Leveraging census data to study migration flows in Latin America and the Caribbean: An assessment of the available data sources

Julieta Bengochea[1], Emanuele Del Fava[2], Victoria Prieto[1], Emilio Zagheni[2]

1. Population Program, University of the Republic, Montevideo, Uruguay
2. Max Planck Institute for Demographic Research, Rostock, Germany

Corresponding author(s): Julieta Bengochea (julieta.bengochea@cienciassociales.edu.uy)

Abstract

Although the scarcity of accurate and accessible data on international migration flows typically prevents a full understanding of migratory patterns, this might not be the case for Latin America, where high-quality census data on migrant flows is publicly available through the project International Migration in Latina America (IMILA). However, such data has mostly been used for research at the regional level because of the fragmented nature of their availability and the lack of English documentation. To tackle this issue, we consolidated data from the IMILA collection to provide a harmonized dataset with five-year flows by country of birth, sex, and age group, for 19 countries of destination and five census waves. Moreover, comparing IMILA to other two available data sources on flows to Latin America, we showed that IMILA provides a more accurate assessment of migration flows from North America and Europe, enables a better quantification of minor migration flows, and enhances the visibility of female migration.

Introduction

In recent years, there has been a number of breakthroughs in indirect methods for the study of global international migration[1–5] that helped overcoming the scarcity and comparability of global migration flow data enabling the comparative analysis of migration trends from 1960 to 2015 for almost all the countries in the world. These include estimates of international flows based on migration stocks elaborated by the Population Division of the United Nations' Department of Economic and Social Affairs (UNDESA)[6,7] or the World Bank (WB). Some of these estimates have incorporated return migration as a component of the global flows[5], or have been broken down by sex[1]. While indirect methods fill important data gaps, their quality, and the quality of the data sources used to generate estimates is difficult to assess, especially in low-income countries. For the case of Latin American and Caribbean (LAC) countries, high-quality census data on both long-term migrant stocks and five-year migrant stocks are typically available. This is essentially golden standard data against which all other estimates

can be assessed. This data has been assembled and harmonized within the project International Migration in Latin America (IMILA)[8], which has provided the basis for major lines of scholarship on the interregional migration. Census data from IMILA has been the main input for the regional research on human mobility in LAC taking place since 1970, providing most of the information on trends, directionality, and demographic composition of the LAC migration system thanks to the international migration stocks by place of birth, place of destination, and migration status on a fixed date[9–11]. Other examples include the analysis of return migration[12,13] and emigration from LAC[14], especially based on the 2010s' census wave, the investigation of the gendered patterns – a topic that has gained increasing attention in the last decades[9,15–21] – and the skilled migration of the LAC international migration[22]. All this research has benefitted from the systematic inclusion in the censuses of relevant questions for the study of international migration, which can provide data on migrant stocks by origin and duration of stay[14]. On the other hand, the fragmented nature of the data, which is made available online through a series of HTML tables which are usually downloaded one at a time, and the fact that the data and their documentation are only available in Spanish may have prevented a more widespread use of it.

In this paper, we argue that census data from IMILA might provide a solid basis to approximate migration flows from figures on recent migration trends. To assess the validity of this argument, first, we provide a harmonized dataset on five-year migrant stocks by country of birth, country of destination and census wave, stratified by age group and sex, and obtained by web scraping the data from the IMILA website; second, we examine the LAC migration system by focusing on intraregional migration from 1995 to 2010 and by comparing IMILA data to other two migration data sources, namely, recent migrant stocks calculated from samples of census data published by the Integrated Public Use of Microdata Series, International (IPUMS-I) and bilateral flows estimates by sex, region of origin, and country of destination produced via indirect methods[1]. Because of the data availability, our analysis is limited to twelve LAC countries in the periods 1995-2000 and 2005-2010; moreover, we assume that recent migration is a good approximation to flows in the form of transitions, namely, the number of people living in the destination country and who had a different place of residence five years prior to the census[23].

Results

IMILA data

So far, the IMILA Project has compiled data for 19 countries of destination from the LAC region and more than 60 countries of birth from all over the world[24]. The main goal of the project is to disseminate information on international migrants by means of a set of tabulations

that report, among others, socio-demographic and socio-economic features of migrants, including sex, age, education, activity status, place of previous residence, and time when the last international migration took place (when available). In particular, this data enables the construction of five-year migration stock matrices by country of birth and country of residence, and also allows a further stratification by sex and age group. Its main limitations include (i) the fact that not all available censuses were included in this project due to data inconsistencies (Table 1), and (ii) the difficulty of extracting data from the project website, in a way that makes it immediately useful for research.

To tackle the latter point, we web scraped the IMILA website and provided a harmonized data set of five-year migrant stocks, which consist of 719 dyads organized by census period and sex, containing information for seven variables: country of birth, country of destination, region of origin (South America; Central America, Mexico, and the Caribbean; USA and Canada; Europe; Rest of the World), and census period (1995 – 2000, 2005 – 2010).

Other available data sources for regional migration in LAC

There are other two data sources worth of consideration in the assessment of intensity, directionality, and demographic profile of migrants for regional migration in LAC.

First, samples of census microdata from IPUMS-I also enable the construction of five-year migration stock matrices by country of birth and country of residence. While IMILA focuses on international migration during a shorter period (1980-2010), IPUMS-I offers data on a wide range of topics covering a larger period (1960 to 2010) (Table 1). As the data is disseminated in the form of microdata at the individual level, only a fraction – from 5% to 10% – of the full censuses is made available. This explains the differences in the figures with IMILA, which uses the full census.

Second, Abel (2018) published a pool of 5- and 10-year global migration origin-destination flow tables by age and sex from 1960 to 2015. His estimates represent the minimum number of transitions from the country of previous residence to the country of next residence (differently from IMILA), consistent with the migrant stocks provided by UNDESA (and with demographic change), and do not explicitly exclude return migration. The main strength of Abel's estimates is the full coverage of migration flows for all countries in a wide temporal range. The main drawback is that the estimates rely on several assumptions, which may lead to discrepancies with actual data on flows for a number of countries.

Critical examination of migration trends in Latin America and the Caribbean

In this section, we examine regional migration in LAC from 1960 to 2010 for a specific set of countries for which the comparison holds based on data availability. By comparing the stock

of recent migrants and estimates on migration flow, we point out how the appraisal of magnitudes and trends varies according to the data sources in use.

The three data sources captured a five-year migration flow with magnitude that varies from 482,123 (according to IMILA) to 594,593 (according to Abel´s estimates for ten LAC destinations in 1995-2000), and from 842,670 (according to IPUMS-I) to 889,638 (according to IMILA) for seven LAC destination countries in 2005-2010 (Figure 1).

