Scholarly Migration Database

Quantifying the Mobility of Scholars

# Scholarly Migration Database Methods and Documentation

Version 2022_V1


- Aliakbar Akbaritabar, Max Planck Institute for Demographic Research (MPIDR)
- Tom Theile, Max Planck Institute for Demographic Research (MPIDR)
- Emilio Zagheni, Max Planck Institute for Demographic Research (MPIDR)


March 14, 2023

# Contents

# 1 Summary

The Scholarly Migration Database (available at: https://www.ScholarlyMigration.org) is prepared, maintained and publicly released by the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany. It uses and re-purposes bibliometric data from scientific publications to prepare *stock* and *flow* data on scholarly migration events worldwide.

The meta-data of publications come from Scopus 2021 provided to the Scholarly Migration Database by the "Kompetenzzentrum Bibliometrie (Competence Centre for Bibliometrics)" (2021) through the Max Planck Digital Library (MPDL). The methodology of re-purposing this data and calculating measures is described in detail in the Scholarly Migration Database Methods and Documentation protocol (current document). The most updated *version* of the database and documentation on **March 14, 2023** is: **2022_V1**.

# 2 Scholarly Migration Database Team

Members of the Scholarly Migration Database *Core Team* who contributed to the version 2022_V1 of the database and methods and documentation include the following list (in alphabetical order):

- Aliakbar Akbaritabar, Max Planck Institute for Demographic Research (MPIDR)
- Tom Theile, Max Planck Institute for Demographic Research (MPIDR)
- Emilio Zagheni, Max Planck Institute for Demographic Research (MPIDR)

The core team has been supported by the following *current and former collaborators* (in alphabetical order):

- Samin Aref (University of Toronto)
- Francesco Billari (Bocconi University)
- Maciej J. Dańko (MPIDR)
- Ridhi Kashyap (University of Oxford)
- Andrea Miranda Gonzalez (University of California, Berkeley)
- László Németh (MPIDR)
- Asli Ebru Sanlitürk (MPIDR)

- Guy Stecklov (University of British Columbia)
- Jevin D. West (University of Washington)
- Xinyi Zhao (MPIDR, University of Oxford)

The core team has been advised by the following *advisory board members* (in alphabetical order):

- Vincent Larivière (Université de Montréal)
- Cassidy R. Sugimoto (Georgia Institute of Technology)

To ask questions or communicate with the Scholarly Migration Database team, please use this e-mail address scholarlymigration@demogr.mpg.de.

# 3 Introduction

Bibliometric data has proven useful for demographic research on the specific case of scholars as a subset of the high-skilled population (Alburez-Gutierrez et al., 2019; Kashyap et al., 2022). Re-purposing these data and using academic affiliation addresses, allows constructing mobility trajectory of academics (Miranda-González et al., 2020; Moed et al., 2013; Moed & Halevi, 2014; Zhao et al., 2022). The Scholarly Migration Database uses these data as a novel source of *digital traces* (Kashyap et al., 2022) and re-purposes them to answer questions regarding high-skilled and, specifically, scholarly migration flows and stocks. The geographic scope of the current database includes all countries worldwide for which the data is available in Scopus.

## 3.1 Examples of research using Scholarly Migration Database

As illustrative examples of the type of research that is possible with the Scholarly Migration Database, consider these examples that have used previous versions of the database. While these studies were focused on specific national contexts, the Scholarly Migration Database currently covers all countries worldwide, which are available in Scopus, and enables comparative studies at a global scale.

Miranda-González et al. (2020) offered one of the few studies of internal scholarly migration using bibliometric data of all Scopus-published Mexican scholars from 1996-2018 and their mobility between regions of Mexico. They found that most of the scholars do not move and the capital, i.e., Mexico City, was the most preferred destination of emigrants.

Zhao et al. (2022) investigated Scopus published German scholars from 1996-2020, finding that fewer migrant women scholars return to Germany.

Subbotin and Aref (2021) investigated Scopus-published Russian scholars from 1996-2020 finding that mobile scholars account for 5% of all scholars affiliated to Russia, and in recent years, the so-called *brain drain* from Russia is replaced with a more balanced *brain circulation*.

