**Bridging Data Gaps: Social Media and Traditional Sources in Assessing Migration in Latin American countries**


## Abstract

This paper assesses the correspondence between Meta web social media data (Facebook, Instagram, and Messenger) and traditional demographic sources—population censuses and household surveys—in estimating total and migrant populations in Latin America. Using data extracted from the Meta Marketing API aligned with census and survey periods for Argentina, Ecuador, Mexico, Paraguay, and Uruguay, the study evaluates the strength and consistency of associations between Meta daily active users and official counts. Multivariate regressions were estimated for both total and migrant populations, controlling for origin, destination, sex, and Meta app penetration rates derived from the Latinobarometer survey. Results reveal that Meta user estimates are a strong and consistent predictor of both census- and survey-based measures. However, the fit is notably higher with census data than with household surveys, both for total populations and for migrant stocks. Among migrant populations, the relationship remains significant even when accounting for Meta penetration at origin and destination, suggesting that coverage and digital habits at either end do not fully explain the association. These findings indicate that Meta data track census figures closely and provide valuable, timely information for migration research in contexts where traditional data are limited or delayed. The results also challenge the assumption of censuses and household surveys as "gold standards," given their growing omissions and inconsistencies, particularly in measuring foreign-born populations.


## Background, aim and contribution

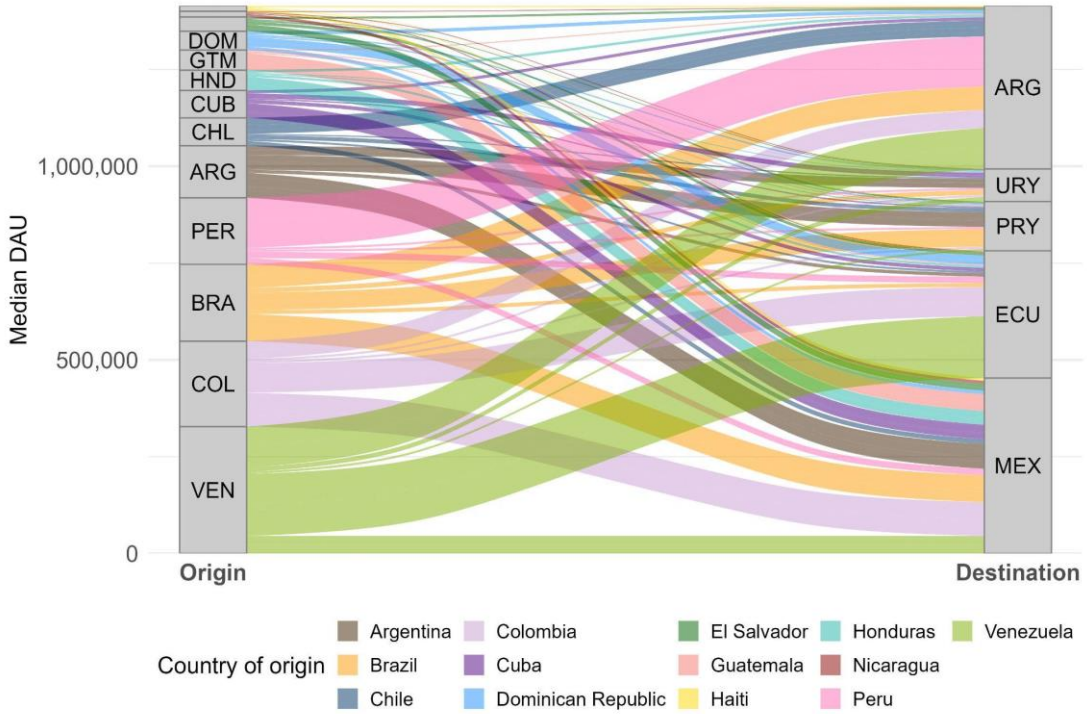
While web social media data offers a real-time alternative for assessing migrant stocks and their sociodemographic characteristics, it has inherent limitations. Several studies have evaluated the biases of social media data, particularly Facebook, by comparing it with high-quality traditional data sources such as censuses and administrative records. These studies have highlighted the potential and the limitations of using digital data to capture migration dynamics. For instance, Zagheni et al. (2017) analyzed Facebook's advertising data to estimate migrant stocks across the United States by comparing figures on monthly expat users to the American Community Survey data. Similarly, Spyratos et al. (2019) used Facebook to estimate international migration from several large diasporas comparing its performance against UNDESA/OECD figures, and Palotti et al. (2020) estimated the international migration of Venezuelans in several destinations providing insights into the discrepancies between digital and official estimates from register data. All these studies have relied on more traditional data sources that enable the assessment of biases and limits of alternative web social media data.