IPUMS-I numbers for the recent migrant stock in ten LAC countries in 1995-2000 – including only the origins for which we have comparable data with IMILA and Abel[1] – differ from IMILA numbers by 4.8% for women and 4.7% for men (Table 2). More remarkable discrepancies are found for the same period when looking at the magnitude of the flows estimated by Abel: they differ from IMILA by 18.0% for women and by 23.9% for men. For 2005 and 2010, and with a different pool of seven destination countries, both IPUMS-I and Abel's data are lower than the number of recent migrants provided by IMILA and the variation by sex within each source is less than 1%.

The mismatch between the number of recent immigrants reported by IMILA and the flow estimates produced by Abel increases in magnitude when looking at the figures by country of destination and origin. This is the case for the Dominican Republic in 1995-2000, where Abel's estimate is 133.1% larger than IMILA's total number and 159.7% larger than IMILA's number for women; Brazil, where the total is more than 51.8% higher; Costa Rica, where this gap peaks to 28%. Conversely, for other countries such as Paraguay, Guatemala, Bolivia or Honduras, Abel's estimates are between 48% to 110% lower than IMILA's number of recent migrants (Table A in Annex). Despite identifying differences by sex in the size of the mismatches (e.g., Chile and Mexico), no consistent pattern of a largest bias for one specific sex is observed (Table 2).

The observation of the percentage log changes between Abel's migration flows and the stock of recent migrants reported by IMILA for 2005-2010 shows that the first estimates perform significantly better than in 1995-2000 and are more similar the IPUMS-I data (Table 3). For the countries included in this period, Abel reports on average lower values than IMILA by 6.2% and IPUMS-I figures are down by around 5.4% than IMILA. However, the gap between Abel's and IMILA is still quite large when it comes to main destination countries. Additionally, the differences between IMILA and IPUMS-I data, though quite smaller, are nonetheless quite relevant for Brazil, where IPUMS figures are about 44.5% lower than those reported by IMILA on recent migrants (Table 2). Again, we do not find any pattern of systematic bias for any specific sex.

When we focus on the regions of origin, we can confirm what has been shown for the disaggregated data by destination, i.e., the gap between the different data sources is significantly larger when looking at the numbers from 1995-2000. Additionally, the

differences are larger when comparing Abel's estimates to IMILA in 1995-2000 and tend to be smaller as we move to the set of countries considered for the period 2005-2010 (Table 3). For 2005-2010, IPUMS-I figures perform rather poorly for the flow coming from Asia, Australia and Africa (-25.6%) and Europe (-18.9%), while the greatest mismatch for Abel (2018) estimates in this period corresponds to the inflow from Europe (-99.1%) and US and Canada (+47.6%).

Another way of visualizing the findings from the comparison between a census sample of recent migration stocks and Abel's estimates is provided by Figure 2, where the focus is on the absolute numbers and the population exchange from large regions of origin to five selected countries that are well established destinations of the South American (Argentina and Brazil), Central American (Costa Rica and Mexico), and Caribbean (Dominican Republic) migration systems. This enables to pinpoint a more disaggregated view of the mismatches between the data sources, as it narrows down the comparison to a pool of dyads for which we have data for both periods under study. Although the migration dynamics in the selected countries of destination under the three data sources are rather similar, some differences emerge when looking at the size of the flows. First, the increase in the recent migrants between the two periods is almost double for IMILA (around 83%) and, to a lesser extent, IPUMS-I, than for Abel's (around 41%). Second, it appears that Abel overestimates the migration flow from the US and Canada to Mexico, which seems to be the main destination of international migration in the LAC region, while it underestimates the magnitude of the South American migration addressing Argentina and Brazil. In contrast, IMILA and IPUMS-I estimates point to Argentina as the main destination of international migration in LAC. Also, Abel underestimates the flow to the Dominican Republic in 2005 – 2010, which seems to be the main destination of migrants born in Central America, Mexico and the Caribbean and is even larger than the flow to Brazil.

Gendered patterns of migration in Latin America and the Caribbean

With respect to the implications for female participation, the comparison between recent stock and flow estimates indicates that these three data sources provide different outcomes. If we take the five countries included in Figure 3 and we look at the regions of origin of the immigrants, we found little differences in the share of female migration between IMILA and IPUMS-I data. On the contrary, larger differences between both census data sources and Abel's estimates stand out (Figure 3). For example, for these five countries, the migration from US and Canada is dominated by women according to Abel's, while both IPUMS-I and IMILA show a recent stock dominated by male migration. Conversely, Abel underestimates female immigration compared to IPUMS-I and IMILA when looking at flows

from Europe to Argentina and Brazil in both study periods, and from all origins (except US and Canada) to Mexico.

Discussion

In this paper, we provide a harmonized dataset on five-years migrant flows to 19 countries in Latin America and the Caribbean over 5 census waves, with the data obtained from the censuses collected by the IMILA project. We also assessed the validity, the scope, and the accessibility of the information on migration flows and recent migrants for the study of migration in the LAC countries provided by IMILA, comparing its data with those from IPUMS-I and Abel[1].

Both IMILA and IPUMS-I have information on recent migrants from census data, the former by providing ready-to-use aggregated data, while the latter micro-data sampled from census. We observed discrepancies between the two sources in the absolute values of recent migrants and the sex composition. We expect these differences to be contained within the confidence intervals (CIs) associated with the IPUMS-I census sample, although it is not possible to verify it, since IPUMS-I only provides CIs for some specific data aggregation.

On the other hand, Abel[1] does not present a clear pattern of overestimation or underestimation with respect to IMILA. As these estimates represent the minimal flow that is compatible with migrant stocks at the beginning and the end of the considered periods, it is expected that they underestimate the number of recent migrants. Thus, it is useful to use Abel's estimates to understand the trends and directionality of migratory systems as long as we acknowledge that their accuracy in terms of magnitude of bilateral flows is less precise. Those indirect estimates are affected by the accuracy of the information provided by UNDESA[7]. It can be seen that the phenomenon of return migration from Europe of LAC migrants – mainly from Spain starting after the 2008 crisis[12,13] – is not captured in Abel's estimates for 2005-2010, while it plays a significant role in IMILA and IPUMS-I estimates, where European children and spouses who accompanied Latin American returnees are captured (Table 2). Also, it is relevant to remark the differences in the estimations of migration flows between Abel's and IMILA, since they show some discrepancies in terms of acknowledging the feminization or masculinization of some flows, such as those from US and Canada to Argentina in 2005 – 2010 or to Costa Rica in 1995 – 2000 and 2005 – 2010. These are predominantly male in Abel's estimates and predominantly female in IMILA. These differences should not only be considered in quantitative terms, but also in terms of theoretical and interpretative implications, for example in assessing whether a certain flow is dominated by one sex or another.

