Published in final edited form as:

Hist Methods. 2015 ; 48(3): 113–127. doi:10.1080/01615440.2014.999150.

Indirectly Estimating International Net Migration Flows by Age and Gender: The Community Demographic Model International Migration (CDM-IM) Dataset

Raphael J. Nawrotzki^{b,c,a} and Leiwen Jiang^b

^bNational Center of Atmospheric Research (NCAR), Climate and Global Dimensions Division (CGD), Integrated Assessment Modeling Group (IAM).

^cUniversity of Colorado at Boulder, Institute of Behavioral Sciences (IBS), CU Population Center.

Abstract

Although data for the total number of international migrant flows is now available, no global dataset concerning demographic characteristics, such as the age and gender composition of migrant flows exists. This paper reports on the methods used to generate the CDM-IM dataset of age and gender specific profiles of bilateral net (not gross) migrant flows. We employ raw data from the United Nations Global Migration Database and estimate net migrant flows by age and gender between two time points around the year 2000, accounting for various demographic processes (fertility, mortality). The dataset contains information on 3,713 net migrant flows. Validation analyses against existing data sets and the historical, geopolitical context demonstrate that the CDM-IM dataset is of reasonably high quality.

Keywords

International net migration; bilateral net migrant flows; age; gender; database

Introduction

As the world has entered “the age of migration” (Castles, de Haas, and Miller 2013), increased numbers of people move between countries with substantial socioeconomic and environmental impacts (Bilsborrow et al. 1997, UNPD 2011). International migration plays an important role in determining the size, growth, and structure of national and regional populations (UNPD 2005). Policy makers need information about the number of migrant inflows and outflows and their demographic characteristics (e.g., age and gender) to plan for future economic development and the expansion of social services (CGD 2009, Bilsborrow et al. 1997). It has been recognized that migrants play an important role in sustainable development, through remittances and skilled labor migration, and a better understanding of the characteristics of migrant flows may help to design targeted policies and programs to maximize the benefits related to the movement of people (UN 2013). Moreover, to study the

^aCorresponding author: Raphael J. Nawrotzki, National Center of Atmospheric Research (NCAR), CGD/IAM, Mesa Lab, 3090 Center Green Drive, Boulder, CO 80301, Raphael.Nawrotzki@colorado.edu; Phone: 303-819-6725.

impacts of economic and environmental changes on human populations of different demographic characteristics, researchers need data on the composition of international migration flows (Brown 2008, Piguet et al. 2010). However, obtaining high quality data on international migration is notoriously difficult for multiple reasons (Bilsborrow et al. 1997, CGD 2009), including governments' unwillingness to release this data (UNECE 2012), under-registration of migrants and limitation of data coverage (Nowok et al. 2006), variations in data quality across countries (Abel 2013; Raymer et al. 2013), and inconsistent definitions of enumerating migrants (Raymer et al. 2013).

In this article, we follow the United Nations (UN) and define an international migrant as "any person who changes his or her country of usual residence" (UN 1998, 17). Moreover, it is important to distinguish between *migrant stock* and *migrant flow*. Censuses, national surveys, and registration systems typically record the number of migrants residing in a particular destination country at any point in time, and this number is referred to as "migrant stock" (UN 1998, 88). The migrant stock data can help compute "migration flow" estimates, which measure the number of migrants moving from an origin to a destination during a specific time period (e.g., per year) (Perruchoud and Redpath-Cross 2011). Migrant stock data is easier to obtain, and therefore, more readily accessible than migrant flow data (Bilsborrow et al. 1997). However, migrant flow data better reflects the dynamics of the migration process and enables researchers to examine migratory behavior under differing socioeconomic, demographic and environmental circumstances (Raymer et al. 2013).

Various organizations have put forth great efforts in compiling migration raw data sets (CGD 2009). As of 2014, the UN Global Migration Database (UNGMD) (Henning and Hovy 2011, Zlotnik et al. 2010, UNDESA 2008), developed by the United Nation Population Division (UNPD), constitutes the most comprehensive collection on international migration stock data disaggregated by country of birth, citizenship, sex, and age. Based on the UNGMD database, the Development Research Group at the World Bank constructed the Global Bilateral Migration data set, containing international migrant stock data for the period of 1960–2000 (Ozden et al. 2011). Using migrant stock data in a log-linear regression framework, Abel (2013) and Abel and Sander (2014) estimated decennial migration flows across 191 and 196 countries, respectively. Abel, Samir, and Sander (2013) have most recently employed this methodology to estimate global bilateral flows by gender. Taking advantage of the better quality of the European migration data, two recent projects, the MIMOSA (MIgration MOdelling for Statistical Analyses) project (de Beer et al. 2010; Raymer et al. 2011) and the IMEM (Integrated Modeling of Europe Migration) project (Raymer et al. 2013) estimated age and sex specific international migration flows within Europe.

The aforementioned data collection efforts and modeling approaches have helped advance the study of international migration. However, each of the available data sets has certain limitations: (1) Global data sets compiled by the UN Population Division (e.g., UNDESA 2008, 2013a, 2013b) contain age and sex specific information regarding migrant stocks but not on migrant flows; (2) Abel and Sander's (2014) data set has global coverage but reports bilateral flow counts that are not disaggregated by age or gender (although Abel, Samir, and Sander [2013] outline a methodology to estimate bilateral flows by gender); (3) the IMEM

data set (Raymer et al. 2013) overcomes both of these flaws and contains bilateral migrant flow data by age and gender, but only for European countries.