However, in Latin America, with no population register data and where statistical data sources often have their limitations in measuring migrant populations, establishing a reliable 'gold standard' for comparison is particularly challenging. For example, Gutiérrez et al. (2020) highlight several challenges in using household surveys to measure and characterize migrant populations in the region: (i) insufficient sample sizes for accurate estimates across multiple levels of disaggregation; (ii) outdated sampling frames that fail to capture significant changes in migration flows between censuses; and (iii) underestimation of migrant stocks, either due to the exclusion of collective dwellings in survey designs or migrants concealing their status during interviews. In addition, during the 2020 census wave, several Latin American countries incorporated self-report strategies using telephone and online methods (Argentina 2022, Costa Rica 2022, Chile 2024, Ecuador 2022, Mexico 2020, and Uruguay

2023), making census collection more flexible but also raising concerns about coverage and the digital divide, especially in rural and poor urban communities or among people on the move (Del Popolo, 2024). And while in cases, such as Uruguay, there was greater use of census data with administrative records (e.g., civil registries, social security databases) to complement and enhance traditional census data when in-person data collection faced limitations, other aspects directly related to the measurement of migrant populations seem to have followed a backward path. For example, Brazil (2023) has included exhaustive data on international migration in the extended questionnaire applied only to a sample of the total population, while recently released data for Argentina (2022) and Uruguay show a high share of non-response for country of birth among the foreign-born population (Koolhaas et al., *forthcoming*). On top of this, preliminary findings point to an increase in census omission and overuse of imputation of population in non-interviewed housing (Del Popolo, 2024). In this context, the lack of consistent, high-quality traditional data across Latin America makes it difficult to establish a reliable gold standard for measuring migration. As a result, web social media data emerges as a relevant alternative, potentially facing similar limitations but offering a timely and flexible supplement to existing sources that can enable exercises of nowcasting migrant populations.

To further this discussion, this paper compares figures on users of Meta web social media (Facebook, Instagram, and Messenger) available at Meta (formerly Facebook) Marketing API against tradiotional statistic data sources. In particular, we compare the number of Meta web social media users by dyad of origin destination (Figure 1), with census and household survey data from a selection of Latin American countries. Using the data from household surveys and population censuses, we assess the correlation between the number of absolute migrants, disaggregated by place of birth and sex, and analogous data on daily user data from Meta platforms, disaggregated by previous residence. We focus on Latin American countries with both annual labor/household surveys (here on HS) that include questions to identify migrant populations and recent census data. At present, these criteria are met by Argentina (2022), Ecuador (2022), Mexico (2020), Paraguay and Uruguay (2023) which are the five destination included in our study. The origins included in the analysis are determined by the availability of Meta data—which do not consistently tag users by previous country of residence—and by their relevance to the intraregional Latin American migration system, ultimately corresponding to those depicted in Figure 1.

**Figure 1**. Estimated migrant stock for each origin–destination dyad. Based on the median of daily active users on social media (DAU).