Şanlıtürk et al. (2022) studied the initial changes in the British academic environment after the Brexit referendum. The study shows evidence that after Brexit, scholars who started their academic careers in the EU countries have a higher probability to leave the UK, while scholars who started their academic careers in the UK have a higher probability to return to the UK. The results signal a compositional change rather than a brain drain in the British academic environment, in the years following the Brexit referendum.

Sanliturk et al. (2023) studied if the migration of scholars worldwide associates significantly with the economic development of countries (in terms of GDP per capita). Emigration propensity, on average, initially increases with economic development. They found the opposite pattern for the migration of scholars. Despite the reported inverse U-shape pattern of migration for the general population, in case of the academics, a U-shape pattern is observed. This means by increasing GDP, the migration of scholars first decreases and then it starts to increase in rich countries which could signal the return migration of graduates to their home countries.

Zhao et al. (2023) provided a gender perspective on the migration of scholars worldwide. They addressed a gap concerning the migration of scholars, whether male and female scholars participate equally in transnational mobility and how these patterns have shifted over time from a global perspective. They found that, while female researchers continued to be underrepresented among internationally mobile researchers and migrate over shorter distances, this gender gap was narrowing at a faster rate than the gender gap in the population of general active researchers.

# 4 Bibliometric database

The current version of the Scholarly Migration Database uses Scopus 2021 publications. The data snapshot usually is frozen in April of the current year and would only include the first four months of the latest year. Every year, with the release of a new data snapshot, our database of scholarly migrations is updated and new rates and estimates are provided publicly.

## 4.1 Selection criteria for publications

Scopus 2021 publications are limited to only *Article* and *Review* document types to ensure the highest quality of metadata. In addition, affiliation addresses delivered by Scopus 2021 are limited to only *author* affiliation addresses to exclude publisher and other types of addresses which are less relevant to trace scholarly migration.

## 4.2 Scientific entity name disambiguation

It is necessary to ensure that the bibliometric data used has sufficient quality. Lack of proper data or lower quality of metadata causes errors in identifying entities (e.g., authors or academic affiliations) and migration events. In other words, failing to identify scholars properly could cause a merger between different individuals' mobility trajectories. Organizational and academic addresses (i.e., affiliations) need to be correct for the migration event identification to work and to be reliable.

### 4.2.1 Author name

For author name disambiguation, we use identification numbers added to each unique author by Scopus (Baas et al., 2020). Authors who do not have a Scopus author ID or are not indicated as *disambiguated* (i.e., active profiles) by Scopus are excluded from our analysis and the Scholarly Migration Database.

The author_id identifies all publications of a single author in 94.4% of cases (recall) and has a precision of 98.1%, which means that records of two different authors could be merged by mistake under one author_id only in 1.9% of the cases. Precision and recall rates are quoted from Scopus and the study published by Baas et al. (2020), which includes more detail on the disambiguation process and Scopus meta-data.

### 4.2.2 Organization name

Organization names are disambiguated using the research organization registry (ROR) API. We use the full affiliation string from Scopus to geo-code it to different granular levels. For instance, "Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany" is one affiliation address. But different authors who use this affiliation might write it with a different set of details e.g., to include or exclude city or country names or some might add department and laboratory names. Hence, different versions of this address need to be unified under a unique affiliation identification number to reduce the error in

identifying a change in affiliation addresses which are used here as a proxy for a residential address change, i.e., a migration event. See Akbaritabar (2021) for a description of the used methodology and a comparison of its performance with other organization name disambiguation methods.

# 5 Pre-processing steps

Bibliometric data as delivered by the database owner to the "Kompetenzzentrum Bibliometrie (Competence Centre for Bibliometrics)" (2021) and through them to us needs pre-processing to allow identification of migration events.

## 5.1 Construction of authorship records

For each author, we obtained the list of all publications and we processed the metadata to assign a date to each author-publication-affiliation-triplet. We call this triplet an authorship record (Aref et al., 2019). For example, a paper authored by John Doe and Jane Doe, where John Doe has the affiliations "1" and "2", while Jane Doe has only the affiliation "2", will result in three authorship records as shown in Table 1.