The difference in geographical coverage by census round between IMILA and IPUMS-I with Abel's is relevant since the latter has information for all LAC countries in both

periods of analysis. On the other hand, IMILA has information for 14 countries in the period of 1995-2000 and for 15 countries in the period 2005–2010, while IPUMS-I has information for 19 countries in the first period and 15 countries in the second period. In this sense, Abel's estimates may be quite useful for research since they have a geographical coverage of 100%. Data accessibility does not constitute a problem in any of the three sources since data is easily available online. As regards to IMILA, we provide, for the first time, a harmonized data set on migration stocks by residence five years before the census, which allows to obtain information on five-year flows to a set of 19 LAC countries. The R scripts made available with this work can be easily modified to extract and harmonize the further available information on stocks by socio-demographic factors from the project. IPUMS-I has a much friendlier mechanism to download the data, where the researcher can decide in which format data can be extracted. The information is at the micro level, with weights that allow to estimate quantities at the whole population level, although there is no possibility of obtaining the CIs associated with the estimates. Of the three data sources, Abel's data are the most user-friendly to download, as the author offers data files in CSV format with CIs for the estimates.

One additional advantage of the data available in IMILA and IPUMS-I is that they contain information on recent migrants stratified by sex and age, while Abel's only provides information by sex. Moreover, these two sources provide other sociodemographic information such as educational level and activity status, among others, although in IMILA they are only available for migrant stocks.

Regarding temporary coverage, Abel's work provides the greater coverage (1960 - 2015) and guarantees comparability over time. On the other hand, both IMILA and IPUMS-I have a temporal coverage ranging from 1960 to 2010, with variations in the years of observation depending on the timing of the censuses in each country. However, our results show that, for some origin-destination dyads, Abel's estimates are not accurate. For its part IMILA has the higher quality since it is census data while IPUMS-I provides a census sample and Abel estimations are based on UNDESA data. IMILA and IPUMS-I allow the analysis by sex and age, which are essential stratifications to obtain an exhaustive study of the migratory systems. Also census data would enable scientists to address return migration. In addition, it provides a large list of demographic and sociodemographic attributes, is available through microdata and tabulations and allows for the identification of recent and long-term changes in patterns of migration. Also, we believe that Abel's estimations could be far more accurate if census data, when available, were directly used for the estimation. IMILA has great potential, especially if it improved the following three aspects: making data access easier, completing the information from the missing censuses, and having sociodemographic information, beyond gender and age, for recent migration. In this sense, we provide the IMILA data in a friendlier format than the one currently available on the website.

In conclusion, each of these sources have strengths and weaknesses, so it is difficult to position one over the other, but our results not fully validate Abel´s statement about the possibility of misrepresenting contemporary migration trends when stock data is used in the case of LAC migration system. On the contrary, Abel´s estimations may not be capturing some migration patterns for the LAC migration system. Therefore, we argue that, in countries with a large tradition of census data, it is possible to rely on the number of recent migrants provided by this data source about every ten years, instead of using migration flow estimates derived from UNDESA input data. In this sense, we believe that, while selecting one source should be based on the specific research question, and an in-depth analysis of the data features, for the study the LAC migration system IMILA is the best data source.

Data and Methods

In this study we use three data sources and two kinds of migration data, i.e., the estimated flows in a five-year period and the stock of recent immigrants (who moved to the destination within five years prior to the census) by place of previous residence and place of birth. Data on stocks of recent immigrants comes from IMILA and IPUMS-I, while the flows data corresponds to the estimates published by Abel[1].

The IMILA Project was created in 1970 by the Population Division of the Economic Commission for Latin America and the Caribbean (CEPAL) with the purpose of collecting data on international migrants captured by the LAC censuses. The data are made available from the project website (https://celade.cepal.org/bdcelade/imila/) and are tabulated by country of residence, census, and country of birth. For each stratum, 13 archives with different stratifications are available. This project enables the construction of migration matrices by country of birth as origin and country of current residence as destination. For reasons of data protection, data is not shown when one stratum contains less than 300 cases, and not all available censuses were included in this project (Table 1). The biggest strength of this data source relies on the quality of census data for the study of migration as it provides information on the recent migrant stocks –a proxy of migration flows – by socio-demographic attributes. Although IMILA has published so far only one table on the five-year count of recent migrants by origin, sex, and age, it has the potential to replicate this analysis also by educational attainment, economic activity, and number of children born, among other variables. Additionally, it would be possible to work on estimates for return migration or third-country migration as CEPAL stores all LAC full census microdata, which is a convenient advantage over IPUMS-I, that works with census samples. On the other hand, IMILA's website is not user-friendly and not up to the latest standards. It also lacks English translation of documentation and of the data tables, thus making access to the data difficult for international scientists. For example, developing an online tabulator as the one available at IPUMS-I would

enable users to produce remotely any cross table of migration by variables of interest and type of migration flow in real time. At its current state, it is challenging for users to extract the available tables of IMILA, unless they do it using *ad hoc* web scraping scripts. For the sake of our analysis, we developed an R script to extract and harmonize the information contained in two archives, namely, Archive 1 ("Population by Sex and Age") and Table 13 ("Population Aged 5 Years or More, by Place of Residence Five Years Earlier, Sex and Age") for the 1980s to 2010s census waves. We then provide this data together with this article.

IPUMS-I has become a very powerful tool in the study of global migration by making available more than 400 census samples for almost a hundred countries covering 60 years. This platform was developed by the Minnesota Population Centre and currently provides harmonized migration variables on place of birth and place of previous residence under different time windows (1 year, 5 year or 10 years prior to census date) using a unique country code system, which enables researchers to locate individuals from the same country of birth in over 98 countries of residence. In contrast to IMILA, the data provided by IPUMS-I is published in the form of microdata at the individual level. For this reason, only a fraction – from 5% to 10% – of the full censuses is made available.

Abel's work was partially based on the methodology developed in previous studies[2,4], where a demographic accounting framework, supported by an iterative proportional fitting statistical procedure, was used to derive origin-destination flows from bilateral migrant stock tables produced by the World Bank[4] (WB) and UNDESA[2]. Since such migrant stock tables do not contain any indication of the time in which transitions took place, the estimates produced represent the minimum number of people who changed their country of residence matching the UNDESA migration stocks by country of birth for 1990, 2000 and 2010. The latest estimates from Abel's differ from his previous work by (i) covering a longer period and incorporating flow estimates by sex; (ii) accounting for inconsistencies between demographic and stock data, that previously limited the estimation for some countries and periods; and (iii) assessing the sensitivity of the estimates to demographic and stock input data coming from different sources (WB and UNDESA).

