The indirect estimation of international migration flows by age and gender

Raphael J. Nawrotzki[abc]
&
Leiwen Jiang [b]

Manuscript prepared for consideration to be presented at the 2014 Annual Meeting of the Population Association of America (PAA), Boston, Massachusetts, May 1-3

Draft – Do not cite

[a] Corresponding author: Raphael J. Nawrotzki, National Center of Atmospheric Research (NCAR), CGD/IAM, Mesa Lab, 3090 Center Green Drive, Boulder, CO 80301, E-mail: nawrotzk@ucar.edu; Phone: 303-819-6725
[b] National Center of Atmospheric Research (NCAR), Climate and Global Dimensions Division (CGD), Integrated Assessment Modeling Group (IAM).
[c] University of Colorado at Boulder, Institute of Behavioral Sciences (IBS), CU Population Center.

Acknowledgments:
We gratefully acknowledge that this project received funding from the U.S. Department of Energy (DoE) grant No. 106952. We thank Ben Sanderson for his help with the data extraction.

The indirect estimation of international migration flows by age and gender

Abstract:
This paper reports on the methods used to generate a novel data set of age and gender profiles for country specific migrant flows with approximately global coverage. We employ raw data from the United Nations Global Migration Database (UNGMD). We selected the best data for two years closest to 2000. The age and gender profiles were computed in three different ways, 1) by using information directly from the selected file; 2) by borrowing information from other files; and 3) by applying gender and age information derived from aggregated region-level streams. Finally, we compute age and gender profiles of migrant flows by employing formal demographic tools to adjust the estimates for mortality and fertility. The generated data set of 3,850 streams will be valuable source to investigate global migration patterns and may be used for population projections.

Keywords:
International migration, Migrant Flows, Age profile, Gender profile, Mortality adjustment

The indirect estimation of international migration flows by age and gender

Table of content:

The indirect estimation of international migration flows by age and gender

I. Introduction

In an era of rapid demographic change and globalization, more people than ever before move across national boundaries with substantial socioeconomic and environmental impacts beyond any single country (UNPD, 2011). International migration plays an increasingly important role in determining the changes in size and structure of national and regional population, especially when large parts of the world have completed the demographic transition (UNPD, 2005). Information on the changes in number of migrants as well as variations in demographic characteristics of the migrant population is needed to improve policies towards better planning of economic development and social services (Raymer et al., 2012). Moreover, to study the impacts of global economic and environmental changes (such as climate change) on future global and regional population distribution, data on international migration that links populations across countries/regions is necessary (Brown, 2008; Piguet et al., 2010). However, to obtain data on migration flows is notoriously difficult due to inconsistent methods of enumeration (e.g., define migrants by country of citizenship or country of birth), and unwillingness of governments to release this data (Bilsborrow et al.,1997; Poulain et al., 2006; Zlotnik et al., 2010; UNECE, 2012).

International organizations and demographers have invested great effort in mining migration data and studying the volume and characteristics of the international migrants, utilizing various indirect estimating methods. The most comprehensive international migration dataset is the UN Global Migration Database (UNGMD) (Zlotnik et al., 2010; UNDESA, 2012), developed by United Nations' Population Division. This dataset is based on empirical records of the number ("stock") of international migrants by age and sex, collected from censuses and other official statistical sources. Using the UNGMD database, the Development Research Group at the World Bank constructs a Global Bilateral Migration data set, matrices of migrant stock between countries for the period of 1960-2000 (Ozden et al., 2011). More recently, Abel (2013) derived the decennial migration flows across 191 countries to reflect the migration intensity of countries during 1960-2000 period, based on the World Bank migration stock tables. Similarly, taking advantage of the better quality of European migration data, the Migration Modeling for Statistical Analysis (MIMOSA) and the Integrated Modeling of Europe Migration (IMEM) projects estimate migration flows across European countries and between the Europe and the rest of the world (de Beermer et al., 2010; Raymer et al., 2011; Raymer et al., 2012). The derived matrix of migration flows provide estimates of total number of international migrants moving across countries during certain period of time. However, this data set does not include information on demographic characteristics such as age and gender profiles of the migrants, which is important for understanding the cause of international migration and their impacts on global and national socioeconomic development and future demographic changes.

To study the age and gender profiles of migrants, Rogers and co-authors developed the Model Migration Schedules (Rogers and Castro, 1981; Rogers et al., 1978, 2005) and used this method to obtain age-gender profiles of internal migration flows for the U.S. within several other countries (Rogers et al., 2002, 2007; Rogers and Raymer 1998, 2007; Rogers and Taylor-Wilson 1996). The dataset constructed under the IMEM project also include information on age and gender of international migrants, but only for countries in the EU and the EFTA. Absent from the available data are age and gender profiles of international migrant flows for the most of the other countries/regions of the world.

The current study, as part of the National Center of Atmospheric Research (NCAR) Community Demographic Model project, is an attempt to fill this void. This paper reports on the methods and techniques used to generate a dataset of age and gender specific international migrant flows around the year 2000. We discuss our preliminary results in the context of recent trends in international migrations and the characteristics of the migration across countries and regions.

II. Concepts and Data

Definitions: In this paper, we use the terms "migration stream" or "stream" to define the unidirectional population migration between two countries/regions, moving from origin to destination. For instance, the migration stream of Mexico – U.S. refers to the migrants moving from Mexico to the US. The size of a migration stream may be measured as the "stock" of migrants living in the country of destination at a particular point in time. In contrast, the "migration flow" adds a temporal dimension and measures the migration from for a country of origin to a country of destination for one unit of time (e.g., per year) (Perruchoud and Redpath-Cross, 2011).

Data: The raw data used in this study comes from the United Nations Global Migration Database (UNGMD). It records the migrant stock for a particular migration stream (e.g., Mexico – U.S.) at different points in time. The complete data set contains 57,047 files covering 7,530 unique migration streams (about 8 files per stream). Migration streams operate at the national level (n=6,432) and regional level (n=1,098) and are reported for years between 1975 and 2009. The migrant stock information is available in a separate .txt file for each year. It typically contains information on the country of origin, the country of destination, the method of enumeration (country of birth vs. country of citizenship), the data source (e.g., demographic yearbook, census, register, survey, etc.), and year of enumeration. In addition, the migration stock data are disaggregated by age and gender groups (see Figure 1 as an example for the stream Mexico – U.S. for the year 2000).

(Figure 1 about here)

However, the information contained in each file varies substantially between migration streams, and even for different years of the same migration stream. For instance, the number of migrants in some files is disaggregated by both age and gender (such as shown in Figure 1), while others are disaggregated only by either age or gender. In the worst case, only total number of migrants is reported. When age information is included, it varies frequently in format. Age information may be recorded evenly or unevenly in 1-year, 5-year, 10-year, 15-year, or even 20-year groups; in some cases, the recording follows the grouping method commonly used in demographic analysis (e.g., 0-4, 5-9, 10-14), while other files use unconventional groupings (e.g., 0-5, 5-10, 10-15); some files have summary groups included between regular age groups (e.g., 0-4, 5-9, 10-14, 15-19, 0-17, 20-24, or 0-4, 5-9, 10-14, 15-19, 18+, 20-24), but may miss one or more age category in between and may or may not have an open-ended category (e.g., 65+). In addition, the values between 1 and 9 migrants are replaced by an asterisk, for confidentiality reasons (UNDESA, 2008). For the same reason, countries of origin with fewer than 100 international migrants are not shown separately. As such, origins are sometimes recorded at regional level (e.g., Southern Europe) or group of countries (e.g., Greece and Italy). In order to overcome the difficulties caused by the irregular data format in the original files and to generate a unified dataset, we performed several steps of data cleaning and processing described in the next section.

III. Method

We use the programming language and statistical environment of the R project (R Core Team, 2013) and take five steps to accomplish the data cleaning and processing:

1. Assess overall data condition by generating an overview table from all raw data files;
2. For each migration stream, select a file closest to year 2000 with most detailed age and gender information for migrant stock (Year 1), and select a second file close to but different from the first year file (Year 2)
3. Derive age and gender profiles of the migrant stock for both Year 1 and Year 2;
4. Generate age and gender profiles of migration flow by calculating the differences of migrants by age and gender between Year 1 and Year 2, after adjusting for mortality and fertility;
5. Derive age and gender profiles of migration flow for 31 global regions by aggregating national level migration streams within each region.

Step 1: Assess overall data condition

From all 57,046 text files, we extract relevant information, including the country of origin and destination, the year of enumeration, the data source (e.g., census, register), the criterion of enumeration (“Country of birth” or “country of citizen”), the gender information availability (yes=1, no=0), the age information availability (yes=1, no=0), and the number of age categories.

Based on the country/region code from the UN Statistical Division (UNSD, 2013), we identify unique combinations of country of origin and destination. When the origin was recorded in the files as a group of countries, we assigned it to the next higher regional level. For example, if both Germany and France are recorded as origin in a given file, we assigned Western Europe as the origin (e.g., UN code “155”).