Table 1: An example of one publication and its respective authorship records

| Publication Title | Author Full Name | Affiliation | Publication Year |
|---|---|---|---|
| Paper 1 title | John Doe | Affiliation 1, Country 1 | 2020 |
| Paper 1 title | John Doe | Affiliation 2, Country 2 | 2020 |
| Paper 1 title | Jane Doe | Affiliation 2, Country 2 | 2020 |

The affiliation includes the address and the country and in some cases an affiliation_id, which identifies the same institutions, even if their names are spelt differently. The address information is available only in 87% of the records, but the country information is available in 99% of the records (Baas et al., 2020).

## 5.2 Country of origin assignment

Bibliometric data does not include the country of origin of the authors. It only indicates the authors' countries and addresses of affiliation at the time of publication that could be used as a proxy for their residential addresses. Previous research has used the author's family name to find proxies for country of origin (Basu, 2013; Begum et al., 2017; Grilli & Allesina, 2017;

Ioannidis et al., 2021; Lewison et al., 2016). We are currently working on a deep learning based method to assign the country of origin to the full name of an author. Our method is similar to Hu et al., 2021 and uses a character-based encoding of the authors' normalized *full* name as input. It does a multi-class classification task on it and exports the country of origin as output. The neural network model is implemented in Tensorflow (Abadi et al., 2016) and has three one-dimensional-convolutional and three dense layers and in total about two million parameters. The model is trained on two data sets that we constructed ourselves: 1) From the Scopus data and 2) an extracted subset from Wikidata[1], which is publicly available. The Scopus data set is built from all author names that appear predominantly in only one country. Since these names do not appear in other countries, we are assuming them as typical names for these countries. The Wikidata data set is built from all entities indicated as *human* in the Wikidata database (Vrandečić, 2012). These entities have a name and an assigned country or country of birth. To evaluate the performance of the model, we extract two subsets of the data (i.e., manually constructed data using Scopus and a subset of Wikidata's human entities) and use them as validation data.

This metadata will be added to the Scholarly Migration Database in future versions if the testing and evaluation of the quality of results are successful. We will present comparison results on the accuracy of our algorithm in assigning the author's country of origin versus our validation data sets and manually assigned gold standards.

## 5.3 Gender assignment

We use a similar neural network model to the one described in the country of origin assignment (5.2) section to infer the gender using an author's name. We used another dataset extracted from Wikidata that includes all entities indicated as *human* with a name and an assigned gender of male or female. The number of females are much smaller than males, hence, we select a subset of the males to have a balanced data set which prevents a biased model. In addition, we used another data set with more than four thousand manually assigned gender from Akbaritabar and Squazzoni, 2020 to evaluate our method's results.

This metadata will be added to the Scholarly Migration Database in future versions if the testing and evaluation of the quality of results are successful. We will present comparison results on the accuracy of our algorithm in assigning the author's gender versus manually assigned gold standards.

[1] https://www.wikidata.org/wiki/Wikidata:Main_Page

# 6 Identifying migration events

Migration events could be identified in different ways. One can include all affiliations per author and any change in those affiliations as a proxy for a *travel* and *mobility*, or consider only long-term *migration* events. For the current version (2022_V1) of the Scholarly Migration Database, we use a mode-based method to identify migration events.

## 6.1 Mode-based method

The migration counts are based on the aggregation of all migration events that we could detect using changes in the affiliation addresses of authors.

After the pre-processing of the raw authorship records, as described in Section 5, we determined the country of residence for every author and every year. More specifically, we considered all the affiliation countries of an author_id in one year. If there is more than one country, we take the mode of all countries. If there is more than one mode, we check whether one of the modes was the previous country of residence and take that country if that was the case. If it were not, we would choose one of the mode countries randomly.

To determine migration events, an algorithm then goes through the years and logs a migration event when the country of residence changes. We assume a *two years* preparation time for all publications to cover disciplinary differences in publication delay (Björk & Solomon, 2013). If there are gaps in publication years (e.g., authors are not publishing continuously), we backward fill each publication year for two years and assume the author's residence to have changed two years earlier. If there is enough evidence (i.e., continuous publication activity), we consider the year when the modal affiliation changes as the migration year.