The quality and validity of the migration stock estimates produced by UNDESA, and later used in the work of Abel[1], Abel and Sander[2], Azose and Raftery[5], and Abel and Cohen[3], might be quite low for those countries where quite strong assumptions are applied to deal with insufficient empirical data[7]. In fact, UNDESA reports that for 94 per cent of LAC countries their migrant stock estimates were based on census data, while this rate falls to 79 per cent in Central and Southern Asia and to 86 per cent for countries in Sub-Saharan Africa[25].

Abel's dataset on bilateral migration flows estimates is available online. IPUMS-I microdata could be downloaded upon registration and processed in a format comparable to that of the estimates[1]: namely, every row of the database corresponding to the migration data

from origin country (either the country of previous residence or birth) to destination country ("dyad") for a given period. IMILA provides tables of aggregated data, which can be downloaded as single spreadsheet files by year, country, and stratification. To obtain this data in a format consistent with Abel's and IMILA layout, we first scraped the data from the HTML files that compose the IMILA database and formatted them similarly to IPUMS-I and Abel's estimates; we then performed a series of data harmonization steps to (i) match the IPUMS and IMILA country coding with that of Abel (2018) and UNDESA, based on ISO 3166-1 alpha-3 codes; and to (ii) replace incorrect data in the IMILA database on recent immigrants for Argentina 2010 with the original census data obtained from the National Institute of Statistics and Census of Argentina (INDEC). We assess the discrepancies between the three data sources per dyad by examining the differences in the magnitude of the flows, overall and broken down by destination, region of origin, and sex. To have a comparable measure of the discrepancy between the three data sources, we calculate the percentage logarithmic change, computed as $\Delta_{i,IM} \approx 100 * \ln(x_i/x_{IM})$, where x_i represents the migration numbers from Abel (2018) or from IPUMS, to depict the differences with respect to the IMILA data (x_{IM}), taken as benchmark in this study.

To deal with census data, it was necessary to sort out the fact that LAC censuses are not conducted every ten years at the end of every decade (Table 1). We thus classified census dates to fit around 2000 and 2010 census waves as follows: censuses conducted from 1995 to 2004 were considered as CIRCA 2000, while censuses conducted from 2005 to 2014 were classified as CIRCA 2010 following UN classification. However, for the purpose of this comparison it is necessary to use some more restrictive criteria to classify the census waves. Thus, we decided to consider only the censuses conducted between 2008 and 2012 to compare with Abel's estimates on the flows around 2005-2010, and 1998 to 2002 to compare with the estimates on the flows around 1995-2000 (Table 4). A detailed table on migration variables included in censuses conducted for 24 LAC countries is available in the Annex.

As we aimed to compare census data on stock of recent migrants from both IMILA and IPUMS-I with the five-year flow estimates by Abel (2018), we limited the study to the countries satisfying the following three criteria: (i) having a census with all three questions on place of birth, place of residence five years ago, and current place of residence; (ii) the census being conducted around 1998-2002 or 2008-2010, which offers a range relatively closer to the one provided by Abel's estimates (1995-2000 and 2005-2010); (iii) being available on both IMILA and IPUMS-I databases. This narrowed down the number of countries ultimately included in the analysis to 12: ten for 2000s and seven for 2010s (Table 2).

Our final data set consisted of 378 dyads organized by period and sex, containing information for seven variables: country of origin (country of previous residence, for Abel's; country of birth, for IPUMS and IMILA), country of destination, region of origin (South

America; Central America, Mexico, and the Caribbean; USA and Canada; Europe; Rest of the World), period (1995 – 2000, 2005 – 2010), and three dummy variables indicating the data source (IPUMS-I, IMILA, and Abel).

Data availability

The migration data that support the analysis of this study are available at IMILA – CELADE, IPUMS International and Abel (2018) and were derived from the following resources: https://celade.cepal.org/bdcelade/imila/, https://international.ipums.org/international/ and https://journals.sagepub.com/doi/suppl/10.1111/imre.12327. Also, the dataset generated during this study are included in its supplementary information files.

Code availability

The code used for creating the data set that were used in the analysis were conducted using Stata. The script used for the visualizations were implemented in R. Both are available at (view-only link): https://osf.io/f75tn/?view_only=d552b113ea8a4a3a856ad797608d15f2.

Acknowledgements

This research was conducted as part of the research project "Using internet-based data to quantify and sample international migrants. Applications to examine recent immigration to Uruguay" funded by a binational collaboration between Population Program from University of the Republic (Uruguay) and the Max Planck Institute for Demographic Research (Germany), funded by Agencia Nacional de Investigación e Innovación (Uruguay) and the Max Planck Society (MPI_ID_2018_1_1008456).

Author contributions

JB contributed to the general idea and conception of the paper, the systematization of the data, the analysis and discussion of results and the writing of the paper. EDF contributed to the general idea and conception of the paper, the data collection, the analysis and discussion of results and the writing of the paper. VP contributed to the general idea and conception of the paper, the analysis and discussion of results and the writing of the paper. EZ contributed to the general idea and conception of the paper, to the discussion of results and to the writing of the paper.

Competing interests

Authors do not have any conflict of financial, commercial, legal, or professional interest.

References

1. Abel, G. J. Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015. *Int. Migr. Rev.* **52**, 809–852 (2018).
2. Abel, G. J. & Sander, N. Migration Flows. *Science (80-.).* **343**, 1520–1523 (2014).
3. Abel, G. J. & Cohen, J. E. Bilateral international migration flow estimates for 200 countries. *Sci. Data* (2019) doi:10.1038/s41597-019-0089-3.
4. Abel, G. J. Estimating global migration flow tables using place of birth data. *Demogr. Res.* **28**, 505–546 (2013).
5. Azose, J. & Raftery, A. Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proc. Natl. Acad. Sci. U. S. A.* **116**, (2019).
6. United Nations Population Deparment of Economic Affairs, P. D. Trends in International Migrant Stock: The 2017 Revision. https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.asp (2017).
7. United Nations Population Department of Economic and Social Affairs, P. D. Trends in International Migrant Stock: The 2015 Revision. https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates15.asp (2015).
8. Latin American and Caribbean Demographic Centre - Population Division. *Migración internacional en América Latina - IMILA*. (2000).
9. Martínez Pizarro, J. & Orrego Rivera, C. *POBLACIÓN Y DESARROLLO Nuevas tendencias y dinámicas migratorias en América Latina y el Caribe*. (2016).
10. Pellegrino, A. *La migración internacional en América Latina y el Caribe: tendencias y perfiles de los migrantes*. (2003).
11. Villa, M. & Martínez Pizzaro, J. Tendencias y patrones de la migración internacional en América Latina y el Caribe. *Notas Poblacion* **773**, 51–99 (2001).
12. Masferrer, C. & Prieto Rosas, V. El perfil sociodemográfico del retorno migratorio reciente. Diferencias y similitudes en contextos de procedencia y acogida en América Latina. in *¿Volver a casa? Migrantes de retorno en América Latina. Debates, tendencias y experiencias divergentes* 67–126 (El Colegio de México, 2019).
13. Prieto Rosas, V., Pellegrino, A. & Koolhaas, M. Intensidad y selectividad de la migración de retorno. in *Retorno en los procesos migratorios de América Latina. Conceptos, debates, evidencias* 55–80 (ALAP, 2015).
14. Maguid, A. M. *La emigración internacional a través de los censos en países de origen: evaluación de resultados y recomendaciones*. (2008).
15. Cerrutti, M. *Gender and Intra-Regional Migration in South America*. (2009).
16. Donato, K. M. & Gabaccia, D. *Gender and International Migration. From the Slavery Era to the Global Age*. (Russell Sage Foundation, 2015).
17. Prieto Rosas, V. Perfiles demográficos de la migración latinoamericana entre 1950 y 2010. *Rev. la Asoc. Española Demogr. Histórica* **37**, 185–215 (2019).
18. Duarte, I., Liz, R. E., Vicens, L. & Féliz Franco, N. *Movimientos migratorios desde y hacia la*