An overview of the data condition is report in Table 1. It shows that the years for which the migration data was recorded ranged from 1975 to 2009. Gender differentiated migrant counts were available for 84% of the files while disaggregation by age was less common with 52% of the files reporting five or more age categories. The reported number of age groups varies widely (from 1 to 102 age groups), with an average of 10 age groups. For the criterion of enumeration, slightly more than half of all files (54%) used “Country of birth” instead of “Country of citizenship” to classify individuals as migrants.

(Table 1 about here)

Step 2: Select the best file for Year 1 and Year 2

We use a decision-tree structure and adopt five criteria to select for each migration stream the best files with most detailed information at two time points (Year 1 and Year 2). For Year 1, the five criteria are listed here according to importance:

1. Temporal distance from the year 2000
2. Information content
3. Criterion for defining migrants
4. Quality of the age grouping
5. Random number

To select a file closest to the year 2000, we generate a difference measure by subtracting the year of enumeration from the year 2000, which allowed us to use the absolute (sign removed) value as selection criterion. To assess the quality of information content, we generate a categorical variable indicating whether a file has both age and gender information (category 1), has gender information but no age (category 2), has age information but no gender (category 3), or has information on total number of migrants (category 4). The file with more detailed information is preferred. In addition, we prioritize “country of birth” (coded 0) over “country of citizenship” (coded 1). We also count the number of age categories used in the files, and choose the file with the largest number of age categories. We then use a nested sorting algorithm to choose the “best” file according to this set of selection criteria. Figure 2 visually displays the priority sorting mechanism, for choosing the file for the migration stream of Armenia – U.S. as an example. After applying the four selection criteria, the best files are placed on top. When the algorithm is unable to produce a unique result, a file is chosen based on a randomly generated number.

(Figure 2 about here)

Step 3: Obtain age and gender profiles for Year 1 and Year 2

Due to the large variation in information content in the raw data, we follow a rather complicated process to derive age and gender profiles with standardized data format for all migration streams.[1] We first assign each file one of four treatment categories according to information content:

Treatment category 1: Files containing both age and gender information;
Treatment category 2: Files with only gender but no age information;
Treatment category 3: Files with only age but no gender information;
Treatment category 4: Files with only total number of migrants but no age or gender information.

For migration streams with data files in treatment category 2, 3, 4 (incomplete age and gender information), there may be additional information available for years other than Year 1 and Year 2. We try to fully use this additional information, by searching through all available files within each migration streams and flagging the files containing complementary information. Using the decision tree structure described above, we select the files with best information for further processing steps. Table 2 shows exemplary for year 1 the counts of migration streams for which supplementary information was derived from another year, listed as sub-category a, b, and c. For example, it is possible to obtain age information from different years for 557 (of 1997) migration streams that have only gender but no age information in the selected file. For all "b" and "c" categories in Table 2, no complementary "real" data could be obtained from within the stream and age and gender profiles were derived from regional level data.

(Table 2 about here)

After each file had been classified, we derive age and gender profiles for treatment category 1 files. Although treatment category 1 files have complete age and gender information, it is still necessary to standardize the format of age category. We standardize all age profiles adopting 16 five-year age categories.[2]

For files with missing information for certain age group (e.g. 5-9), which contain information on a summary age category (e.g., 0-17), we use simple algebra to compute the missing information. In some age groups migrant counts between 1 and 9 are replaced by an asterisk for confidentiality reasons. Whenever enough information is available (e.g., male="*", female=11, total="16"), we calculate the migrant count (e.g., male=16-11=5). In the case when insufficient information for calculating the migrant count is available, we assign both male and female columns a default value of 4.5 (median of possible values 1 to 9) and the total column the sum of those values (e.g., male=4.5, female=4.5, total=9.0). For age groups with migrant counts on total column but no information on male and female columns, we allocate the total migrant counts evenly to males and females.

---

[1]We wrote a set of 21 functions for the processing of the data and compiled these in an R source package. The scripts, functions, and package are available upon request from the corresponding author.
[2] The sixteen age categories are: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+.

We then we spread the migrant counts to single year group and collapse the data to the desired 16 five-year age groups. In this way we are able to unify the age categories. An example of a successfully processed and cleaned file is presented in Table 3.

(Table 3 about here)

In a consecutive step, we derive age and gender profiles for migration streams with data files of treatment categories 2a, 3a, and 4a. These migration streams in the "a" categories have data files missing age or gender information, but complementary information can be obtained from data files from a different year. Figure 3 demonstrate the process flow for a hypothetical category 2a case, for which only gender information is available in the selected file, but for which age information can be obtained from an earlier year.

(Figure 3 about here)

We then use regional level information to derive age and gender profiles for the migration streams with data files of treatment category 2b, 3b, 4b, and 4c. These "b" and "c" category files neither have gender/age information available in the selected file nor can complementary information be found at any other year. We construct age and gender profiles at the next higher geographic level (e.g., region, continent) based on the available national level migration information within corresponding regions or continents and apply this "artificial" age and gender profile to the stream lacking this information.

For example, if the age profile for the migration stream from Honduras to the U.S. is missing, we construct the age profiles of migration streams of Central America to the US, using the available information of all migration streams from Central American countries (e.g., El Salvador, Nicaragua, Belize, Guatemala, etc.) to the U.S. and apply the derived age profile at regional level to the migration stream at the country level (e.g., Honduras to the U.S.). To generate the upper-level profiles, we specify a hierarchical structure of countries nested within regions using the nesting structure suggested by the UN (UNSD, 2013). The nesting structure has five levels including country, sub-continent, continent supra-continent, and global (see Figure 4). We use data files of lower-level migration streams to generate the age and gender profiles of the upper-level migration streams.

(Figure 4 about here)

The generated upper level profiles are then applied to files of "b" and "c" treatment categories for gender and/or age profiles could not be derived from within the stream. The employed algorithm first obtains the unique stream ID (e.g., 840-484 for migration from Mexico to the U.S.) of the "b" and "c" category file with the missing profile information. The algorithm then tests at each level whether the required profile can be obtained. Figure 5 visually displays the process of profile selection.

(Figure 5 about here)

In finding data files with complementary information for Year 2 for all “a” or “b” treatment categories, we try to use a different file than the one used for Year 1. Only in the case when no other information is available, was the same file employed and the profiles for Year 1 and 2 were derived from identical gender and age information.

Step 4: Generate age and gender profiles of migration flow through calculating the differences of migrants by age and gender between Year 1 and Year 2, after adjusting for mortality and fertility

We compute migrant flows by subtracting the count of migrants in the earlier year (e.g., Year 1) from the count of migrants in the later year (e.g., Year 2) and divide the quantity by the period length to derive the average annual changes. However, the differences in the migrant stocks between the two time points can be, in parts, attributed to deaths and births of the migrants. To obtain an accurate estimation of net migration, it is necessary to account for the impacts of mortality and fertility of the migrants, as formally described in the following.

Equation 1:

$$M_t^x = M_{t-n}^{x-n} + (I^{x-n} - O^{x-n}) - \left(M_{t-n}^{x-n} - \frac{O^{x-n}}{2}\right) * (1 - S^x) - (I^{x-n} - O^{x-n}) * (1 - S^x)$$

In equation 1 M_t^x and M_{t-n}^{x-n} are the number of migrants of age x at time t and of age x-n at time t-n, respectively. S^x is the conditional survival probability of population from age x-n to x, I^{x-n} and O^{x-n} are the immigrants and emigrants of age x-n to x during the period t-n to t. Therefore, $\left(M_{t-n}^{x-n} - \frac{O^{x-n}}{2}\right) * (1 - S^x)$ is the number of deaths of migrants in the country of destination, and $(I^{x-n} - O^{x-n}) * (1 - S^x)$ is the number of deaths of net immigrants moving in during the period of t-n to t. Here, we assume the age-specific mortality rates are the same for immigrants and emigrants.

As we do not know the number of emigrants from the data files, we use $(M_{t-n}^{x-n}) * (1 - S^x)$ to replace $\left(M_{t-n}^{x-n} - \frac{O^{x-n}}{2}\right) * (1 - S^x)$ for the number of deaths of non-moving migrants, and trust that the difference is relatively small.

Denoting the number of net migrants of age x-n to x during the period t-n to t $as\ N^x$ ($N^x = I^{x-n} - O^{x-n}$), equation 1 simplifies to equation 2:

Equation 2:
$$M_t^x = M_{t-n}^{x-n} + N^x - M_{t-n}^{x-n} * (1 - S^x) - N^x * (1 - S^x)$$
$$M_t^x = M_{t-n}^{x-n} * S^x + N^x * S^x$$
$$N^x = M_t^x / S^x - M_{t-n}^{x-n}$$

When migrants are defined by "country of birth," the number of net migrants of age 0 to n is calculated using equation 3, where N^{-n} denotes the net migrants of age 0 to n, and M_t^{-n} represents the number migrants of age 0 to n at time t.