As such, researchers have not yet compiled a data set containing bilateral migration flow information by age and gender with global coverage. Detailed information on the age and gender composition of migrant flows would improve global population projections that require historical migration flow data as baseline input to estimate future population changes. In addition, detailed information on the age and gender composition of international migration flows would allow researchers to analyze the impact of global factors, including economic crises and climate change on the migration behavior of various vulnerable sub-populations (e.g., migrant women and girls), which has been recently identified by the UN General Assembly as a key area of concern (UN 2013).

In an attempt to begin filling the void in publically available data sources, we created a novel data set as part of the National Center of Atmospheric Research (NCAR) Community Demographic Model (CDM) project. The CDM International Migration (CDM-IM) data set is publically available and contains detailed information on age and gender specific bilateral net migrant flows around the year 2000. In this paper, we report the methods and techniques used to generate and evaluate the CDM-IM data set.

Data

The raw data used in this study originates from the United Nations Global Migration Database (UNGMD) (Henning and Hovy 2011, UNDESA 2008), and was extracted during the period of May 2010 to May 2011. The complete UNGMD data set contains 46,431 useable files, covering 6,759 unique origin destination combinations (about 7 files per origin destination combination), operating at the national (n=6,193) and regional level (n=566). We could not use files that reported multiple countries as origin (e.g., “Algeria, Tunisia, Morocco”), files that referred to regions without an existing UN code (e.g., “Africa (other)”, “Melanesia (unknown)”, “Tonga (partly)”, “Born at sea”), and files that used the same country as origin and destination. The UNGMD records migrant stock information for each available year (1975 to 2009) in a separate text file. Gender differentiated migrant counts are available for 85% of the files, while disaggregation by age is less common with 52% of the files reporting five or more age categories. The reported number of age groups vary widely (from 1 to 102 age groups), with an average of 10 age groups. The UNGMD defines migrants’ country of usual residence as “country of birth” (54%) or “country of citizenship” (46%), depending on the nation-specific method of enumeration. However, information that would allow for a more detailed classification of migrant status by duration of stay (e.g., long-term versus short-term migrants, UN [1998]) is not available.

Method

We used the programming language and statistical environment of the R project (R Core Team 2014) and adopted three steps to accomplish the data cleaning and processing: (1) we selected high quality files of two years closest to the year 2000 for each origin destination combination (e.g., migrants from Mexico [origin] residing in the U.S. [destination]); (2) we

standardized the data files; and (3) we used migrant stock information from the two years selected in step 1 to compute fertility and mortality adjusted profiles of bilateral net migration flows. The method used for each step is described below (for technical details see Nawrotzki and Jiang, 2014).

Selection of the best file for two years

We used a decision-tree structure and adopted five criteria to select the files containing the most detailed information at two time points (referred to as “Year 1” and “Year 2”) for each origin destination combination. The five criteria include: the temporal distance from the year 2000, the information content (e.g., we preferred files with age and gender information over files containing age information only), the criterion used to define migration status (we prioritized “country of birth” over “country of citizenship”), the quality of the age grouping (we favored a larger number of age groups), and a random number to select among cases of equal quality. For Year 1, we chose the highest quality file closest to the year 2000. For Year 2, we selected the file closest to the year 2000, but different from Year 1.

Standardization of the data files

Due to the significantly varied information content in the raw data, we developed a protocol to derive age and gender profiles with a standardized data format for all origin destination combinations. We first assigned each file to one of four treatment categories: *category one* files contain both age and gender information; *category two* files contain gender but no age information; *category three* files contain age but no gender information; and *category four* files contain only the total number of migrants and lack age or gender information. After distributing the migrant counts across single year groups, we collapsed the data to 16, five-year age groups (e.g., 0–4, 5–9, ..., 70–74, 75+). For some migrant stock records that lacked age and/or gender information (e.g., *category two, three, four*), we obtained complementary information from data files for a different year than selected for Year 1 and Year 2. For origin destination combinations that did not contain such complementary information, we employed region-level gender and/or age profiles that we constructed by aggregating the available nation-level migration information.

Mortality and fertility adjustments

We computed bilateral net 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); we then divided the quantity by the period length to derive the average annual changes. As formally described in equation 1, to obtain an accurate estimation of the bilateral net migration flow, we excluded the impacts of mortality and fertility. This approach conceptually follows Rogers and von Rabenau (1971), but instead of using a multiregional matrix model, we employ a classical residual method to estimate net changes of migrants in the stocks after accounting for the effects of fertility and mortality. The adjustments are based on the assumption that migrants quickly adopt the fertility and mortality patterns of the destination country (Goldstein 1973, Visaria 1969, White et al. 2005).