República Dominicana. (Fondo para el Fomento de la Investigación Económica y Social, 2011).

19. Prieto, V. & López Gay, A. Push and Pull Factors of Latin American Migration. in *Demographic Analysis of Latin American Immigrants in Spain. From Boom to Bust.* (eds. Domingo, A., Sabater, A. & Verdugo, R.) 1–27 (Springer International Publishing, 2015). doi:10.1007/978-3-319-12361-5_1.
20. Giorguli - Saucedo, S. E. & Angoa, M. A. International Migration, Gender and Family: A Miroir from Latin America. in *International Handbook of Migration and Population Distribution* (ed. M., W.) 543–572 (Springer, Dordrecht, 2016). doi:10.1007/978-94-017-7282-2_25.
21. Giorguli-Saucedo, S. E., García-Guerrero, V. M. & Masferrer, C. *A migration system in the making: Demographic dynamics and migration policies in North America and the Northern Triangle of Central-America*. (2016).
22. Pellegrino, A. Éxodo, movilidad y circulación: nuevas modalidades de la migración calificada. *Notas Poblacion* **73**, 129–162 (2001).
23. Nowok, B. & Willekens, F. A probabilistic framework for harmonisation of migration statistics. *Popul. Space Place* **17**, 521–533 (2011).
24. Moya, O. P. Proyecto IMILA. Investigación de la migración internacional en América Latina. in 1–23 (1993).
25. United Nations Population Department of Economic and Social Affairs, P. D. The 2019 Revision of World Population Prospects. https://population.un.org/wpp/ (2019).

Figures

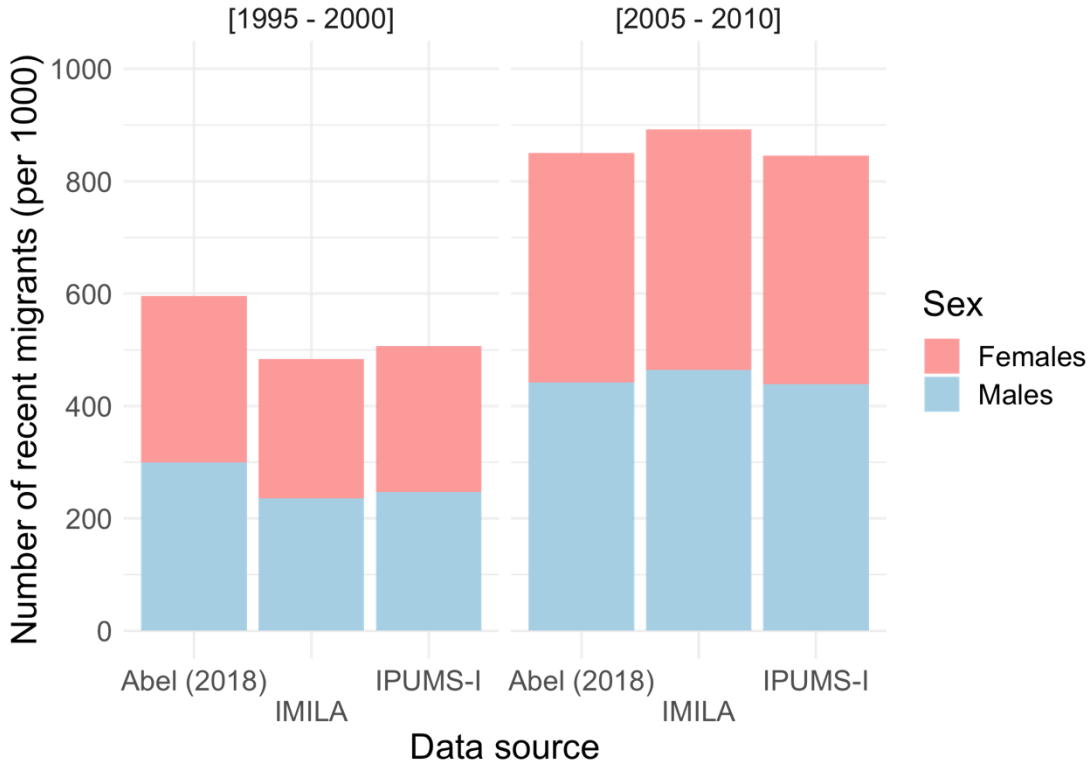

Figure 1. Total number of recent immigrants received by LAC countries by sex and data source, 1995-2000 and 2005-2010. Elaboration based on IPUMS-I census microdata, IMILA tabulations on recent migrants by country of destination, and estimates of migration flows from Abel (2018).