Equation 3:
$$N^{-n} = M_t^{-n} + M_t^{-n} * (1 - S^n)/2$$

On the other hand, when migrants are defined by "country of citizen," female migrants of reproductive age give birth to children in the country of destination, who are counted as migrants although they never moved. Therefore, we have to exclude the number of births given by the female migrants from the number of migrants of age 0 to n. This can be formally described as follows (Equation 4).

Equation 4:
$$N^{-n} = (M_t^{-n} + M_t^{-n} * \frac{1 - S^n}{2}) - \sum_{i=15}^{49} (M_{t-n,f}^i * b^i + M_{t-n,f}^i * S^i * b^{i+n}) * n/2$$

In equation 4, the parameter b^i represents the fertility of women age i. Assuming a sex ratio of birth equal to one, we allocate the total new births evenly to the male and female groups.[3]

Step 5: Derive age and gender profiles of migration flows for 31 global regions

In order to be able to use the generated migrant flow measure for multiregional population projections of NCAR's Community Demographic Model, we aggregate the country level flows to 31 global regions.[4] For the aggregation we used only those migration flows that operate at the country level. In addition, we included only those county level flows that had no missing values on all age and gender categories to ensure accurate representation of the migration flows. We aggregated both the total flow measures and the gender and age specific flow to the 31 global regions.

[3] The age, gender, and country specific survival rates of year 2000 have been provided by the UN Population Division (UNPD, 2009b). Because the survival rate data was provided in five-year intervals, we used a linear interpolation to obtain one-year survival rates. The age and gender specific survival rates for the country of destination are used, assuming that migrant's mortality rates adjust quickly to the average of the host country. The age-specific fertility of year 2000 used here came also from the UN Population Division (UNPD, 2009a). We assume a constant birth rate for female migrants across the period between Year 1 and Year 2.

[4] Appendix Table 1 lists the 31 global regions and the respective countries included within these regions.

V. Preliminary Results and Discussion

(Note that the below paragraphs constitute preliminary results. We are currently testing the robustness of our data set and are working on some further analyses and are in the process of improving the interpretation of the results in light of country specific contexts)

Using the above described methodology, we were able to generate a data set of gender and age specific migrant flows at the country and region level. Migration flows were constructed at the national level for 3,850 migration streams for which data files of two time points are available. The numbers of migration streams for various countries differ substantially due to different migration intensity and data availability. For instance, there are 190 migration streams derived for the U.S. as destination, while there is no data for China as migration destination. Adding net migration flows of all countries within the 31 global regions, we obtain the figures at regional level. Of the possible 961 regional migration flows with unique origin and destination combinations (31 x 31), we are able to obtain age and gender profiles of net migrants for 476 (50%) of the streams.

The regional net migration counts allow for the identification of major international migration senders and receivers. A map of net international migration flows (Figure 6) illustrates that the U.S.A., Canada (orange lines) and Australia (red lines) are the major receivers. Other international migration receivers include Western Europe, the Gulf oil rich countries in Western Asia, South Africa in the Africa continent, and Japan for South, Southeast, and East Asian emigrants. The developing regions including Latin America and the Caribbean, South and Southeast Asia, and Africa are the main senders of international migrants. However, it is noteworthy that a large number of net immigrants to Australia and the rest of Oceania originate from Northern Europe Western Europe.

(Figure 6 about here)

In contrast, one of the largest negative net migration flows originated in Russia (yellow green lines), mainly because of the returning migrants of Eastern European countries during the period close to 2000. The results are generally in line with the findings of other researchers and organizations (e.g. see UNECE, 2012, chart 2).

Informed by the global representation of lows, we examine the demographic characteristics of migrants for a selection of country level flows. The characteristics of these flows are determined by numerous influences including historical background, political situation, socioeconomic conditions, etc. (Brown and Bean, 2006). We focus our analysis on five distinct migration flows: Philippies to Japan, Italy to the U.S.,

(Figure 7 about here)

The age and gender compositions of the international immigrants in the stock registered at the time of censuses or surveys can substantially differ from that of the net migration flow. For instance, while the age and gender profiles of immigrants from Philippines to Japan recorded in year 2000 and 2001 are quite similar to a typical migration model schedule (Rogers et al., 2005), the age and gender profiles of the net migration flows between the two years is quite different (Figure 7, Panel A). Similarly, the age and gender profiles of immigrants from Italy to the U.S. (Panel B) in year 2000 and 2002 shows that a large proportion of the immigrants are above age 55. However, the age and gender profiles of net migration flows between 2000 and 2002 reveal that net immigrants are largely young individuals under age 30, while net emigration among older individuals characterizes the flow. This is because most of the Italian immigrants came to the US decades ago and aged over the years, while the age and gender profiles of the net migration flows reflect the demographic characteristics of the more recent immigrants from Italy. Therefore, one cannot simply apply the age and gender composition of migrant stock reported in a census to indicating the profiles of the recent net migration flows.

Moreover, because of the cumulative nature of migrant stock, the age and gender composition of migrants reported in different censuses or surveys may hardly change in a short period of time. For instance, the Mexican migrants to the U.S. (Panel C) show a very similar pattern of age and gender in both 2000 and 2002. However, our analysis of the age and gender composition of the migration flow demonstrates that the recent Mexican immigrants have a quite different age-gender profiles than the stock would suggest. While in both the migrant stocks and migration flow, most immigrants are working age adults (aged 20-50), there are more children but less teenagers in the recent migration flows. Interestingly, there are more migrating children aged 5-10 than those aged 0-4 in the recent flows, indicating that the Mexican immigrants more likely bring school age children than bring the youngest ones. If the age and gender profiles of current migration flow persist, it will gradually change the age and gender compositions of the migrant stock in the future. This effect is obvious in the migration stream Mozambique – South Africa (Panel D): Because of the negative net migration for age 45-64 of male immigrants during the period 1996-2001, the age profiles of male migrant stock in 2001 significantly differ from the one in 1996.

The age-gender profiles of migration flows are affected by the differences in socioeconomic conditions and the changes between the origins and destinations. Changes in immigration policies in the migrants receiving countries also have an important impact on the age-gender profiles of international migrants. The results of the migration streams Kyrgyzstan to Ukraine (Panel E) between the year 1989 and 2001 clearly demonstrated that the dissolution of the former Soviet Union induces large flow of returning migrants of Kyrgyzstan ethnicity and caused considerable negative net migration in the Ukraine. While the negative migration flow appears for all age and gender groups, the returning Kyrgyzstan migrants is less likely for school age males (aged 10-25) and the elderly (except the very old aged 75 and above). The analysis of age-gender profiles of migration flow of the Philippines to Japan (Panel A) offers another example.

The migrant stocks of 2000 and 2001 shows that female Filipinos immigrants (mostly employed in domestic work and entertainment) in Japan are much more prevalent than male migrants. While the feature of unbalanced gender composition is also reflected in the profile of migration flows between 2000 and 2001, we see an interesting two-peak shaped curve in the age profiles of female migration flow. The peak in the later age group (35-45) is perhaps the result of a large number of female Filipinos filling employment positions as domestic workers. However, the peak in the earlier age group (15-29) may reflect a recent change in the employment opportunities of female migrants. The Japanese government issued large number of visas to young Filipinos to work as entertainers in Japan (Anderson 2005).

This brief discussion demonstrates that the age-gender profiles of migration flows can be much more volatile than that of migrant stock data. The net changes of age and gender compositions of the immigrants and emigrants are impacted by a wide range of temporal push and pull factors operating in the origin and destination countries. A careful study on the cause of changes in the age-gender profiles of migration flows is important and helpful for projecting the changes in the profiles and their impacts on future demographic dynamics in the origin and destination countries under different socioeconomic scenarios. Moreover, to evaluate and interpret the derived profiles, it is important to recognize that profiles differ in quality dependent on the available information of the source file. In particular, the reliability of the profiles is dependent on the treatment category. We construct a number of indicators for the evaluation of the quality of the resulting stream measure which is listed in Appendix Table 2.

Finally, based on the available country level data, we generated region-level age-gender profiles as additive flow measures. Figure 8 shows the profiles of migrant streams from selected regions to the U.S.A.

(Figure 8 about here)

Figure 8 shows some substantial variation in the profiles of migration from various regions to the U.S. While for Western Europe the number of older migrants declined, particularly for females, the number of older migrants increased for the flows originating in the Caribbean. Only limited migration occurred from the Rest of Eastern Europe to the U.S. with no clear age pattern. For the migration from China, India, and Northern Africa to the U.S., an increase of young migrants was distinguishable. However, it needs to be recognized that the representativeness of region level streams is dependent on the number and quality of country-level migration streams contributing to the observed numbers. To gain more complete pictures of the age-gender profiles of regional migration flows, we will compare the total number of regional net migration flows derived from our results to other existing tables, such as the global migration flow tables estimated by Abel (2013), since it is based on data of total number of migrants only which is more accessible and complete than data with age and gender information.