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

In equation 1, M_t^x and M_{t-n}^{x-n} represent the number of migrants of age x at time t and of age $x-n$ at time $t-n$, respectively. The parameter S^x reflects the conditional survival probability of population from age $x-n$ to x , while I^{x-n} and O^{x-n} denote the immigrants and emigrants, respectively, 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 who reside in the country of destination at $t-n$. The term $(I^{x-n} - O^{x-n}) * (1 - S^x)$ reflects the number of deaths of net immigrants who moved to the destination during the period of $t-n$ to t . This formalization constitutes a simplification that we deemed sufficiently accurate. A more detailed method of accounting would assume half of the net immigrants arrived exactly at the beginning of the period, while the other half arrived at the end of the interval (Preston et al. 2001). 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 $\left(M_{t-n}^{x-n} - \frac{O^{x-n}}{2} \right) * (1 - S^x)$ to compute the number of deaths of non-moving migrants, and assume 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 ($N^x = I^{x-n} - O^{x-n}$), we simplify equation 1 to equation 2:

$$N^x = M_t^x / S^x - M_{t-n}^{x-n} \quad (2)$$

When the UN data defined migrants by “country of birth”, we calculated the number of net migrants of age 0 to n using equation 3, where N^{0-n} represents the net migrants of age 0 to n , and M_t^{0-n} reflects the number of migrants of age 0 to n at time t .

$$N^{0-n} = M_t^{0-n} + M_t^{0-n} * (1 - S^n) / 2 \quad (3)$$

When migrants are defined by “country of citizenship”, government officials typically count the children born to female migrants in the country of destination as migrants, although they never moved. Therefore, we have to exclude the number of female migrants’ births from the number of migrants of age 0 to n (equation 4).

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

In equation 4, the parameter b^i represents the fertility of women, age i . Assuming that the sex ratio of births is equal to one, we allocate the total new births evenly between the male and female groups. We obtained the age, gender, and country-specific survival and fertility rates from the UN Population Division (UNPD 2009a; UNPD 2009b)

Results and Discussion

Using the described methodology, we generated a data set of 3,713 country-level bilateral net migration flows by age and gender. Table 1 summarizes the distribution of these age and gender profiles of bilateral net migration flows at the national level across 31 global regions (see Appendix Table 1 for information on the country composition of each region); it also shows the proportion of available bilateral net migration flows relative to the total possible number, which varied greatly between regions. For example, for regions such as Australia and the U.S., the dataset provides comprehensive coverage, while no data exists for the *Rest of Eastern Asia, India, and Indonesia*.¹

In addition, we generated a map as a visual depiction of bilateral net migration flows between global regions (Figure 1). Panel (a) depicts positive bilateral net migrant flows between the origin (population loss) and destination (population gain), showing North America (orange lines) and Europe (light green lines) as the major migrant receivers. Other important migration destinations include: Australia and New Zealand (red lines), South Africa on the Africa continent (dark green lines), and Japan for East Asia (green lines).

In contrast, Panel (b) shows Russia (light green lines) as the destination with the largest negative bilateral net migration flows. The return of Russian migrants from Eastern European countries to Russia during the post-1990 era, following the breakdown of the former Soviet Union, likely produced these negative net flows (Pilkington 1998).

Three cases illustrate how the region-specific context uniquely shapes the age and gender profile of a bilateral net migrant flow (Brown and Bean 2006). Age and gender profiles for bilateral net migrant flows may substantially deviate from profiles derived from migrant stock data because net flow profiles investigate changes over time (Rogers 1990). As the first example, our results suggest that children less than 19 years comprise the majority of U.S.-born immigrants residing in Mexico (Figure 2a and 2b). Although quite different from the bell-shape curve of a usual migration schedule, the observed shape falls in line with reports of large numbers of U.S.-born children returning to Mexico (Giorguli Saucedo and Gutiérrez Vázquez 2011).

Most likely, these young U.S. citizens are the children of Mexican migrant workers born in the U.S. Mexican migrants are typically target earners who stay only temporarily in the U.S. (Lindstrom 1996). Once the migrant has achieved a saving target, they return to their origin, accompanied by their U.S.-born children. Under conditions of an extended stay, the Mexican migrant may choose to send the U.S.-born child back to live with their family in Mexico. In addition, studies suggest an increase in the number of migrant women and

¹Unfortunately, the UN Global Migration Database does not provide sufficient migrant stock data to use our methodology to estimate detailed net migration flow profiles for India, Indonesia, and Rest of Eastern Asia (North Korea, Mongolia) as destinations. However, we do estimate net migrant profiles for those regions as origins. These countries either provide no detailed age profiles for their migrant stock data (e.g., India, Mongolia), have detailed gender and age profiles but only for one year (e.g., Indonesia), or there is simply no migrant stock data recorded (e.g., North Korea). Since our methodology of computing net migrant flows requires age and gender specific migrant stock data for a minimum of two time points, we were unable to compute net migrant flows for these countries as destinations.

movements of entire families, increasing the likelihood of child birth among Mexican migrants residing in the U.S. (Marcelli and Cornelius 2001).

An interesting phenomenon, which is not apparent from the migrant stock data, becomes visible in the age and gender-specific profile of the bilateral net migrant flow (Figure 2c): The positive net migration flow from the U.S. to Mexico only exists for young children below age 10. Between ages 10 to 30, a significantly negative net flow suggests that U.S.-born children of Mexican migrants move back from Mexico to the U.S. for better educational and employment opportunities (c.f., Kemper 2005). This finding echoes other studies that show an enrollment increase among children from Spanish speaking families in American schools (Ruiz-de-Velasco, Fix, and Clewell 2000).

The net migration flow from Mozambique to South Africa provides the second example and reveals a different pattern (Figure 3). Historically, South Africa has employed large numbers of migrant workers from Mozambique in its mining, agriculture, and construction sectors (Crush 1999; McDonald et al. 2000). Because of the hard manual labor in those industries, Mozambican migrants were traditionally male (Crush and McDonald 2001). In addition, the migrant stock data also reveal an increase in the total number of migrants between 1996 and 2001. This increase can be attributed to the influx of refugees after the Mozambique civil war (Hargreaves et al. 2004; Polzer 2007), and political and legal changes (e.g., new immigration acts) in South Africa after the fall of the Apartheid system (Crush and McDonald 2001; Polzer 2007)

The bilateral net migrant flow diagram (Figure 3c) demonstrates an increase in net inflow, particularly among young male Mozambican migrants ages 20–30. In contrast, a negative net migration among males ages 45–64 suggests that older males returned to Mozambique. This observation parallels other research confirming that very few Mozambican migrants intend to settle permanently in South Africa (Crush 1999; McDonald et al. 2000).