Source: Data collected from Facebook API during census periods: Argentina (March 16 – May 18, 2022); Ecuador (October 1 – December 31, 2022); Mexico (January 1 – March 31, 2020); Paraguay (January 1 – December 31, 2022); Uruguay (April 29 – September 29, 2023).

Recently, Varona et al. (2024) and Montiel (2024) conducted similar assessments of Facebook data against the Mexican census and household surveys for migrants in several Latin American countries, respectively. Nevertheless, this is the first simultaneous comparison for the same period where three types of data sources are available, which could provide further evidence to understand how much the direction and magnitude of bias of web social media data vary when we change the standard of comparison. In addition, our analysis here is extended to the whole population, which allows us to understand whether the discrepancy between the different data sources is migrant-specific.

The comparative approach contributes to a broader understanding of the validity and limitations of digital data in migration research for Latin America, while problematizing the idea of a gold standard for traditional statistics, at least in the assessment of migrant populations by simultaneously comparing against HS and census data.

**Data and methods**

For this analysis, we rely on three types of data sources. First, the most recent wave of published census results, which is available for a limited number of countries—Argentina, Ecuador, Uruguay, Paraguay and Mexico. For Argentina and Paraguay, we use published tabulations, for Ecuador and Uruguay, published microdata, and in the case of Mexico, we use a 10% sample of the total census because the full sample is only accessible through the INEGI computing center. All are *de jure* censuses conducted over several months and weeks. A detailed description of the census data and the main innovations introduced in the 2020 census round is provided in the Appendix (Table A1 and A2).

Second, we incorporate household or labor force survey microdata for the same countries, selecting the trimester that overlaps with the census. From both, census and HS data we use information on the total population and the population by place of birth, broken down by sex.

Third, we utilize data extracted from the Meta (formerly Facebook) Marketing API on the daily number of Facebook, Messenger, and Instagram users (here on referred to as Meta data)[1], disaggregated by place of previous residence, gender, and country of current residence. These extracts cover 13 Latin American and Caribbean countries of origin for various destinations in the region, though our analysis focuses on Argentina, Ecuador, and Mexico. To estimate a unique number of users by country of previous residence, aligned with the census enumeration period (Table 1), we calculate a median value from multiple weekly extracts in each destination, using bootstrap estimation to generate confidence intervals around the median.

**Table 1.** Period of reference for used data sources

| Country | Census period | Household Survey period | Facebook period |
|---|---|---|---|
| Argentina | Mar 16-May 18, 2022 | Second quarter of 2022 of Permanent Household Survey (*Encuesta Permanente de Hogares, EPH*). | Mar 16-May 18, 2022 |
| Ecuador | October-December, 2022 | Fourth quarter of 2022 of National Survey of Employment, Underemployment and Unemployment (*Encuesta Nacional de Empleo, Subempleo y Desempleo, ENEMDU*) | October 1-December 31, 2022 |
| Mexico | March 2 to 27, 2020 | First quarter of 2020 of National Survey of Occupation and Employment (*Encuesta Nacional de Ocupación y Empleo, ENOE*) for total population<br>Third quarter of 2021 of National Survey of Occupation and Employment (*Encuesta Nacional de Ocupación y Empleo, ENOE*) for migrant dyads | January 1-March 31, 2020 |
| Paraguay | November 9, 2022 | Continuous Household Survey *(Encuesta Permanente de Hogares Continua; EPHC)* | Annual, 2022 |
| Uruguay | April 29- September 29, 2023 | April to September of 2023 of Continuous Household Survey (*Encuesta Continua de Hogares, ECH*) | April 29-September 29, 2023 |

Note: For Mexico, two data sources were employed, as the 2021 ENOE is the first wave to offer comprehensive information disaggregated by country of origin.

As for the definition of migration used to work with census and HS data, we limit it to absolute migrants, i.e. people born abroad. This is because not all countries have at the moment published information on population by place of residence on a fixed previous date (five years earlier) that would allow us to work with the census definition of recent migrants which, as Varona et al. (2024) showed, has higher levels of fit with the Facebook data.