Despite some limitations, this novel data set may proof beneficial for demographers and organizations interested in projecting changes in population. In addition our data set is uniquely positioned to answer research questions such as the following: Are there distinct patterns of changes in sex and age distribution for international migration? Which countries have witnessed the most drastic increase in vulnerable groups such as migrant children and women? Answers to these questions are of high policy relevance in times were the forces of globalization and climate change lead to increased international mobility.

VI. References

Abel, G. J. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research, 28,* 505-546. doi: 10.4054/DemRes.2013.28.18

Adamo, S. B., & de Sherbinin, A. (2011). The impact of climate change on the spatial distribution of populations and migration *Population distribution, urbanization, internal migration and development: An international perspective* (pp. 161-195). New York: Unite Nations, Department of Economic and Social Affairs, Population Division.

Anderson, J. (2005). Filipinos in Japan. In M. Ember, C. R. Ember & I. Skoggard (Eds.), *Encyclopedia of Diasporas*. New York: Springer.

Bilsborrow, R. E., Hugo, G., Oberai, A., & Zlotnik, H. (1997). International migration statistics: Guidelines for improving data collection systems. Geneva, Switzerland: International Labour Office.

Brown, O. (2008). Migration and climate change. Geneva, Switzerland: International Organization for Migration.

Brown, S. K., & Bean, F. D. (2006). International Migration. In D. Posten & M. Micklin (Eds.), *Handbook of population* (pp. 347-382). New York: Springer Publishers.

de Beer, J., Raymer, J., van der Erf, R., & van Wissen, L. (2010). Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe. *European Journal of Population-Revue Europeenne De Demographie, 26*(4), 459-481. doi: 10.1007/s10680-010-9220-z

Hugo, G. (2011). Future demographic change and its interactions with migration and climate change.*Global Environmental Change-Human and Policy Dimensions, 21,* S21-S33. doi: 10.1016/j.gloenvcha.2011.09.008

IOM. (2011). *Migration profiles: Making the most of the process*. Geneva, Switzerland: International Organization for Migration.

OECD. (2012). *International migration outlook 2012*. Danvers, MA: OECD Publishing.

Ozden, C., Parsons, C. R., Schiff, M., & Walmsley, T. L. (2011). Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000. *World Bank Economic Review, 25*(1), 12-56. doi: 10.1093/wber/lhr024

Perruchoud, R., & Redpath-Cross, J. (2011). *Glossary on migration* (2nd ed.). Geneva, Switzerland: International Organization for Migration.

Piguet, E. (2010). Linking climate change, environmental degradation, and migration: a methodological overview. *Climate change, 1*(4), 517-524.

Poulain, M., Perrin, N., & Singleton, A. (2006). *HESIM: Towards harmonized european statistics on international migration*. Louvain-la-Neuve, Belgium: UCL Presses Universitaires de Louvain.

Raymer, J., de Beer, J., & van der Erf, R. (2011). Putting the pieces of the puzzle together: Age and sex-specific estimates of migration between EU/EFTA countries, 2002-2007. *European Journal of Population, 27,* 185-215.

Raymer, J., Foster, J. J., Smith, P. W. F., Bijak, J., & Wisniowski, A. (2012). Integrated modelling of European migration: Background, specification and results (pp. 1-17). London, U.K.: NORFACE Research Programme on Migration.

Raymer, J., & Rogers, A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography, 44*(2), 199-223. doi: 10.1353/dem.2007.0016

RCoreTeam. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org/.

Rogers, A., & Castro, L. (1981). Model migration schedule (pp. 1-153). Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rogers, A., Castro, L., & Lea, M. (2005). Model migration schedules: Three alternative linear parameter estimation methods. *Mathematical Population Studies, 12*(1), 17-38.

Rogers, A., Jones, B., Partida, V., & Muhidin, S. (2007). Inferring migration flows from the migration propensities of infants: Mexico and Indonesia. *Annals of Regional Science, 41*(2), 443-465. doi: 10.1007/s00168-006-0107-1

Rogers, A., Raquillet, R., & Castro, L. J. (1978). Model migration schedules and their applications.*Environment and Planning A, 10*(5), 475-502. doi: 10.1068/a100475

Rogers, A., & Raymer, J. (1998). The spatial focus of U.S. interstate migration flows. *International Journal of Population Geography, 4*, 63-80.

Rogers, A., Raymer, J., & Willekens, F. (2002). Capturing the age and spatial structures of migration.*Environment and Planning A, 34*(2), 341-359. doi: 10.1068/a33226

Rogers, A., & Wilson, R. T. (1996). Representing structural change in US migration patterns. *Geographical Analysis, 28*(1), 1-17.

Thomas-Hope, E., Knight, P., & Noel, C. (2010). Migration in Jamaica: A country profile 2010. Geneva, Switzerland: International Organization for Migration (IOM).

UNDESA. (2008). United Nations Global Migration Database (UNGMD). New York: United Nations, Department of Economic and Social Affairs, Population Division.

UNDESA. (2012). The age and sex of migrants 2011. New York: United Nations Department of Economic and Social Affairs.

UNECE. (2012). *Review of sources and quality of statistics on international migration in selected countries of the Commonwealth of Independent States*. New York: United Nations Economic Commission for Europe.

UNPD. (2005). World population prospects 2004 revision. New York: United Nations Department of Economic and Social Affairs, Population Division.

UNPD. (2009a). Age-specific fertility rates by major area, region, and country *World Population Prospects: The 2008 Revision*. New York: United Nations, Department of Economic and Social Affairs, Population Division.

UNPD. (2009b). Male life table survivors at exact age, l(x), by major area, region and country *World Population Prospects: The 2008 Revision*. New York: United Nations, Department of Economic and Social Affairs, Population Division.

UNPD. (2011). International migration report 2009: A global assessment. New York: United Nations Department of Economic and Social Affairs, Population Division.

UNSD. (2013). Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. New York: United Nations Statistical Division, retrieved June 18, 2013, from http://unstats.un.org/unsd/methods/m49/m49regin.htm.

Wing, I. S., & Fisher-Vanden, K. (2013). Confronting the challenge of integrated assessment of climate adaptation: a conceptual framework. *Climatic Change*, *117*(3), 497-514. doi: 10.1007/s10584-012-0651-x

Zlotnik, H., Guest, P., Hovy, B., & Henning, S. (2010). Data and analysis: Partnering to better understand and address the human development implications of migration. Geneva, Switzerland: United Nations Department of Economic and Social Affairs.

Table 1: Summary statistics of relevant information across all 57,046 raw data files of the UN Global Migration Data Base

Variable	N	mean	Std.Dev.	min	max
Year of enumeration	57046	1994.92	8.83	1975	2009
Age information available	57046	0.52	0.5	0	1
Number of age categories	57046	10.66	12.84	1	102
Gender information available	57046	0.84	0.37	0	1
Criterion of enumeration	57046	0.54	0.5	0	1

Note: For criterion of enumeration "Country of birth" was coded 1 while "Country of citizenship" was coded 0.

Table 2: Treatment categories used for generating age and gender profiles for "year 1"

Main category	Sub category	Description	Main (N)	Sub (N)
1		**Gender, age, total**	**4,745**	
2		**Gender, total, no age**	**1,997**	
2	a	Age in same stream but different year		557
2	b	No age in any year		1,447
3		**Age, total, no gender**	**97**	
3	a	Gender same stream but different year		61
3	b	No gender in any year		37
4		**Total, no gender or age**	**678**	
4	a	Gender and age in two other years		133
4	b	Gender or age in different year but not both		154
4	c	No gender or age in any year		391

Note: Total number of migration streams 7,530. For sub category "a" true age or gender information could be obtained from another file within the same stream.