Migration from Indonesia to Malaysia provides the third and final example (Figure 4). The gender profiles of the migrant stock reveal that the initial gender difference (more male migrants) in 1991 (Figure 4b) leveled off in 2000 (Figure 4a); Indonesia's migration history helps to explain this change. In Indonesia, labor migration has long served as a remedy to the problem of unemployment and as a source of foreign exchange through remittances (Tsai and Tsay 2004). A number of factors make neighboring Malaysia a particularly desirable destination including: geographical proximity, linguistic, religious and ethnic similarities, a growing economy with high wages, and a history of reliance on migrant workers in response to local labor shortages (Hugo 1995, 2002). During the Asian financial crisis of 1997/98, the Indonesian government actively encouraged labor emigration (Tsai and Tsay 2004), leading to the doubling of total number of international labor migrants between 1991 and 2001 (Tsai and Tsay 2004).

In the early 1990s, the nature of employment opportunities (e.g., manual work in plantation and construction sectors) has led to a male-dominated gender distribution of the migrant stock (Hugo 1995). However, a tendency of joint migration with male family members, emerging employment opportunities (e.g., factory work, prostitution), and the changing roles

and social status of women help to explain the increased proportion of females in the year 2000 migrant stock and estimated bilateral net migrant flow (Hugo 2002). In summary, the shape of the age and gender profiles of the examined three bilateral net migration flows accurately reflects the complex socio-historical context in which these flows emerged.

Validation

For a quantitative assessment of the data quality, we compare the CDM-IM data with two published sources of international migration data: (1) the migrant profiles computed as part of the IMEM project (Raymer et al. 2013), and (2) the table of global transition flows generated by Abel and Sander (2014). While both data sets allow for a comparison of the total bilateral net flows, only the IMEM project data (available for European countries only) can validate the net migrant age and gender profiles.

Total migration

To compare the total bilateral net migration flows, we standardize the flow measures in all three data sets to reflect the average bilateral net migration per year. From Abel and Sander's (2014) migrant transition flow table, we derived net migration flows by subtracting outflows from inflows. From the publically available IMEM data base, we extracted age and gender-specific median net migration flow values. IMEM provides estimates for seven years (2002 to 2008), and we employed the average value across those years. Using a pairwise t-test, we first compared the size of the average total bilateral net migrant flows (see Table 2). To evaluate the shape of the net migrant profiles, we then conducted a separate t-test for each age and gender group, which allows for a more nuanced evaluation in differences/similarities compared to tests that contrast entire distributions (e.g., Kolmogorov-Smirnov test).

For the t-test, we restrict the comparisons to high quality cases in the CDM-IM data set (category 1 files). To guard against the potentially biasing influence of unstable estimates for small net flows, we also present comparisons that only use cases composed of more than 100 net migrants (Table 2, panel B). Although the mean values are not identical, the t-test revealed no significant differences among the three data sets. The variations in the mean values may stem from: the differences in computation methods, differences in the time period of data reporting, and inconsistencies in the reporting of migration flows (c.f., Raymer et al. 2011).

Migrant profiles by age and gender

To compare the shape of migrant profiles, we select those bilateral net migration flows for which data is available in both, the CDM-IM and IMEM data sets. We use the proportion of migrants in a particular age and gender category as input for a pairwise t-test (see Table 3); we perform the t-test for the age categories 0–9, 20–44, and 50–69, which typically show a distinct shape within migration profiles.

The average age and gender profiles are largely similar for high quality cases (Panel A), and no statistically significant differences emerged. However, a few differences become evident when comparing large, high quality net migrant flows (Panel B). For middle-aged males

(age 20–44), IMEM predicts a larger proportion of net migrants, while for older females (age 50–69), IMEM predicts a smaller proportion of net migrants compared to CDM-IM. We graphically depict the average proportion of net migrants in each 5-year age category to help evaluate the shape differences (Figure 5).

The shapes of the curves largely follow a similar trajectory. However, the IMEM data shows a more pointed peak around age category 25–29, while the CDM-IM data is more dispersed in the tail for older ages. To evaluate the differences in the shapes for specific bilateral net migrant flows, Figure 6 shows the age and gender profiles for three randomly selected cases.

In summary, the validation tests demonstrate a high degree of similarity between total bilateral net migrant flows and age and gender profiles in the CDM-IM data set and other publically available data sources. The observed differences likely stem from the different computation methods employed. While we compute the net migrant flow estimates in the CDM-IM data directly from the raw data, the IMEM project uses a Bayesian modeling strategy to obtain the estimates (Raymer et al. 2013). A modeling approach can produce smoother curves, but relies on various assumptions, such as Bayesian priors to include subjective expert opinions (Raymer et al. 2013), which may or may not appropriately reflect reality. Numerous sources of uncertainty, inherent in a complex modeling approach, may lead to estimates that substantially differ from the observed flows. A more simplistic approach using raw data and basic demographic adjustments (as in the CDM-IM data set), may produce a more conservative picture of the true net migration flows.