The analysis focuses on population between 18 to 65 years old as these were the age breaks we include in the data extracts from Meta users.

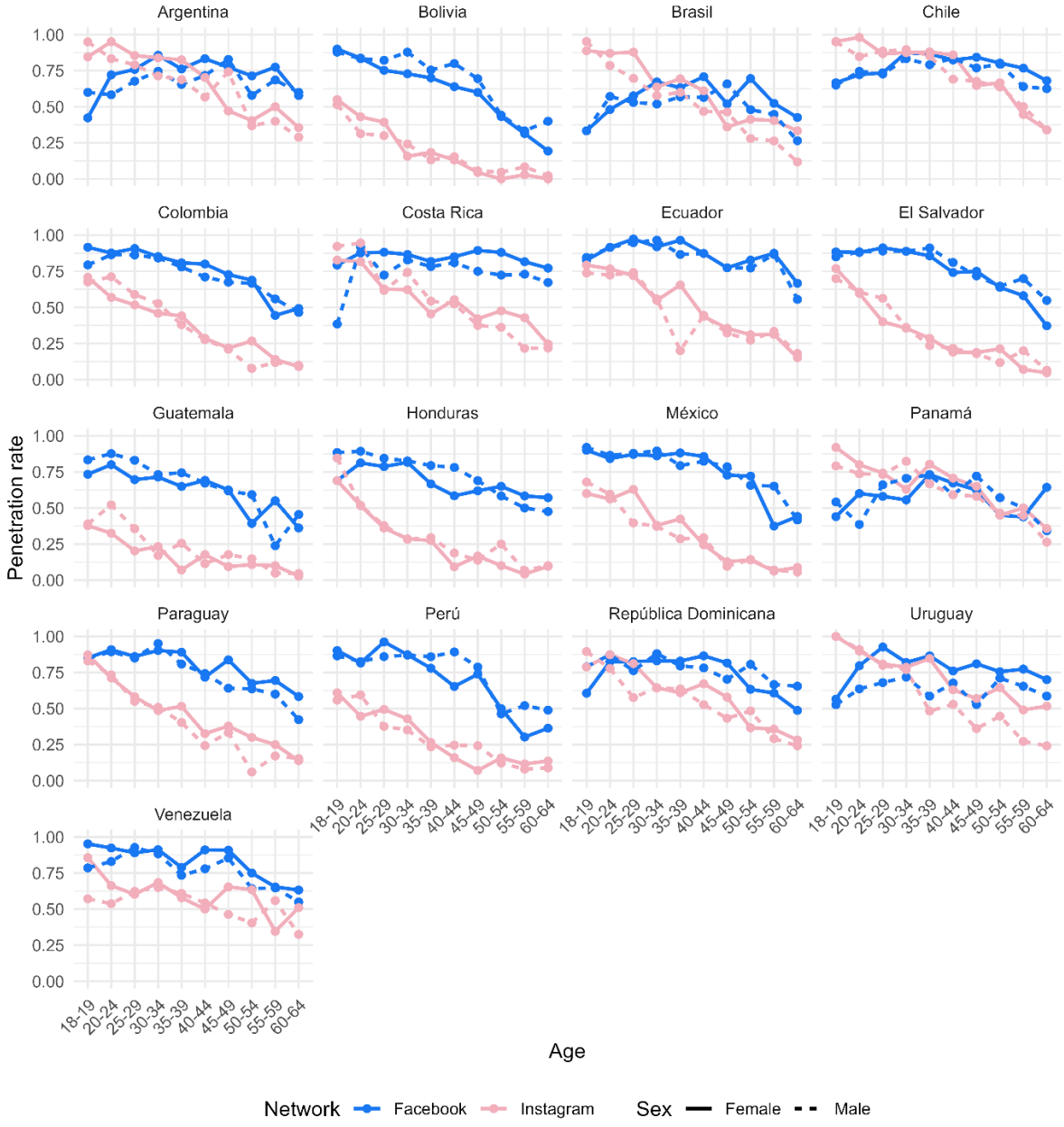
We first conducted a set of bivariate analyses to illustrate the similarities in the total and migrant populations captured by the three information sources—Facebook, census, and household surveys—always distinguishing results by country of destination or enumeration and by sex. We then estimated multivariate linear regression models to examine the correspondence between Facebook daily user data and official demographic estimates following the approach of Zagheni et al. (2017), Montiel