Table 3: Example of cleaned migrant stock profile for migrants from Mexico to the U.S. obtained for the year 2000

Age category ID	Age category	Male	Female	Total
1	0-4	89,096	76,393	165,489
2	5-9	158,855	146,906	305,761
3	10-14	231,155	212,822	443,977
4	15-19	415,857	286,392	702,249
5	20-24	716,640	468,546	1,185,186
6	25-29	787,913	581,997	1,369,910
7	30-34	735,407	561,529	1,296,936
8	35-39	598,345	474,767	1,073,112
9	40-44	448,259	371,460	819,719
10	45-49	309,322	267,913	577,235
11	50-54	208,533	193,787	402,320
12	55-59	133,231	137,060	270,291
13	60-64	89,324	101,094	190,418
14	65-69	62,838	76,498	139,336
15	70-74	40,979	53,759	94,738
16	75+	58,725	82,085	140,810
Total		5,084,479	4,093,008	9,177,487

Note: Criteria of enumeration: Country of birth; Source of information: Census; The table was derived as part of treatment category 1 with a gender specific profile

Table 4: Excerpt of level-1 (region) table of summed migrant counts into the U.S.A. for selected regions

ID	UNdest	UNorig	count	female.1	female.2	female.3	female.4
840-143	840	143	2	115	854	894.4	1006
840-145	840	145	14	5512	11363	16641	19334
840-15	840	15	7	1938.1	3370.7	4351.9	4104
840-151	840	151	11	11902	22891	33324	37229
840-154	840	154	9	4325.1	8222.6	9426.2	11532.3
840-155	840	155	8	6418.3	9641.8	9945.8	13765.9
840-17	840	17	1	173	439	670	804

Note: UN IDs for unique combination of country of destination (here U.S.: ID=840) and regions (e.g., 143=Central Asia; 15=Northern Africa; 17=Middle Africa; 155=Western Europe) are displayed. The “count” variable was used as a quality check and indicates the number of country-level stream that contributed to the upper-level stream. For example, seven country-level streams were summed to form the upper-level stream from Northern Africa to the U.S. The displayed data is in wide format and “female.1” represents age category #1 (0-4 years) for females.

Table 5: “Hierarchy” table displaying unique country of destination and origins operating at various levels

UNdest	UNorig	unreg0	unreg1	unreg2	id0	id1	id2
840	484	484	13	419	840-484	840-13	840-419
840	492	492	155	150	840-492	840-155	840-150
840	496	496	30	142	840-496	840-30	840-142
840	498	498	151	150	840-498	840-151	840-150
840	500	500	29	419	840-500	840-29	840-419

Note: UN IDs were employed to obtain a table of unique country of destination (e.g., U.S.A. = 840) and origin (e.g., Mexico=484) streams and all possible upper-level IDs (e.g., Central America to the U.S.A. = 840-13). UNdest = UN ID for destination country; UNorig = UN ID for origin country; unregX = UN region code for respective level X; idX = ID combination of country of destination plus region of origin operating at level X.

Figure 1: Raw data file for the migration stream from Mexico to the U.S. for the year 2000

Country or area of enumeration			United States of America
Country of birth or aggregate			Mexico
Criterion			Country of birth
Year			2000
Source			Census
Footnote			
Age Group	Male	Female	Total
Total	5,084,479	4,093,008	9,177,487
00-00	17,766	12,113	29,879
01-04	71,330	64,280	135,610
05-09	158,855	146,906	305,761
10-14	231,155	212,822	443,977
15-19	415,857	286,392	702,249
20-24	716,640	468,546	1,185,186
25-29	787,913	581,997	1,369,910
30-34	735,407	561,529	1,296,936
35-39	598,345	474,767	1,073,112
40-44	448,259	371,460	819,719
45-49	309,322	267,913	577,235
50-54	208,533	193,787	402,320
55-59	133,231	137,060	270,291
60-64	89,324	101,094	190,418
65-69	62,838	76,498	139,336
70-74	40,979	53,759	94,738
75-79	31,210	37,000	68,210
80-84	14,647	21,862	36,509
85-89	8,007	15,029	23,036
90-94	3,452	5,804	9,256
95-99	1,128	1,883	3,011
100+	281	507	788

Note: File ID: "840_2000_5_Melanesia.txt"

Figure 2: Schematic depiction of employed priority sorting algorithm to select the "best" case among streams with multiple files

File	Stream ID	Year	Quality data	Criteria	Quality age prof	Random
1	840-51	2001	Gender & Age	Cntry of birth	1 year groups	1
2	840-51	2001	Gender & Age	Cntry of birth	1 year groups	5
3	840-51	2001	Gender & Age	Cntry of birth	5 year groups	
4	840-51	2001	Gender & Age	Cntry of birth	10 year groups	
5	840-51	2001	Gender & Age	Cntry of citizenship		
6	840-51	2001	Gender only			
7	840-51	2001	Age only			
8	840-51	2001	Total only			
9	840-51	1998				
10	840-51	1998				
11	840-51	1975				

Note: Hypothetical example for 12 files for the migration stream Armenia to the U.S. File 1 on top was selected for total migrant stock information.

Figure 3: Schematic representation of working steps to derive and apply an age profile for a hypothetical category 2a case that contains only gender differentiated total information

Age profile (year 1992)

Age group	Male	Female	Total
0-4	10	20	30
5-9	40	80	120
10-14	90	120	210
...	...	...	...
Total	800	1000	1800

Age profile transformed to column percent (year 1992)

Age group	Male	Female	Total
0-4	0.013	0.020	0.017
5-9	0.050	0.080	0.067
10-14	0.113	0.120	0.117
...	...	...	...
Total	1.000	1.000	1.000

Gendered total (year 2000) missing age information

Age group	Male	Female	Total
Total	200	300	500

Gendered total (year 2000) with applied age profile

Age group	Male	Female	Total
0-4	2.6	6.0	8.5
5-9	10.0	24.0	33.5
10-14	22.6	36.0	58.5
...	...	...	...
Total	200	300	500

Figure 4: Hierarchy of countries, regions, continents for the example of the Americas as used by the UN (UNSD, 2013) and employed in the generation of upper-level profiles

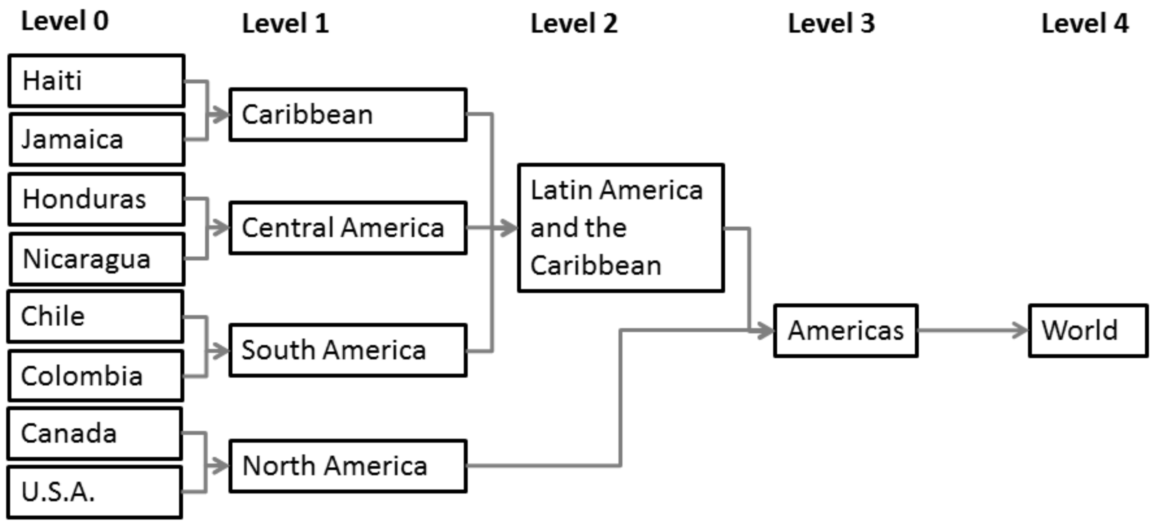


Figure 5: Conceptual visualization of the employed algorithm to select a suitable upper-level profile for streams that miss age and/or gender profiles