Conclusions

The CDM-IM data set constitutes a novel collection of age and gender-specific bilateral net migrant flows. Such a data set is presently unavailable and may be used for multiple purposes. First, models to project national and regional population change require historical migration flow data as the base year input. Conventional approaches (e.g., UNPD 2013) employ a single standard migrant profile for all countries, likely resulting in biased projections, given that migration serves as a major demographic force with large variations by country and region. Using the CDM-IM data to appropriately represent the shape of gender and age profiles for unique bilateral net flows will allow researches, non-governmental organizations, and governmental agencies to produce more accurate population projections for future years.

Second, detailed information on the age and gender composition of net migrant flows will allow researchers to conduct multi-country analyses of the impact of macro-level forces (e.g., economic crises, environmental factors, socio-political change) on the migration behavior of various sub-populations. For example, a researcher may be interested in the connection between occurrences of natural disasters and the international net migration of working age adult males in African countries. Presently, the CDM-IM data set is the only data source that would allow for such a detailed multi-country analysis of a migration response among a specific age/gender sub-population. More generally, any study that requires country-specific, international net migration information disaggregated by age/gender sub-populations (e.g., as control variables) will benefit from the CDM-IM data set.

Finally, researchers may use the CDM-IM data set as the base for fitting migration model schedules (Rogers et al. 2005), or as a standard to benchmark the accuracy of data sets derived from modeling approaches.

However, this data set is not without limitations. First, the available raw data limit the computation of bilateral net migration flows by age and gender. Although we use the most comprehensive migration data set currently available (the UNGMD), it does not include information on all possible origin destination combinations. For example, a lack of available data prevented the computation of detailed net migrant profiles for India, Indonesia, and Rest of Eastern Asia as destinations; although, we were able to estimate net migrant flows originating from these regions.

Second, we computed net flows of differing data quality to generate the largest number of migration flows possible. For example, we computed a small number of net flows by employing data based on different enumeration methods (“country of birth” versus “country of citizenship”), or by employing aggregated, region-level age and gender profiles. However, we believe that using these close approximations is better than not having data available for the particular migrant flow. The CDM-IM data set contains a number of data quality flags, which will allow users to select a subset of cases to fit their research needs based on quality considerations.

Finally, measures of net migration should not be mistaken for gross migration measures (emigration and immigration) and are much more limited in their applicability (Rogers 1990). For example, a large immigration flow in combination with a large emigration flow may result in only a small net migration flow. In addition, the age structure of the net migration flow may differ from the gross-migration flows (Rogers 1990). Furthermore, our bilateral net migrant flow measures do not account for third country moves.² Using gross migrant flow measures by age and gender would remedy these limitations but such data is presently not available.

With these advantages and limitations in mind, we trust that the CDM-IM data set will help improve population projection models, will allow researchers to gain a deeper understanding of the complex nature of migration patterns, and will facilitate cross sectional analysis of the determinants of specific migrant flows. Moreover, we hope that this paper contributes to the discussion on improving the collection, measurement, and analysis of bilateral migration flow data.

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 in extracting the raw data, and we express our thanks to Jamie Jones for her careful editing and helpful suggestions. We are grateful for the insightful comments of the journal editor and three anonymous reviewers.

²We assume that a bilateral net flow between origin country “A” and destination country “B” can be measured by an increase/decrease of the stock of people of country “A” residing in country “B”. However it is possible, that the change in the stock of “A”s residing in “B” is the result of a movement of “A”s that were living in a third country “C” to “B” and not from a direct move from “A” to “B”.