[1] Note that WhatsApp is also a relevant application within the Meta group, but its data are not available for extraction through the Facebook Business API.

(2024), and Varona et al. (2024). Separate models were specified for total and migrant populations, using as dependent variables the counts derived from the census and household surveys. The models progressively incorporate controls for destination country (and origin country in the case of migrant populations), sex, and the level of Meta app penetration, in order to observe how the association between Facebook and traditional data sources varies once structural and compositional differences across countries are taken into account.

The Meta app penetration rate was estimated using data from the Latinobarometer survey corresponding to the same year as the household surveys and census used in the analysis, for all countries except Cuba, Haiti, and Nicaragua, where this survey is not conducted regularly. Latinobarometer includes questions on the regular use of different applications, specifically asking about Facebook and Instagram, the two Meta platforms included in our daily extraction of Facebook user data (Figure 2). Because reported usage varies by age—Instagram being more common among younger respondents and Facebook among adults—we first identified, for each age group, the higher penetration value between the two apps. We then calculated an age-standardized media**n** of these maximum values for the population aged 18–65. The resulting penetration estimates are origin–destination–sex specific, allowing us to account for differences in Meta platform usage patterns across both ends of the migration corridor and between men and women. This indicator was used to capture variation in Meta platform access that may influence migrants' presence on the network, depending not only on the digital habits prevailing in the destination but also on those "imported" from the country of origin. In the multivariate analysis, origin countries for which this information was unavailable were treated as missing, which explains the smaller number of observations included in Models 9 and 10 of the migrant-stock regressions.