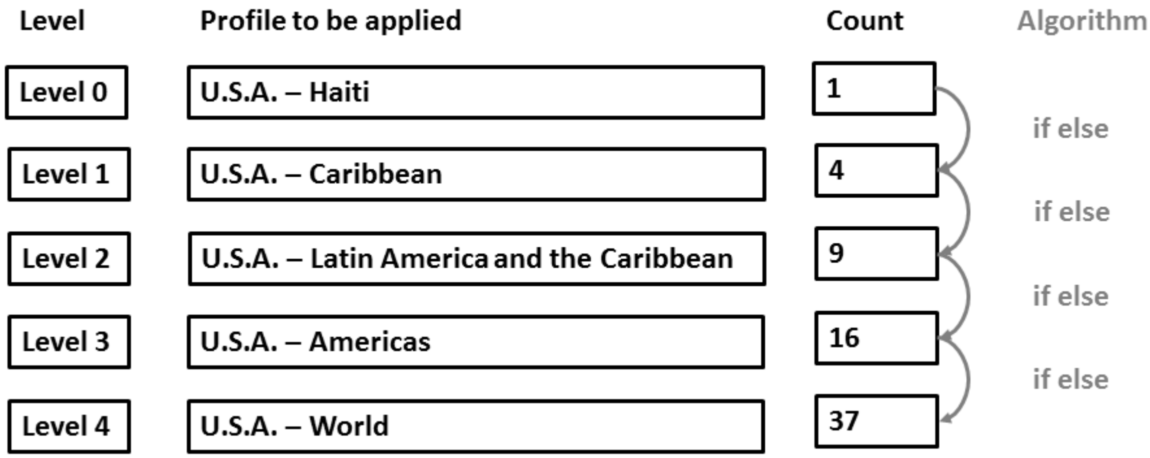
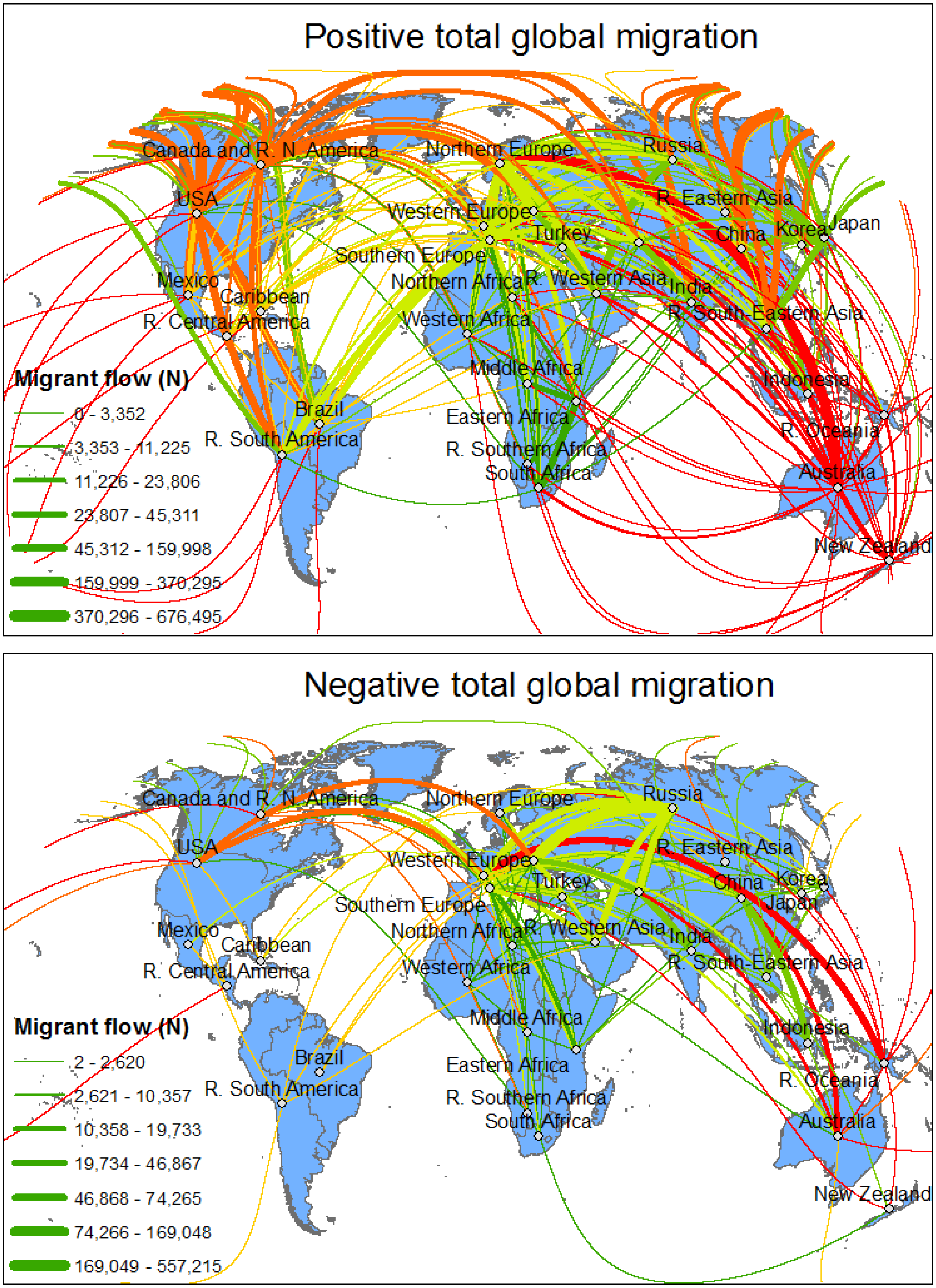


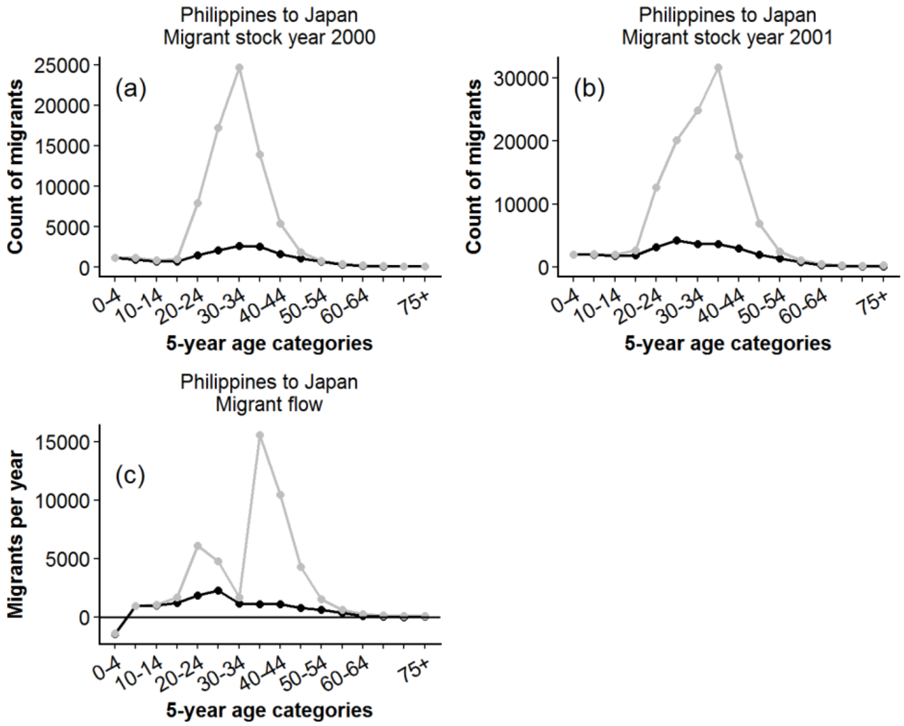
Figure 6: Global representation of positive and negative migration flows between 31 regions.



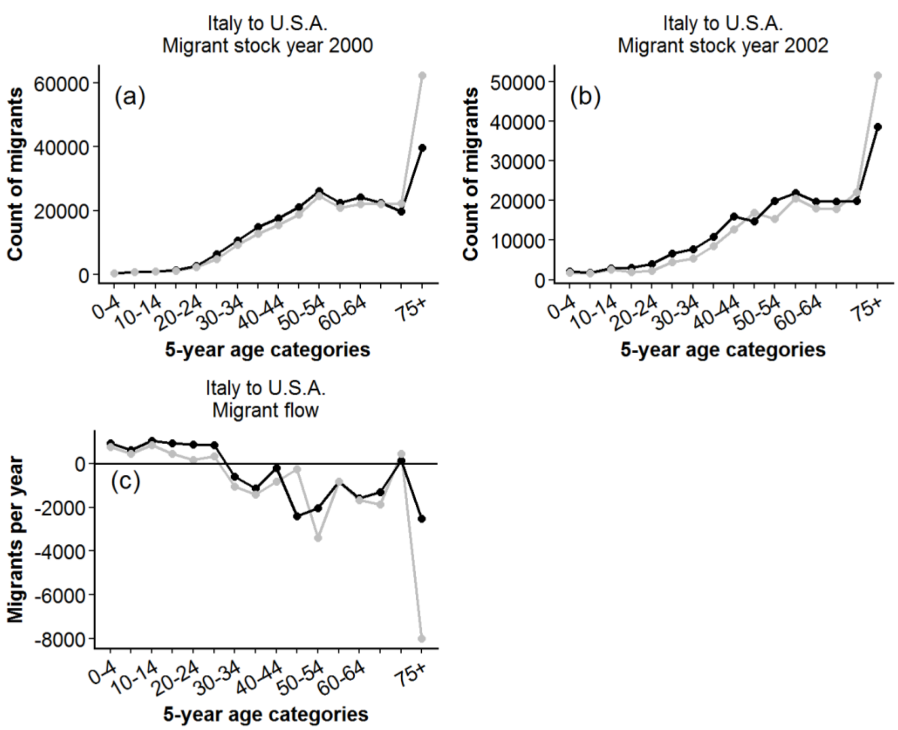
Note: Grouping of migrant flows by destination region is visualized by using the same line color

Figure 7: Migrant stock data for two years as well as resulting flows for five distinct migration streams