References

- Abel GJ. Estimating global migration flow tables using place of birth data. *Demographic Research*. 2013; 28:505–546.
- Abel GJ, Sander N. Quantifying Global International Migration Flows. *Science*. 2014; 343(6178): 1520–1522. [PubMed: 24675962]
- Abel, GJ.; Samir, KC.; Sander, N. Joint Eurostat/UNECE Work Session on Demographic Projections. Rome, Italy: 2013. Examining the role of international migration in global population projections.
- Bilsborrow, RE.; Hugo, G.; Oberai, A.; Zlotnik, H. International migration statistics: Guidelines for improving data collection systems. Geneva, Switzerland: International Labour Office; 1997.
- Brown, O. Migration and climate change. Geneva, Switzerland: International Organization for Migration; 2008.
- Brown, SK.; Bean, FD. International Migration. In: Posten, D.; Micklin, M., editors. *Handbook of population*. New York: Springer Publishers; 2006. p. 347–382.
- Castles, S.; de Haas, H.; Miller, MJ. The age of migration: International population movements in the modern world. New York: The Guilford Press; 2013.
- CGD. Migrants count: Five steps toward better migration data. Washington D.C.: Center for Global Development; 2009.
- Crush J. The discourse and dimensions of irregularity in post-apartheid South Africa. *International Migration*. 1999; 37(1):125–151. [PubMed: 12322067]
- Crush J, McDonald DA. Introduction to special issue: Evaluating South African immigration policy after apartheid. *Africa Today*. 2001; 48(3):1–13.
- de Beer J, Raymer J, van der Erf R, van Wissen L. Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe. *European Journal of Population-Revue Europeenne De Demographie*. 2010; 26(4):459–481. [PubMed: 21124647]
- Giorguli Saucedo SE, Gutierrez Vazquez EY. Ninos y jovenes en el contexto de la migracion internacional entre Mexico y Estados Unidos. *Coyuntura Demografica*. 2011; 1:21–25.
- Goldstein S. Interrelations between migration and fertility in Thailand. *Demography*. 1973; 10(2):225–241. [PubMed: 4789434]
- Hargreaves JR, Collinson MA, Kahn K, Clark SJ, Tollman SM. Childhood mortality among former Mozambican refugees and their hosts in rural South Africa. *International Journal of Epidemiology*. 2004; 33(6):1271–1278. [PubMed: 15319405]
- Henning S, Hovy B. Data sets on international migration. *International Migration Review*. 2011; 45(4):980–985.
- Hugo G. International labor migration and the family: Some observations from Indonesia. *Asian and Pacific Migration Journal*. 1995; 4(2–3):273–301. [PubMed: 12320103]
- Hugo G. Effects of international migration on the family in Indonesia. *Asian and Pacific Migration Journal*. 2002; 11(1):13–46.
- Kemper, RV. Mexicans in the United States. In: Ember, M.; Ember, CR.; Skoggard, I., editors. *Encyclopedia of Diasporas*. New York: Springer; 2005.
- Lindstrom DP. Economic opportunity in Mexico and return migration from the United States. *Demography*. 1996; 33(3):357–374. [PubMed: 8875068]
- Marcelli EA, Cornelius WA. The changing profile of Mexican migrants to the United States: New evidence from California and Mexico. *Latin American Research Review*. 2001; 36(3):105–131.
- McDonald DA, Zinyama L, Gay J, de Vletter F, Mattes R. Guess who's coming to dinner: Migration from Lesotho, Mozambique and Zimbabwe to South Africa. *International Migration Review*. 2000; 34(3):813–841.
- Nawrotzki, RJ.; Jiang, L. NCAR Technical Note TN 508+STR. Boulder, CO: National Center of Atmospheric Research; 2014. The Community Demographic Model International Migration (CDM-IM) dataset: Generating age and gender profiles of international migration flows; p. 1–47.
- Nowok, B.; Kupiszewska, D.; Poulain, M. Statistics on international migration flows. In: Poulain, M.; Perrin, N.; Singleton, A., editors. *THESIM: Towards Harmonised European Statistics on*

- International Migration. Louvain-la-Neuve: UCL Presses Universitaires de Louvain; 2006. p. 203-231.
- Ozden C, Parsons CR, Schiff M, Walmsley TL. Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000. *World Bank Economic Review*. 2011; 25(1):12–56.
- Perruchoud, R.; Redpath-Cross, J. Glossary on migration. Geneva, Switzerland: International Organization for Migration; 2011.
- Piguet E. Linking climate change, environmental degradation, and migration: a methodological overview. *Climate change*. 2010; 1(4):517–524.
- Pilkington, H. Migration, displacement and identity in post-Soviet Russia. London: Routledge; 1998.
- Polzer T. Adapting to changing legal frameworks: Mozambican refugees in South Africa. *International Journal of Refugee Law*. 2007; 19(1):22–50.
- Preston, S.; Heuveline, P.; Guillot, M. *Demography, Measuring and Modeling Population Process*. New York: Blackwell; 2001.
- Raymer J, de Beer J, van der Erf R. Putting the pieces of the puzzle together: Age and sex-specific estimates of migration between EU/EFTA countries, 2002-2007. *European Journal of Population*. 2011; 27:185–215.
- Raymer J, Wisniewski A, Forster JJ, Smith PWF, Bijak J. Integrated Modeling of European Migration. *Journal of the American Statistical Association*. 2013; 108(503):801–819.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. <http://www.R-project.org/>.
- Rogers A. Requiem for the Net Migrant. *Geographical Analysis*. 1990; 22(4):283–300.
- Rogers A, Castro L, Lea M. Model migration schedules: Three alternative linear parameter estimation methods. *Mathematical Population Studies*. 2005; 12(1):17–38.
- Rogers A, von Rabenau B. Estimation of interregional migration streams from place-of-birth-by-residence data. *Demography*. 1971; 8(2):185–194. [PubMed: 5163989]
- Ruiz-de-Velasco, J.; Fix, M.; Clewell, BC. Overlooked and underserved: Immigrant students in U.S. secondary schools. Washington, D.C.: The Urban Institute; 2000.
- Tsai, PL.; Tsay, CL. Foreign direct investment and international labour migration. In: Ananta, A.; Arifin, EN., editors. *International Migration in Southeast Asia*. Pasir Panjang, Singapore: Institute of Southeast Asian Studies; 2004. p. 94-136.
- UN. Recommendations on statistics of international migration. New York: United Nations; 1998.
- UN. Declaration of the high-level dialogue on international migration and development (68/4). New York: United Nations; 2013.
- UNDESA. United Nations Global Migration Database (UNGMD). New York: United Nations, Department of Economic and Social Affairs, Population Division; 2008.
- UNDESA. Trends in international migrant stock: Migrants by age and sex. New York: United Nations Department of Economic and Social Affairs; 2013a.
- UNDESA. Trends in international migrant stock: Migrants by destination and origin. New York: United Nations Department of Economic and Social Affairs; 2013b.
- UNECE. 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; 2012.
- UNPD. World population prospects 2004 revision. New York: United Nations Department of Economic and Social Affairs, Population Division; 2005.
- UNPD. World Population Prospects: The 2008 Revision. New York: United Nations, Department of Economic and Social Affairs, Population Division; 2009a. Age-specific fertility rates by major area, region, and country.
- UNPD. World Population Prospects: The 2008 Revision. New York: United Nations, Department of Economic and Social Affairs, Population Division; 2009b. Male life table survivors at exact age, $l(x)$, by major area, region and country.
- UNPD. International migration report 2009: A global assessment. New York: United Nations Department of Economic and Social Affairs, Population Division; 2011.