**Figure 2.** Instagram and Facebook penetration rates by age. Selected countries, 2022

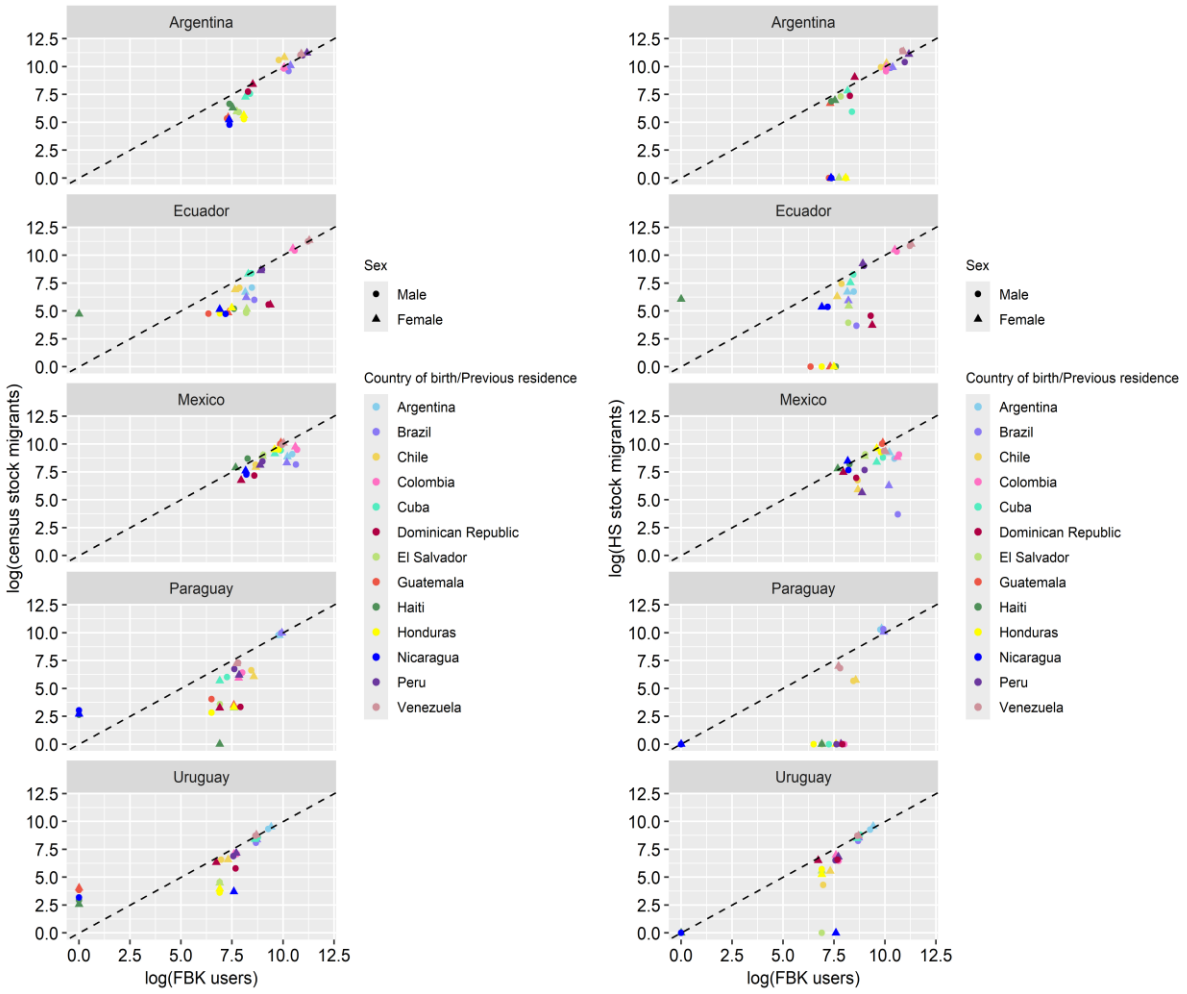



Note: none available data in Latinobarometer for Haiti, Honduras and Cuba.
Source: own elaboration based on data from Latinobarometer annual surveys for 2022.

**Preliminary results**

Figure 3 shows a general alignment between the population figures of interest obtained from Facebook data, surveys, and censuses, with most of the values closer to the identity line. Generally, Facebook data exhibits a better fit with census figures (left column) than with survey data (right column), a pattern particularly evident in Argentina, Ecuador and Mexico.