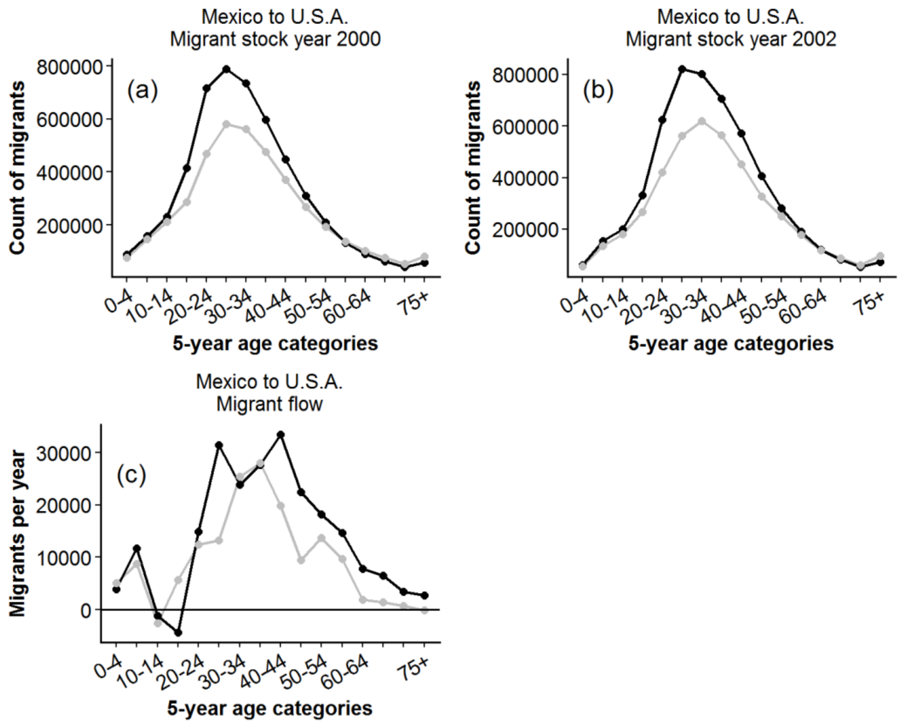
Panel A: Philippines to Japan



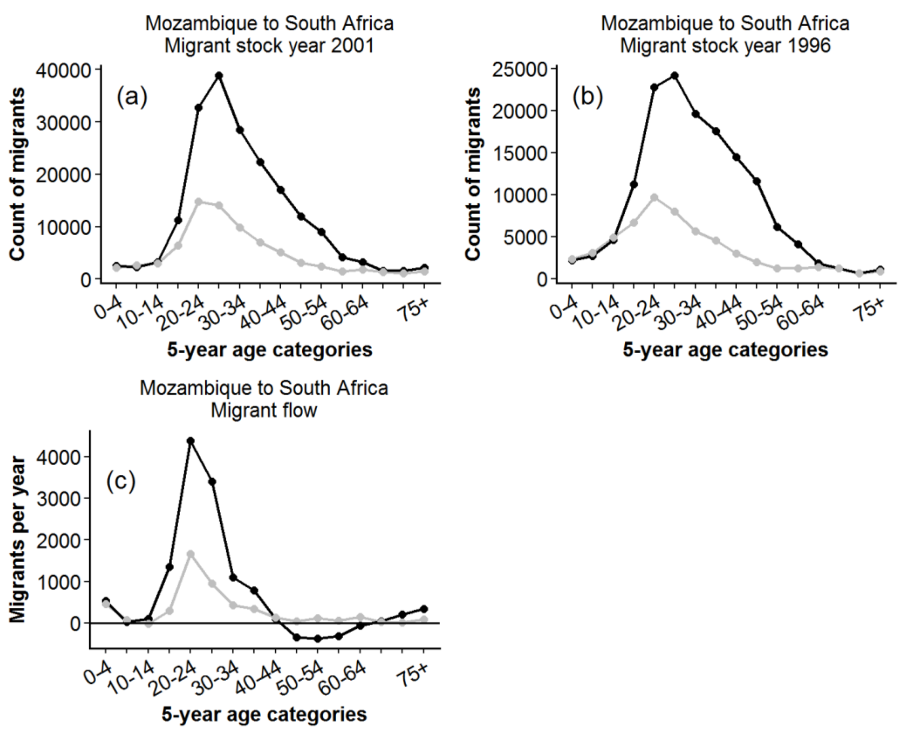
Panel B: Italy to the U.S.A.



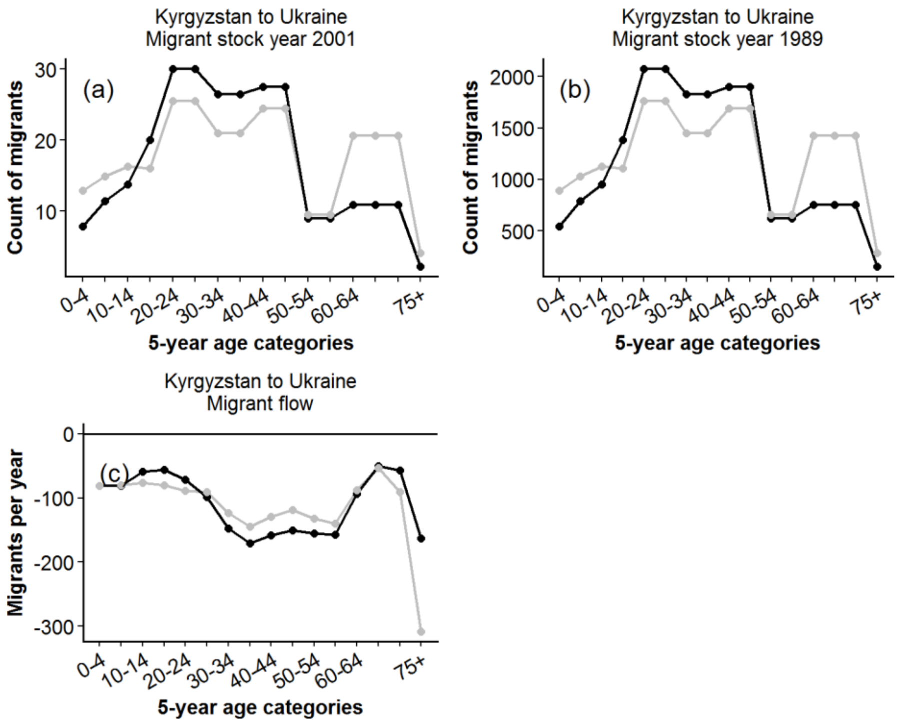
Panel C: Mexico to the U.S.A.