- UNPD. World Population Prospects 2012 revision. New York: United Nations Department of Economic and Social Affairs, Population Division; 2013.
- UNSD. Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. New York: United Nations Statistical Division; 2013. retrieved June 18, 2013, from <http://unstats.un.org/unsd/methods/m49/m49regin.htm>.
- Visaria, P. Urbanization, migration and fertility in India. In: Cambell, AA., et al., editors. The family in transition. Washington, D.C.: National Institute of Health; 1969. p. 257-284.
- White MJ, Tagoe E, Stiff C, Adazu K, Smith DJ. Urbanization and the fertility transition in Ghana. Population Research and Policy Review. 2005; 24(1):59–83.
- Zlotnik, H.; Guest, P.; Hovy, B.; Henning, S. 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; 2010.

Appendix Table 1: Definition of 31 global regions

Region	Letter ID	Number ID	Countries in Region
Eastern Africa	EAF	910	Burundi; Comoros; Djibouti; Eritrea; Ethiopia; Kenya; Madagascar; Malawi; Mauritius; Mayotte; Mozambique; Réunion; Rwanda; Seychelles; Somalia; South Sudan; Uganda; United Republic of Tanzania; Zambia; Zimbabwe
Middle Africa	MAF	911	Angola; Cameroon; Central African Republic; Chad; Congo; Democratic Republic of the Congo; Equatorial Guinea; Gabon; Sao Tome and Principe
Northern Africa	NAF	912	Algeria; Egypt; Libya; Morocco; Sudan; Sudan; Tunisia; Western Sahara
Western Africa	WAF	914	Benin; Burkina Faso; Cape Verde; Cote d'Ivoire; Gambia; Ghana; Guinea; Guinea-Bissau; Liberia; Mali; Mauritania; Niger; Nigeria; Saint Helena; Senegal; Sierra Leone; Togo
R. Southern Africa	RSF	913a	Botswana; Lesotho; Namibia; Swaziland
South Africa	ZAF	710	South Africa
Mexico	MEX	484	Mexico
R. Central America	RCM	916a	Belize; Costa Rica; El Salvador; Guatemala; Honduras; Nicaragua; Panama
Caribbean	CRB	915	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
R. South America	RSM	931a	Argentina; Bolivia (Plurinational State of); Chile; Colombia; Ecuador; Falkland Islands (Malvinas); French Guiana; Guyana; Paraguay; Peru; Suriname; Uruguay; Venezuela (Bolivarian Republic of)
Brazil	BRZ	76	Brazil
Canada and R. N. America	RNM	905a	Bermuda; Canada; Greenland; Saint Pierre and Miquelon
USA	USA	840	United States of America
R. South-Central Asia	RSC	921a	Kazakhstan; Kyrgyzstan; Tajikistan; Turkmenistan; Uzbekistan; Afghanistan; Bangladesh; Bhutan; Iran (Islamic Republic of); Maldives; Nepal; Pakistan; Sri Lanka

Region	Letter ID	Number ID	Countries in Region
China	CHN	156	China; China, Hong Kong Special Administrative Region; China, Macao Special Administrative Region
R. Eastern Asia	REA	906a	Democratic People's Republic of Korea; Mongolia
Japan	JPN	392	Japan
Korea	KOR	410	Republic of Korea
India	IND	356	India
R. South-Eastern Asia	RSE	920a	Brunei Darussalam; Cambodia; Lao People's Democratic Republic; Malaysia; Myanmar; Philippines; Singapore; Thailand; Timor-Leste; Viet Nam
Indonesia	IDN	360	Indonesia
R. Western Asia	RWS	922a	Armenia; Azerbaijan; Bahrain; Cyprus; Georgia; Iraq; Israel; Jordan; Kuwait; Lebanon; Occupied Palestinian Territory; Oman; Qatar; Saudi Arabia; Syrian Arab Republic; United Arab Emirates; Yemen
Turkey	TUR	792	Turkey
Northern Europe	NEU	924	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
Southern Europe	SEU	925	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
Western Europe	WEU	926	Austria; Belgium; France; Germany; former Federal Rep. of Germany; Liechtenstein; Luxembourg; Monaco; Netherlands; Switzerland
R. Eastern Europe	REU	923a	Belarus; Bulgaria; Czech Republic; former Czechoslovakia; former German Democratic Republic; Hungary; Poland; Republic of Moldova; Romania; Slovakia; Ukraine; former USSR
Russia	RUS	643	Russian Federation
Australia	AUS	36	Australia
New Zealand	NZL	554	New Zealand
R. Oceania	RON	909a	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

Note: The composition of the 31 global regions follows largely the UN's regional hierarchy (UNSD 2013). However, the 31 global regions differ slightly by incorporating considerations regarding socio-economic differences among countries located in the same region. For example, South Africa is socio-economically more advanced than its neighboring countries and we designate South Africa as a unique region that is distinct from the "Rest of Southern Africa" region (Botswana, Lesotho, Namibia, and Swaziland).

Figure 1.