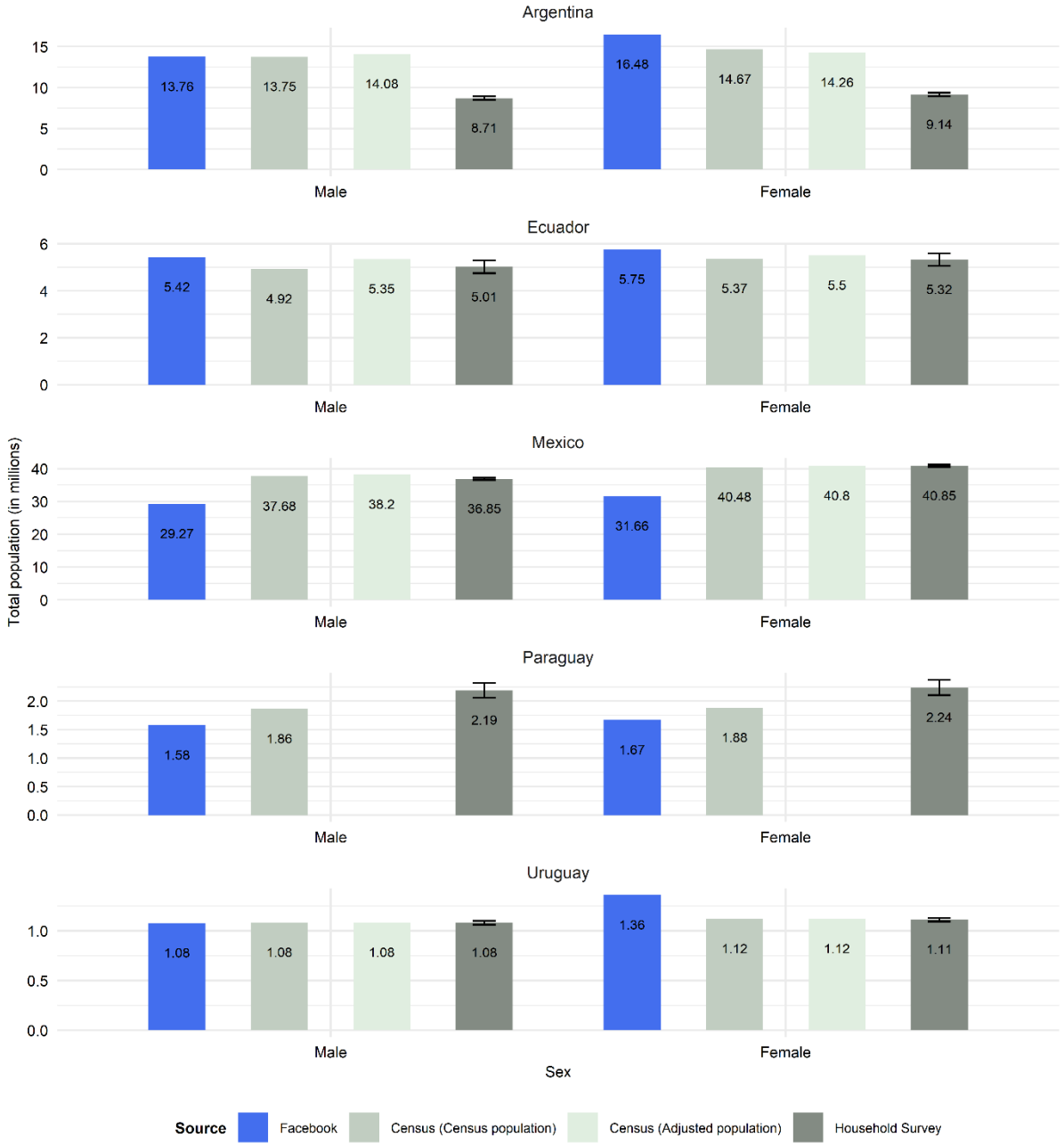
**Figure 2.** Scatter plot for Facebook data against census (left) and household survey data (right) by origin and sex. Argentina, Ecuador, Mexico, Paraguay and Uruguay, circa 2020




Source: own elaboration based on Facebook API extracts, HH survey and census data.

A closer look at the magnitudes from different data sources for selected dyads reveals remarkable similarities between the population estimates from census data and Facebook, while household surveys show less alignment. However, the degree of alignment varies depending on the community of origin; notably, there is a very strong congruence for the Venezuelan community across all three destination countries or other large communities of origin, regardless of whether Facebook data is compared against census or survey (Figures 3 and 4).