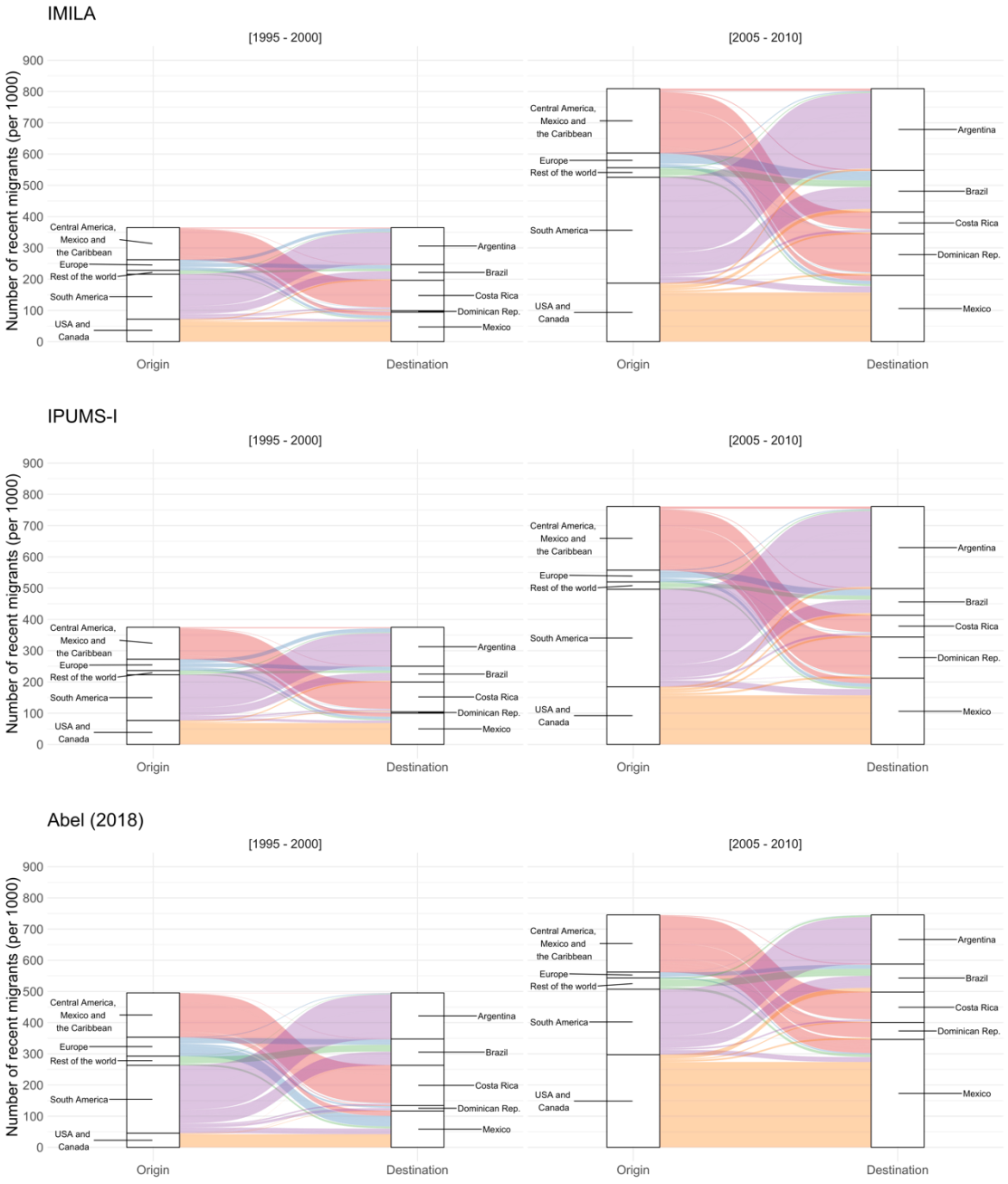

Figure 2. Number of recent immigrants by world area of origin, country of destination and data sources, 1995 – 2000 and 2005 – 2010. Elaboration based on IMILA tabulations on recent migrants by country of destination, IPUMS-I census microdata and estimates of 5-year migration flows from Abel (2018).

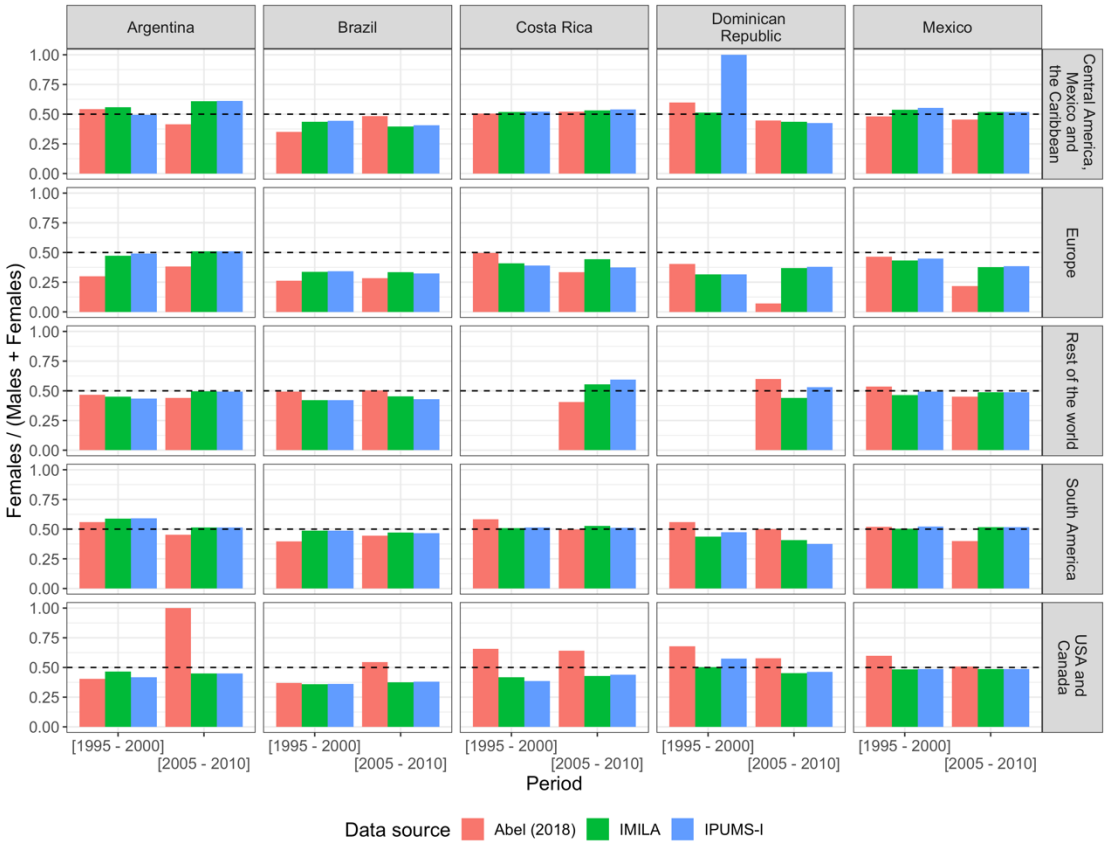

Figure 3. Female proportion of total recent migrants and estimated flows by world area of origin, country of destination, and data source, 1995 – 2000 and 2005 – 2010. Values above the horizontal line at 0.5 indicate a predominance of female migrants. Elaboration based on IPUMS-I census microdata, IMILA tabulations on recent migrants by country of destination, and estimates of migration flows from Abel (2018).

Tables

Table 1 Available census data for Latin America and the Caribbean, CIRCA 1960 to 2010. Own elaboration based on revision of census samples available at IPUMS International and migration matrices from IMILA CELADE, 2019.