## 6.2 Nominator and denominator populations

Once we have detected all migration events, we aggregate them by country and year into emigration counts. To generate measures of exposures (i.e., the denominators for migration rates, or the population size of researchers per country and year), we counted the number of active scholars for each year and country.

Active scholars in a given year include those who publish at least an article or a review during that year. In addition, and to deal with missing observations, we assumed that an author who did not publish in a particular year was still part of the population of active scholars if he or she published

one or two years before. Finally, we excluded authors who had *only* one indexed publication during their entire career from the denominator of the scholars' population. The reason for this exclusion is twofold: first, these scholars could be junior researchers who have graduated or those who leave academia. Since we do not have a live *census* of all academics globally, we cannot consider them as part of the pool of *active scholars*. Furthermore, in each given year, there is a fraction of scholars who enter the pool of active publishers (by having their first publication in the sample) and exit this pool of publishing scholars in the next years (Akbaritabar et al., 2020). Counting them among active scholars would over-inflate the population of scholars and cause our measures to be artificially smaller. Second, because by definition of the mode country per year, these scholars who had publications only in one year could not have migrated (i.e., contributed to the nominator), hence, it is reasonable to exclude them from the denominator.

# 7 Measures

To evaluate the exposure of populations to migration events, we calculate different measures that are presented in the visualizations in the Scholarly Migration Database.

We calculate *in-migration* (equation 1), *out-migration* (equation 2), and *net migration* count (equation 3) and rate (equation 4) (Bell et al., 2002; Miranda-González et al., 2020) as follows:

$$IMR_{i,t} = \frac{I_{i,t}}{N_{i,t}} \tag{1}$$

$$EMR_{i,t} = \frac{E_{i,t}}{N_{i,t}} \tag{2}$$

$$NM_{i,t} = I_{i,t} - E_{i,t} \tag{3}$$

$$NMR_{i,t} = \frac{I_{i,t} - E_{i,t}}{N_{i,t}} \tag{4}$$

where $i$ is the country, $t$ is the year, $I_{i,t}$ is the inflow of scholars entering a country and $E_{i,t}$ is the outflow of scholars exiting that country over the total number of scholars in the country in a given year, i.e., $N_{i,t}$.

# 8 How to use, cite, and contribute

The Scholarly Migration Database must be used for research purposes only. No commercial use is allowed. It is not allowed to use this data for reporting purposes.

If you use our replication data-sets, we ask you to cite this document (see 8.4) and our publications (introduced on Scholarly Migration Database website), and copy and mention the text in the "Data license", 8.1, section which clarifies our data source.

Do not pass on replication data-set files to other researchers. Instead, please ask them to check the Scholarly Migration Database website (available at: https://www.ScholarlyMigration.org) to access the data.

We continuously update the website and documentation based on new developments, correct errors and introduce new publications with replication data. Using the latest version of the replication data introduced on our website will ensure replicability and prevent mistakes. This practice helps to prevent multiple outdated or incorrect versions. It also ensures that each researcher referring to our website has full access to the information regarding the data source, citation procedures, and any other useful documentation.

## 8.1 Data license

The original bibliometric data is licensed by Elsevier Scopus 2021 and provided to the Scholarly Migration Database by the “Kompetenzzentrum Bibliometrie (Competence Centre for Bibliometrics)” (2021) (project "Kompetenzzentrum Bibliometrie" grant number 16WIK2101A) through the Max Planck Digital Library (MPDL). Under the original license terms, only aggregated results based on the raw metadata which accompany our scientific publications as replication materials can be made publicly available. No individual data from Elsevier Scopus 2021 can be shared.

## 8.2 How to use the replication data

On the introduction to data page of the Scholarly Migration Database (available at: https://www.ScholarlyMigration.org), you can find links to the replication data. They include different granularity, e.g., flow and stock data at the country level.