Panel D: Mozambique to South Africa

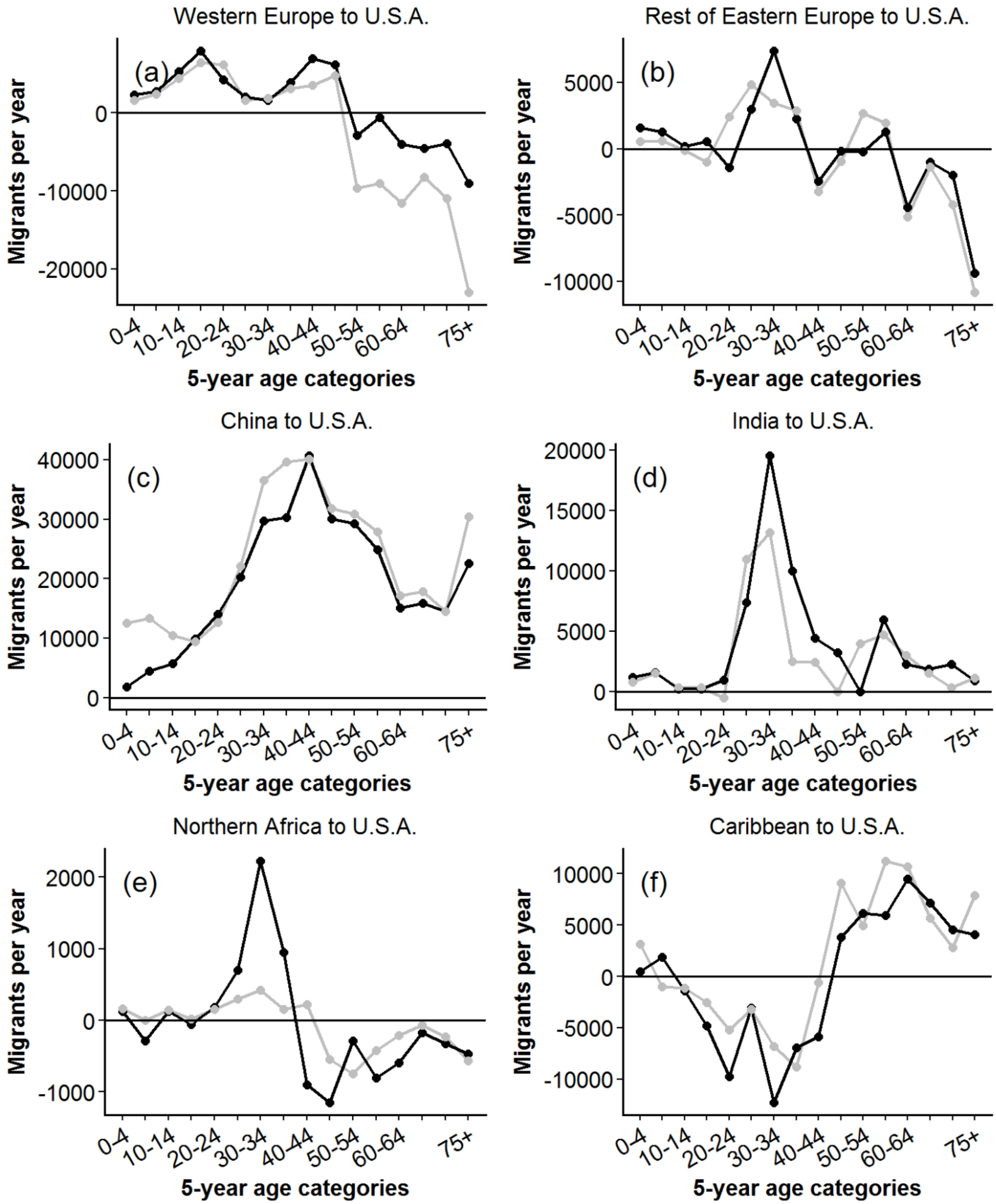


Panel E: Kyrgyzstan to Ukraine



Note: Gray line represents female migrants; Black line represents male migrants.

Figure 8: Gender and age specific profiles of migration from selected IAM regions to the U.S.A.



Note: Gray line represents female migrants; Black line represents male migrants.

Appendices:

Appendix Table 1: 31 global regions and Countries

No.	Region code	Region name	Countries in region
1	910	Eastern Africa	Burundi; Comoros; Djibouti; Eritrea; Ethiopia; Kenya; Madagascar; Malawi; Mauritius; Mayotte; Mozambique; Réunion; Rwanda; Seychelles; Somalia; South Sudan; Uganda; United Republic of Tanzania; Zambia; Zimbabwe
2	911	Middle Africa	Angola; Cameroon; Central African Republic; Chad; Congo; Democratic Republic of the Congo; Equatorial Guinea; Gabon; Sao Tome and Principe
3	912	Northern Africa	Algeria; Egypt; Libya; Morocco; Sudan; Sudan; Tunisia; Western Sahara
4	9130	R. Southern Africa	Botswana; Lesotho; Namibia; Swaziland
5	710	South Africa	South Africa
6	914	Western Africa	Benin; Burkina Faso; Cape Verde; Cote d'Ivoire; Gambia; Ghana; Guinea; Guinea-Bissau; Liberia; Mali; Mauritania; Niger; Nigeria; Saint Helena; Senegal; Sierra Leone; Togo
7	915	Caribbean	Anguilla; Antigua and Barbuda; Aruba; Bahamas; Barbados; Bonaire, Saint Eustatius and Saba; British Virgin Islands; Cayman Islands; Cuba; Curaçao; Dominica; Dominican Republic; Grenada; Guadeloupe; Haiti; Jamaica; Martinique; Montserrat; Puerto Rico; Saint-Barthélemy; Saint Kitts and Nevis; Saint Lucia; Saint Martin (French part); Saint Vincent and the Grenadines; Sint Maarten (Dutch part); Trinidad and Tobago; Turks and Caicos Islands; United States Virgin Islands; Netherlands Antilles
8	9160	R. Central America	Belize; Costa Rica; El Salvador; Guatemala; Honduras; Nicaragua; Panama
9	484	Mexico	Mexico
10	9310	R. South America	Argentina; Bolivia (Plurinational State of); Chile; Colombia; Ecuador; Falkland Islands (Malvinas); French Guiana; Guyana; Paraguay; Peru; Suriname; Uruguay; Venezuela (Bolivarian Republic of)
11	76	Brazil	Brazil
12	9050	Canada and R. N. America	Bermuda; Canada; Greenland; Saint Pierre and Miquelon
13	840	USA	United States of America
14	9210	R. South-	Kazakhstan; Kyrgyzstan; Tajikistan; Turkmenistan;

		Central Asia	Uzbekistan; Afghanistan; Bangladesh; Bhutan; Iran (Islamic Republic of); Maldives; Nepal; Pakistan; Sri Lanka
15	156	China	China; China, Hong Kong Special Administrative Region; China, Macao Special Administrative Region
16	9060	R. Eastern Asia	Democratic People's Republic of Korea; Mongolia
17	392	Japan	Japan
18	410	Korea	Republic of Korea
19	356	India	India
20	9200	R. South-Eastern Asia	Brunei Darussalam; Cambodia; Lao People's Democratic Republic; Malaysia; Myanmar; Philippines; Singapore; Thailand; Timor-Leste; Viet Nam
21	360	Indonesia	Indonesia
22	9220	R. Western Asia	Armenia; Azerbaijan; Bahrain; Cyprus; Georgia; Iraq; Israel; Jordan; Kuwait; Lebanon; Occupied Palestinian Territory; Oman; Qatar; Saudi Arabia; Syrian Arab Republic; United Arab Emirates; Yemen
23	792	Turkey	Turkey
24	9230	R. Eastern Europe	Belarus; Bulgaria; Czech Republic; former Czechoslovakia; former German Democratic Republic; Hungary; Poland; Republic of Moldova; Romania; Slovakia; Ukraine; former USSR
25	643	Russia	Russian Federation
26	924	Northern Europe	Åland Islands; Channel Islands; Denmark; Estonia; Faeroe Islands; Finland; Guernsey; Iceland; Ireland; Isle of Man; Jersey; Latvia; Lithuania; Norway; Sark; Svalbard and Jan Mayen Islands; Sweden; United Kingdom of Great Britain and Northern Ireland
27	925	Southern Europe	Albania; Andorra; Bosnia and Herzegovina; Croatia; Gibraltar; Greece; Holy See; Italy; Malta; Montenegro; Portugal; San Marino; Serbia; Slovenia; Spain; The former Yugoslav Republic of Macedonia; former Yugoslavia; Serbia and Montenegro
28	926	Western Europe	Austria; Belgium; France; Germany; former Federal Rep. of Germany; Liechtenstein; Luxembourg; Monaco; Netherlands; Switzerland
29	36	Australia	Australia
30	554	New Zealand	New Zealand
31	9090	R. Oceania	Norfolk Island; Fiji; New Caledonia; Papua New Guinea; Solomon Islands; Vanuatu; Guam; Kiribati; Marshall Islands; Micronesia (Federated States of); Nauru; Northern Mariana Islands; Palau; American Samoa; Cook Islands; French Polynesia; Niue; Pitcairn; Samoa; Tokelau; Tonga; Tuvalu; Wallis and Futuna Islands

Appendix Table 2: Indicator variables for the evaluation of the quality of the generated migrant streams

Variable	Description	Coding
A. Indicator variables for year 1 and year 2 migrant stock data		
trcat	Identifies the treatment category of the particular profile. Category 1 files are of highest quality because the used profiles were available directly from the particular year and stream. Categories 2a, 3a, and 4a are of slightly lower quality because profile information was derived from another year but the same stream. Categories 2b, 3b, 4b, 4c are of lowes quality because profiles were derive using upper level stream information.	1 = total, age & gender 2 = total & gender 3 = total & age 4 = total
profyear	Identifies the year for which the profile was derived. Allows judging the time difference between profile and raw data (relevant for "b" categories).	Numeric
profID	File name of the source data from which the profile was derived. For "b" categories the ID allows identifying the hierarchical level for which the profile was generated (e.g., regional, continental, global, etc.)	String
profUpSt	Indicates whether the profile was derived from upper level streams.	1 = profile from upper-level streams 0 = profile from same stream
unifGendProf		
profGendSpec	Indicates whether the derived profile has gender differentiated age groups.	1 = gender differentiated age groups 0 = age groups not gender differentiated
profCount	Identifies the number of country level streams that contributed to the particular upper-level profile. Higher numbers indicate better representativeness of the region-level age and gender profiles	Numeric
unifGendProf	Identifies whether a uniform gender distribution was used. Applies to cases where an age profile was available but no gender information was available.	1 = uniform gender distribution 0 = all other cases
B. Indicator variables for migrant stream measures		

trcat1Both	Indicates whether the flows were computed using category 1 files for both years.	1 = flows computed from two category 1 files 0 = flow computation involves other categories
criterionSame	Indicates whether both files use the same criterion of migrant enumeration (country of birth vs. country of citizenship).	1 = files use same criterion 0 = files use different criterion
profileDif	Indicates whether both profiles come from different years.	1 = profiles from different years 0 = profiles from same year