Country	1960	1970	1980	1990	2000	2010
Argentina		1970(IP)	1980(IP)	1991(IP)(IM)	2001(IP)(IM)	2010(IP)(IM)
Belize			1980(IM)	1990(IM)		
Bolivia			1976(IP)	1992(IP)(IM)	2001(IP)(IM)	2012(IM)
Brazil	1960(IP)	1970(IP)	1980(IP)	1991(IP)(IM)	2000(IP)(IM)	2010(IP)(IM)
Chile	1960(IP)	1970(IP)	1982(IP)	1992(IP)(IM)	2002(IP)(IM)	*
Colombia	1964(IP)	1973(IP)		1985(IP)		2005(IP)
				1993(IP)(IM)		
Costa Rica	1963(IP)	1973(IP)	1984(IP)		2000(IP)(IM)	2011(IP)(IM)
Cuba					2002(IP)	2012(IM)
Dominican Republic	1960(IP)	1970(IP)	1981(IP)		2002(IP)(IM)	2010(IP)(IM)
Ecuador	1962(IP)	1974(IP)	1982(IP)	1990(IP)(IM)	2001(IP)(IM)	2010(IP)(IM)
El Salvador				1992(IP)(IM)		2007(IP)(IM)
Guatemala	1964(IP)	1973(IP)	1981(IP)	1994(IP)(IM)	2002(IP)(IM)	
Haiti		1971(IP)	1982(IP)		2003(IP)	
Honduras	1961(IP)	1974(IP)		1988(IP)(IM)	2001(IP)(IM)	2013(IM)
Jamaica			1982(IP)	1991(IP)	2001(IP)	
Mexico	1960(IP)	1970(IP)		1990(IP)(IM)	1995(IP)	2005(IP)
					2000(IP)(IM)	2010(IP)(IM) +
Nicaragua		1971(IP)		1995(IP)(IM)		2005(IP)(IM)
Panama	1960(IP)	1970(IP)	1980(IP)	1990(IP)(IM)	2000(IP)(IM)	2010(IP)(IM)
Paraguay	1962(IP)	1972(IP)	1982(IP)	1992(IP)(IM)	2002(IP)(IM)	*
Peru				1993(IP)(IM)		2007(IP)(IM)
Saint Lucia			1980(IP)	1991(IP)		
Trinidad and Tobago		1970(IP)	1980(IP)	1990(IP)	2000(IP)	2011(IP)
Uruguay	1963(IP)		1975(IP)	1985(IP)(IM)	1996(IP)(IM)	2006(IP)
						2011(IP)(IM) +
Venezuela		1971(IP)	1981(IP)	1990(IP)(IM)	2001(IP)(IM)	2011(IM)
IPUMS samples	12	16	16	18	19	15
IMILA samples	-	-	1	19	14	15
% countries with IPUMS samples	50.0%	66.6%	66.6%	75.0%	79.2%	62.5%
% countries with IMILA samples	-	-	4.2%	79.2%	58.3%	62.5%

Notes: Census waves were defined as follows: CIRCA 1960 includes 1955-1964; 1970 includes 1965-1974; 1980 includes 1975-1984; 1990 includes 1985-1994; 2000 includes 1995-2004; 2010 includes 2005-2014; 2020 includes 2015-2019. “IP”, stands for microdata available in IPUMS International; “IM”, stands for aggregated data available in IMILA; (+) in cases where we had more than one census available for the same period we chose the most recent; (*) indicates cases where population census was conducted, but the data was later discharged due to consistency issues, for example, Chile and Paraguay census for 2012.

Table 2 Comparison of recent immigrants and migration flows by country of destination, sex, and data source, in terms of absolute numbers and logarithmic change with respect to IMILA. Latin American countries, 1995-2000 and 2005-2020

Destination	1995-2000 (10 countries)						2005-2010 (7 countries)					
	IPUMS-I		Abel (2018)		IMILA		IPUMS-I		Abel (2018)		IMILA	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Argentina	71,530	52,980	81,773	66,076	67,941	50,307	134,939	127,406	71,428	86,115	134,438	127,068
	5.1%	*5.2%*	*18.5%*	*27.3%*	*Ref.*	*Ref.*	*0,4%*	*0.3%*	*-63.2%*	*-38.9%*	*Ref.*	*Ref.*
Brazil	21,678	28,806	33,226	51,168	21,540	28,742	35,627	49,782	40,787	49,086	57,105	76,137
	0.6%	*0.2%*	*43.3%*	*57.7%*	*Ref.*	*Ref.*	*-47.2%*	*-42.5%*	*-33.7%*	*-43.9%*	*Ref.*	*Ref.*
Costa Rica	49,200	46,380	65,706	63,130	50,001	47,371	36,520	33,030	50,902	46,699	36,136	33,226
	-1.6%	*-2.1%*	*27.3%*	*28.7%*	*Ref.*	*Ref.*	*1.1%*	*-0.6%*	*34.3%*	*34.0%*	*Ref.*	*Ref.*
Dominican Rep.	1,780	2,410	9,051	8,292	1,833	2,747	55,130	74,250	24,120	28,339	57,033	74,042
	-2.9%	*-13.1%*	*159.7%*	*110.5%*	*Ref.*	*Ref.*	*-3.4%*	*0.3%*	*-86.1%*	*-96.0%*	*Ref.*	*Ref.*
Mexico	49,131	50,174	61,146	54,547	45,357	47,967	103,085	108,560	165,381	168,522	102,922	108,513
	8.0%	*4.5%*	*29.9%*	*12.9%*	*Ref.*	*Ref.*	*0.2%*	*0.0%*	*47.4%*	*44.0%*	*Ref.*	*Ref.*
Total	259,193	246,528	295,950	298,643	247,077	235,106	405,501	437,168	402,908	433,479	427,018	462,620
	4.8%	*4.7%*	*18.0%*	*23.9%*	*Ref.*	*Ref.*	*-5.2%*	*-5.7%*	*-5.8%*	*-6.5%*	*Ref.*	*Ref.*

Table 3 Comparison of recent immigrants and migration flows by region of origin, sex, and data source. Latin American countries, 1995-2000 and 2005-2010

Region of origin	1995-2000 (10 countries)						2005-2010 (7 countries)					
	Recent immigrants by IPUMS-I		Migration flow by Abel (2018)		Recent immigrants by IMILA		Recent immigrants by IPUMS-I		Migration flow by Abel (2018)		Recent immigrants by IMILA	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
South America	131,270	110,922	152,041	149,054	125,951	105,384	185,666	182,212	131,035	160,113	197,717	195,593
Central America, Mexico and the Caribbean	61,519	54,919	73,786	72,683	60,746	55,416	97,092	111,052	84,041	87,185	98,833	111,296
US and Canada	40,462	45,371	27,863	18,981	37,230	42,150	92,969	100,412	163,705	151,220	93,288	102,354
Asia, Australia and Africa	7,189	9,239	16,734	19,007	6,593	8,397	11,377	13,435	17,869	19,497	14,847	17,192
Europe	18,753	26,078	25,526	38,918	16,557	23,759	18,398	30,057	6,258	15,464	22,333	36,185
Total	259,193	246,528	295,950	298,643	247,077	235,106	405,501	437,168	402,908	433,479	427,018	462,620

Table 4 Number of census samples available in IMPUMS-I and IMILA, and including data on country of birth, and migration status five years ago. Selected waves, 2000 (1998-2002) and 2010 (2008-2012). Own elaboration based on revision of census samples available at IPUMS International and migration matrices from IMILA CELADE, 2019.