## 8.3 How to explore the data

If you wish to explore our replication data in your browser, you can use the interactive visualisations which are included with the Scholarly Migration Database (available at: https://www.ScholarlyMigration.org). It allows you to explore different scenarios and see the interrelation between variables and how they affect visualizations.

## 8.4 How to cite

If you use the Scholarly Migration Database, please cite this document. APA style citation is as follows, and the Bibtex code and the handle are available further below.

**APA style citation**:

```
Akbaritabar, A., Theile, T., and Zagheni, E. (2022).
Scholarly Migration Database, Version 2022_V1.
https://www.ScholarlyMigration.org.
Accessed: INCLUDE-YOUR-ACCESS-DATE
```

**Bibtex code**:

```
@misc{ScholarlyMigrationDatabase2022,
  title = {Scholarly Migration Database, Version 2022_V1},
  copyright = {Scholarly Migration Database},
  url = {https://www.ScholarlyMigration.org},
  abstract = {The Scholarly Migration Database
  (available at: https://www.ScholarlyMigration.org)
  is prepared, maintained and publicly released by the Max
  Planck Institute for Demographic Research (MPIDR) in Rostock,
  Germany. It uses and re-purposes bibliometric data from
  scientific publications to prepare stock and flow data on
  scholarly migration events. The publications come from Scopus
  2021 provided to the Scholarly Migration Database by the
  German Competence Centre for Bibliometrics
(project "Kompetenzzentrum Bibliometrie" grant number 16WIK2101A)
  through the Max Planck Digital Library (MPDL).
  The methodology of re-purposing
  this data and calculating measures is described in detail in
  the Scholarly Migration Database Methods and Documentation
  protocol (available at: https://www.ScholarlyMigration.org).
```

```
      The most updated version of the database and documentation on
      April 27, 2022 is:0.1.},
      urldate = {INCLUDE-YOUR-ACCESS-DATE},
       author = {Akbaritabar, Aliakbar and Theile, Tom, and Zagheni, Emilio},
      year = {2022}
}
```

## 8.5 How to contribute

The Scholarly Migration Database team will be happy to hear about your interest to contribute as a researcher, collaborator, data provider, advisor or financial supporter. Please use the following e-mail address (scholarlymigration@demogr.mpg.de) to contact us, and we will get back to you as soon as possible.

# References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467 [cs]*. Retrieved April 29, 2022, from http://arxiv.org/abs/1603.04467

Akbaritabar, A. (2021). A quantitative view of the structure of institutional scientific collaborations using the example of Berlin. *Quantitative Science Studies*, *2*(2), 753–777. https://doi.org/10.1162/qss_a_00131

Akbaritabar, A., & Squazzoni, F. (2020). Gender patterns of publication in top sociological journals [Publisher: SAGE Publications Inc]. *Science, Technology, & Human Values*. https://doi.org/10.1177/0162243920941588

Akbaritabar, A., Traag, V. A., Caimo, A., & Squazzoni, F. (2020). Italian sociologists: A community of disconnected groups. *Scientometrics*. https://doi.org/10.1007/s11192-020-03555-w

Alburez-Gutierrez, D., Zagheni, E., Aref, S., Gil-Clavel, S., Grow, A., & Negraia, D. V. (2019). Demography in the digital era: New data sources for population research.

Aref, S., Zagheni, E., & West, J. (2019). The demography of the peripatetic researcher: Evidence on highly mobile scholars from the web of science. *International Conference on Social Informatics*, 50–65.

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386.

Basu, A. (2013). Some differences in research publications of indian scientists in india and the diaspora, 1986–2010. *Scientometrics*, *94*(3), 1007–1019. https://doi.org/10.1007/s11192-012-0800-z

Begum, M., Roe, P., Webber, R., & Lewison, G. (2017). UK ethnic minority cancer researchers: Their origins, destinations and sex. *Proceedings of ISSI 2017* (pp. 568–579). University of Wuhan, China.