Global representation of positive and negative net migration flows between 31 regions

Note: Lines represent net migration flows between 31 global regions. Panel a shows positive net migration flows, while panel b depicts negative net migration flows. Line color indicates the continent of destination. For example, the orange line connecting the U.S. and Northern Europe in Panel a shows a positive net migration flow from Northern Europe (origin) to the U.S. (destination). In contrast, a green line connecting Russia with R. Western Asia in Panel b shows a negative net migration flow from R. Western Asia (origin) to Russia (destination),

indicating a return of Russian migrants to their home country. Line thickness represents the size of the migration flow. Due to data limitations, net migrant flows to India, Indonesia, and Rest of Eastern Asia (North Korea, Mongolia) as destinations are not displayed.

Figure 2.

Migrant stock data and resulting net flows for migration from the U.S. to Mexico

Note: Gray line represents female migrants; black line represents male migrants. Criterion of enumeration: Country of birth. Data source: Census.

Figure 3.

Migrant stock data and resulting net flows for migration from Mozambique to South Africa

Note: Gray line represents female migrants; black line represents male migrants. Criterion of enumeration: Country of birth. Data source: Census.

Figure 4.

Migrant stock data and resulting net flows for migration from Indonesia to Malaysia

Note: Gray line represents female migrants; black line represents male migrants. Criterion of enumeration: Country of birth. Data source: Census.

Figure 5.

Average proportion of net migrants by age and gender comparing the CDM-IM and IMEM data sets

Note: We computed the average proportions for net migration flows with more than 100 migrants derived from high quality data (category 1 files).

Figure 6.

Comparing the age and gender profiles of bilateral net migrant flows using three randomly selected origin destination combinations in the CDM-IM and IMEM datasets

Note: Gray lines represent female migrants; black lines represent male migrants.

Table 1
Distribution of available age and gender profiles of bilateral net migration flow data across 31 global regions

Region	Ctries	Pos. Str.	Obs. Str.	% Cov.	% High Qual.
Eastern Africa	20	4840	69	1.4	34.8
Middle Africa	9	2178	20	0.9	15.0
Northern Africa	7	1694	64	3.8	48.4
R. Southern Africa	4	968	21	2.2	23.8
South Africa	1	242	72	29.8	98.6
Western Africa	17	4114	71	1.7	28.2
Brazil	1	242	58	24.0	96.6
Canada and R. N. America	4	968	148	15.3	10.8
Mexico	1	242	54	22.3	100.0
R. Central America	7	1694	139	8.2	61.9
R. South America	13	3146	318	10.1	79.2
U.S.A.	1	242	190	78.5	61.1
China	4	968	15	1.5	93.3
India	1	242	–	–	–
Indonesia	1	242	–	–	–
Japan	1	242	48	19.8	0.0
Korea	1	242	19	7.9	100.0
R. Eastern Asia	2	484	–	–	–
R. South-Central Asia	13	3146	20	0.6	25.0
R. South-Eastern Asia	10	2420	39	1.6	25.6
R. Western Asia	17	4114	50	1.2	6.0
Turkey	1	242	19	7.9	0.0
Caribbean	29	7018	162	2.3	34.6
Northern Europe	18	4356	684	15.7	70.9
R. Eastern Europe	9	2178	76	3.5	21.1
Russia	1	242	18	7.4	0.0
Southern Europe	16	3872	434	11.2	26.5
Western Europe	9	2178	546	25.1	44.0

Region	Ctries	Pos. Str.	Obs. Str.	% Cov.	% High Qual.
Australia	1	242	163	67.4	0.6
New Zealand	1	242	118	48.8	17.8
R. Oceania	23	5566	78	1.4	25.6

Note: Ctries = The number of countries located in a particular global region; Pos. Str. = The number of possible streams flowing into the particular region (Ctries * 242); Obs. Str. = The number of observed streams for which net migrant flow data by age and gender is available in the CDM-IM data set; % Cov. = The coverage of possible streams in the CDM-IM data set; % High Qual. = The percentage of age and gender profiles computed based on high quality raw data (category 1 files).

Table 2

Pairwise two-sample mean comparison (t-test) of total net migration flows, contrasting the CDM-IM, IMEM, and Abel and Sander (2014) data sets.

DF	CDM-IM		Abel & Sander (2014)		IMEM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
<i>Panel A: High quality data (category 1)</i>						
238			-2	1606	-255	4427
238	284	2223	-2	1606		
238	284	2223			-255	4427
<i>Panel B: High quality data (category 1) & Large flows (> 100)</i>						
67			292	2056	33	5799
67	846	4051	292	2056		
67	846	4051			33	5799

Note: Significance (sig.) levels for t-test:

- * p< 0.05;
- ** p< 0.01;
- *** p< 0.001

Table 3

Pairwise two-sample mean comparison (t-test) of net migrants' age and gender profiles in the CDM-IM and IMEM data sets.

Age group	DF	Male			Female			Total	
		A	B	sig.	A	B	sig.	A	B sig.
Panel A: High quality data (category 1)									
0-9	210	0.111	0.115		0.099	0.113		0.210	0.228
20-44	210	0.280	0.306		0.300	0.285		0.580	0.591
50-69	210	0.102	0.088		0.107	0.093		0.210	0.181
Panel B: High quality data (category 1) & Large flows (>100)									
0-9	87	0.084	0.104		0.078	0.100		0.161	0.204
20-44	87	0.282	0.332	*	0.336	0.305		0.618	0.638
50-69	87	0.104	0.079		0.117	0.080	*	0.220	0.158 *

Note: A = CDM-IM; B = IMEM; Significance (sig.) levels for t-test:

* p<0.05;

** p<0.01;

*** p<0.001