Destination country	Native or foreign born		Country of birth		Recent migration (within 5 years)	
	2000	2010	2000	2010	2000	2010
Argentina	X	X	X	X	X	X
Bolivia	X	NFM	X	NFM	X	NFM
Brazil	X	X	X	X	X	X
Chile	X	NC	X	NC	X	NC
Costa Rica	X	X	X	X	X	X
Dominican R.	X	X	X	X	X	X
Ecuador	X	X	Not asked	X	X	X
Guatemala	X	*	X	*	X	*
Honduras	X	NFM	X	NFM	X	NFM
Mexico	X	X	X	X	X	X
Paraguay	X	NC	X	NC	X	NC
Uruguay	*	X	*	X	*	X

Notes: "X" indicates whether the country include the question; "NC" stands for "No census", indicating that the country conduct a census within this wave but the census was not valid; "NFM" stands for "Not fully matched", which indicates that there was either one missing estimate from IMILA or IPUMS; (*) in this cases we exclude the Uruguay census conducted in 1996 which falls way far from the period 1998-2002 used for the sake of the comparison to Abel's estimates for 1995-2000.

Supplementary

Table A: Comparison of recent immigrants and migration flows by country of destination, sex, and data source, in terms of absolute numbers and logarithmic change with respect to IMILA. Latin American countries, 1995-2000 and 2005-2010

	1995-2000 (10 countries)						2005-2010 (7 countries)					
	IPUMS-I		Abel (2018)		IMILA		IPUMS-I		Abel (2018)		IMILA	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Argentina	71,530	52,980	81,773	66,076	67,941	50,307	134,939	127,406	71,428	86,115	134,438	127,068
Bolivia	10,390	12,190	4,302	6,252	9,176	10,270						
Brazil	21,678	28,806	33,226	51,168	21,540	28,742	35,627	49,782	40,787	49,086	57,105	76,137
Chile	36,760	34,610	31,961	38,753	33,129	28,528						
Costa Rica	49,200	46,380	65,706	63,130	50,001	47,371	36,520	33,030	50,902	46,699	36,136	33,226
Dominican Rep.	1,780	2,410	9,051	8,292	1,833	2,747	55,130	74,250	24,120	28,339	57,033	74,042
Ecuador							36,500	40,500	47,994	52,939	35,990	40,174
Guatemala	5,230	3,900	1,718	1,340	4,929	3,911						
Honduras	2,590	2,170	885	794	2,489	2,571						
Mexico	49,131	50,174	61,146	54,547	45,357	47,967	103,085	108,560	165,381	168,522	102,922	108,513
Paraguay	10,904	12,908	6,182	8,291	10,682	12,692						
Uruguay							3,700	3,640	2,296	1,779	3,394	3,460
Total	259,193	246,528	295,950	298,643	247,077	235,106	405,501	437,168	402,908	433,479	427,018	462,620
Logarithmic change (IMILA standard)*												
Argentina	5.1%	5.2%	18.5%	27.3%	*Ref.*	*Ref.*	0.4%	0.3%	-63.2%	-38.9%	*Ref.*	*Ref.*
Bolivia	12.4%	17.1%	-75.8%	-49.6%	*Ref.*	*Ref.*					*Ref.*	*Ref.*
Brazil	0.6%	0.2%	43.3%	57.7%	*Ref.*	*Ref.*	-47.2%	-42.5%	-33.7%	-43.9%	*Ref.*	*Ref.*
Chile	10.4%	19.3%	-3.6%	30.6%	*Ref.*	*Ref.*					*Ref.*	*Ref.*
Costa Rica	-1.6%	-2.1%	27.3%	28.7%	*Ref.*	*Ref.*	1.1%	-0.6%	34.3%	34.0%	*Ref.*	*Ref.*
Dominican Rep.	-2.9%	-13.1%	159.7%	110.5%	*Ref.*	*Ref.*	-3.4%	0.3%	-86.1%	-96.0%	*Ref.*	*Ref.*
Ecuador							1.4%	0.8%	28.8%	27.6%	*Ref.*	*Ref.*
Guatemala	5.9%	-0.3%	-105.4%	-107.1%	*Ref.*	*Ref.*						
Honduras	4.0%	-17.0%	-103.4%	-117.5%	*Ref.*	*Ref.*						
Mexico	8.0%	4.5%	29.9%	12.9%	*Ref.*	*Ref.*	0.2%	0.0%	47.4%	44.0%	*Ref.*	*Ref.*
Paraguay	2.1%	1.7%	-54.7%	-42.6%	*Ref.*	*Ref.*						
Uruguay							8.6%	5.1%	-39.1%	-66.5%	*Ref.*	*Ref.*
Total	4.8%	4.7%	18.0%	23.9%			-5.2%	-5.7%	-5.8%	-6.5%		

Table B: Dimensions of assessment between sources: IMILA, IPUMS – International and Abel (2018), 1995-2000 and 2005-2010.

Dimensions of our assessment	**Recent migrant stock**		**Migration flow estimates**
	IMILA	**IPUMS-I**	**Abel (2018)**
Descriptor	Aggregated numbers from whole census data	Microdata for 5-10 % census sample	Dyadic estimates of migration flow by sex, based on UN migration stock estimates and other demographic data.
Geographical coverage by destination	2000´s = 14 2010´s = 15	2000´s = 19 2010´s = 15	All LAC countries
Temporal coverage and comparability	1980-2010 varying by country Variation in the years of observations	1960-2010 varying by country Variation in the years of observations	1960-2015 Harmonized dates of observation
Data access	✓ Data is easily available online: Excel format ✗ Data downloading is time consuming as dyads tables should be download one by one	✓ Data is easily available online: CSV, STATA, R, SPSS format. ✗ No CI is available to know how much the figures in the sample differ from actual population	✓ Data is easily available online: CSV format
Availability of sociodemographic variables	✓ Recent migration stocks by sex and age ⚠ Potentially over sociodemographic attributes could be potentially included (however, only for the whole stock, not for the recent subset).	✓ Recent migration stocks by sex and age ✓ All sociodemographic variables included in census	✗ Migration flows by sex