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., & Hugo, G. (2002). Cross-national comparison of internal migration: Issues and measures [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-985X.t01-1-00247]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *165*(3), 435–464. https://doi.org/10.1111/1467-985X.t01-1-00247

Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, *7*(4), 914–923. https://doi.org/10.1016/j.joi.2013.09.001

Grilli, J., & Allesina, S. (2017). Last name analysis of mobility, gender imbalance, and nepotism across academic systems. *Proceedings of the National Academy of Sciences*, *114*(29), 7600–7605. https://doi.org/10.1073/pnas.1703513114

Hu, Y., Hu, C., Tran, T., Kasturi, T., Joseph, E., & Gillingham, M. (2021). What's in a name? gender classification of names with character based machine learning models. *Data Mining and Knowledge Discovery*, *35*(4), 1537–1563.

Ioannidis, J. P., Koutsioumpa, C., Vakka, A., Agoranos, G., Mantsiou, C., Drekolia, M. K., Avramidis, N., Contopoulos-Ioannidis, D. G., Drosatos, K., & Baas, J. (2021). Comprehensive mapping of local and diaspora scientists: A database and analysis of 63951 greek scientists. *Quantitative Science Studies*, 1–28. https://doi.org/10.1162/qss_a_00136

Kashyap, R., Rinderknecht, R. G., Akbaritabar, A., Alburez-Gutierrez, D., Gil-Clavel, S., Grow, A., Kim, J., Leasure, D. R., Lohmann, S., Negraia, D. V., Perrotta, D., Rampazzo, F., Tsai, C.-J., Verhagen, M. D., Zagheni, E., & Zhao, X. (2022). *Digital and Computational Demography* (tech. rep.) [type: article]. SocArXiv. https://doi.org/10.31235/osf.io/7bvpt

Kompetenzzentrum bibliometrie (competence centre for bibliometrics) [Accessed: 2021-09-24]. (2021).

Lewison, G., Kumar, S., Wong, C.-Y., Roe, P., & Webber, R. (2016). The contribution of ethnic groups to malaysian scientific output, 1982–2014, and the effects of the new economic policy. *Scientometrics*, *109*(3), 1877–1893. https://doi.org/10.1007/s11192-016-2139-3

Miranda-González, A., Aref, S., Theile, T., & Zagheni, E. (2020). Scholarly migration within Mexico: Analyzing internal migration among researchers using Scopus longitudinal bibliometric data. *EPJ Data Science*, *9*(1), 34. https://doi.org/10.1140/epjds/s13688-020-00252-9

Moed, H. F., Aisati, M., & Plume, A. (2013). Studying scientific migration in scopus. *Scientometrics*, *94*(3), 929–942.

Moed, H. F., & Halevi, G. (2014). A bibliometric approach to tracking international scientific migration. *Scientometrics*, *101*(3), 1987–2001.

Sanliturk, E., Zagheni, E., Dańko, M. J., Theile, T., & Akbaritabar, A. (2023). Global patterns of migration of scholars with economic development [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *120*(4), e2217937120. https://doi.org/10.1073/pnas.2217937120

Şanlıtürk, A. E., Aref, S., Zagheni, E., & Billari, F. C. (2022). *Homecoming after Brexit: Evidence on academic migration from bibliometric data* (tech. rep. WP-2022-019) [Edition: 0]. Max Planck Institute for Demographic Research. Rostock. https://doi.org/10.4054/MPIDR-WP-2022-019

Subbotin, A., & Aref, S. (2021). Brain drain and brain gain in Russia: Analyzing international migration of researchers by discipline using Scopus bibliometric data 1996–2020. *Scientometrics*. https://doi.org/10.1007/s11192-021-04091-x

Vrandečić, D. (2012). Wikidata: A new platform for collaborative data collection. *Proceedings of the 21st international conference on world wide web*, 1063–1064.

Zhao, X., Akbaritabar, A., Kashyap, R., & Zagheni, E. (2023). A gender perspective on the global migration of scholars. *Proceedings of the National Academy of Sciences*, *120*(10), e2214664120. https://doi.org/10.1073/pnas.2214664120

Zhao, X., Aref, S., Zagheni, E., & Stecklov, G. (2022). Return migration of german-affiliated researchers: Analyzing departure and return by gender, cohort, and discipline using scopus bibliometric data 1996–2020. *Scientometrics*. https://doi.org/10.1007/s11192-022-04351-4