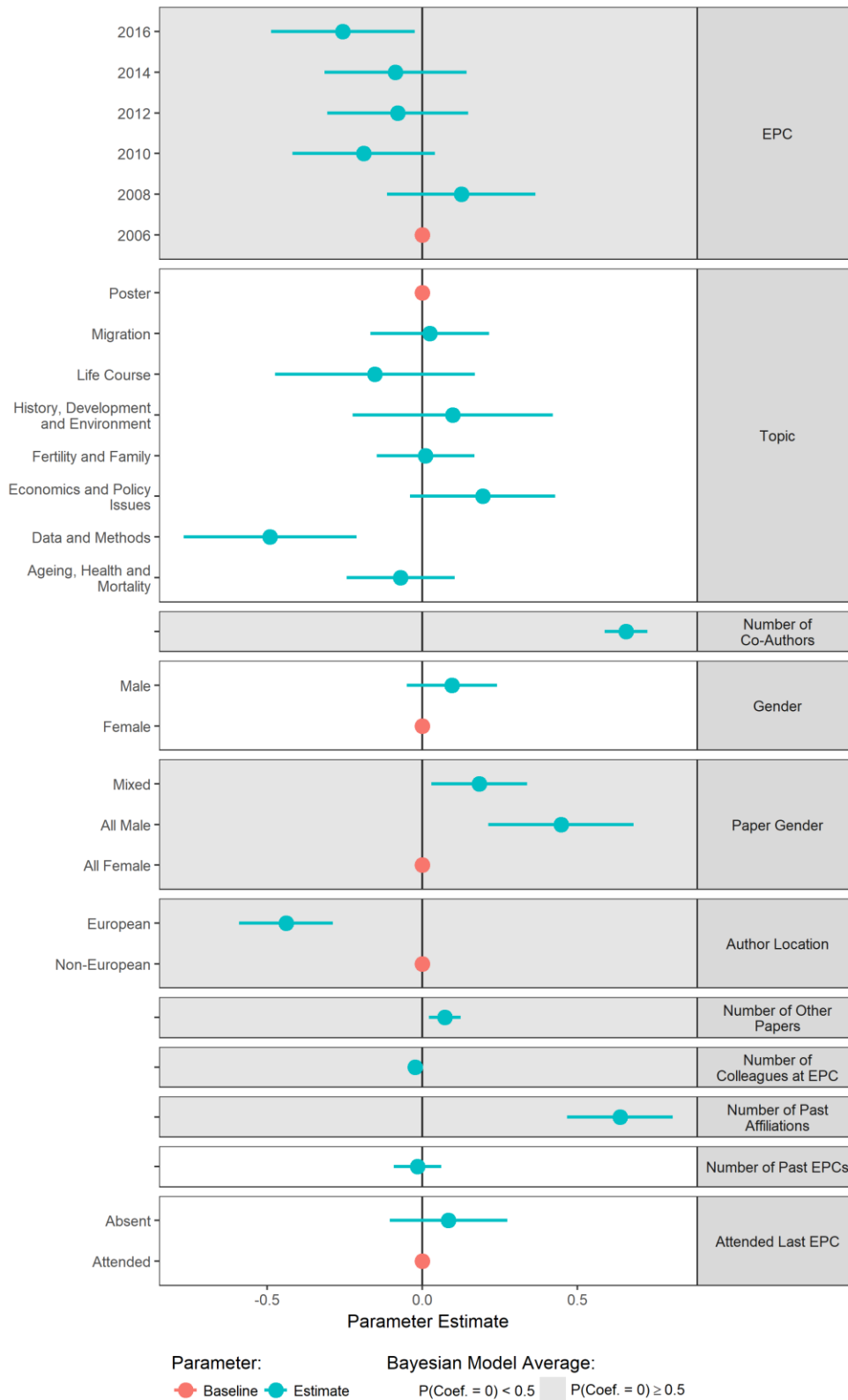


Table 1: Summary counts of authors, papers and author-paper panel data from past EPCs

Variable	Papers	Authors	Author-Year	Author-Paper ¹	Author-Paper ²
<i>Outcome Variable</i>					
Single ³				1303	1768
Multiple ⁴				7820	6049
<i>EPC</i>					
2006	506	757		1023	823
2008	669	1037		1339	1090
2010	677	1107		1418	1211
2012	787	1334		1771	1547
2014	787	1225		1750	1522
2016	771	1267		1822	1624
<i>Topic</i>					
Poster Sessions	1445	1923		2897	2352
Ageing, Health and Mortality	634	924		1465	1326
Data and Methods	164	281		396	346
Economics and Policy Issues	281	477		684	588
Fertility and Family History	888	1108		1997	1755
Development and Environment	153	245		317	267
Life Course	124	220		267	233
Migration	507	755		1100	950
<i>Gender</i>					
Male		1866		4348	3753
Female		2108		4775	4064
<i>Paper Gender</i>					
All Female	1360				1538
All Male	1117				1236
Mixed	1714				5043
<i>Paper Country</i>					
Same	3377				5304
Mixed	819				2513
<i>Author Location</i>					
European		2565		6740	5808
Non-European		1450		2383	2009
<i>Number of Other Papers</i>					
0		5104		5058	4099
1		1087		2158	1895
2		357		1073	1016
3+		179		834	807
<i>Number of Colleagues at EPC</i>					

0	1026	1144	814
1	350	829	721
2	186	734	640
3+	511	6416	5642
<i>Number of Past Affiliations</i>			
1	5958	7758	6615
2	697	1204	1056
3	69	148	133
4+	5	13	13
<i>Number of Past EPCs</i>			
1		3999	3915
2		1333	1585
3		710	1030
4+		685	1287
<i>Attended Last EPC</i>			
Yes		2728	3915
No		3999	3902

Notes: ¹ Panel data for models based on outcome variable for single or multiple authorship

² Panel data for models based on outcome variable for single or multiple institutions in multiple author papers

³ In penultimate column refers to count of single author papers. In final column refers to authors with co-authors in the same institution.

⁴ In penultimate column refers to count of author appearances in multiple author papers. In the final column refers to authors with co-authors in different institutions.

Authors:**1. Guy J. Abel***

Asian Demographic Research Institute (ADRI), Shanghai University, Shanghai, China
Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria

2. Raya Muttarak*

School of International Development, University of East Anglia, UK and Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis, Laxenburg, Austria.

3. Valeria Bordone

Institut für Soziologie, Ludwig-Maximilians-Universität München, Munich, Germany

4. Emilio Zagheni

Department of Sociology, University of Washington, Seattle, USA

*Notes: Guy J. Abel and Raya Muttarak are co-first authors