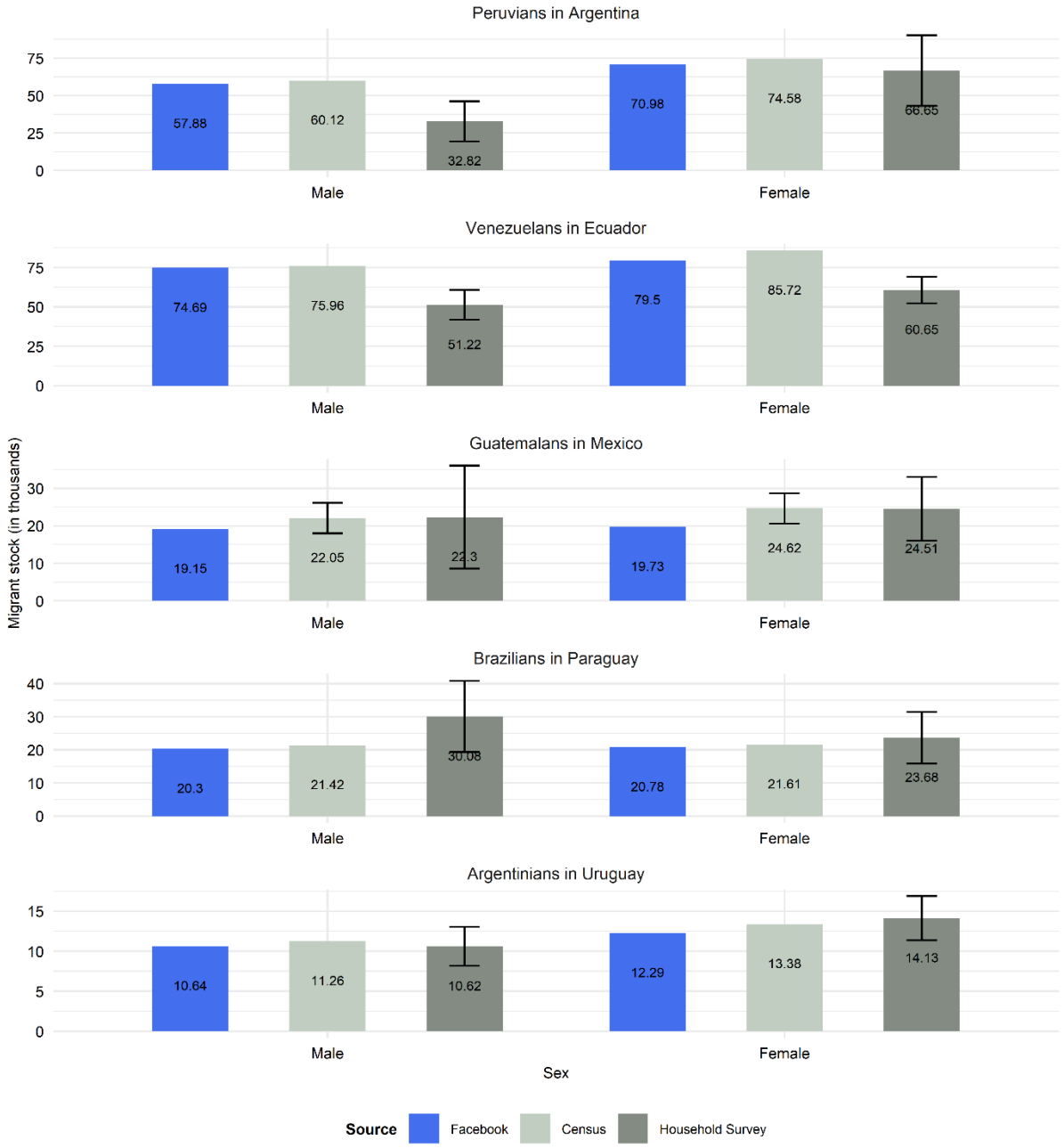
**Figure 3.** Number of people/users born in/previously living in Venezuela according to different data sources. Argentina, Ecuador, Mexico, Paraguay and Uruguay, circa 2020




Note: Here, *Facebook* refers to all Meta social media platforms included in the analysis (Instagram, Facebook, and Messenger). We include estimates of the adjusted total population, which incorporate corrections for census omissions using administrative data imputation procedures (as in Uruguay) or demographic estimation adjustments (as in Ecuador). These adjusted estimates are used only in this figure and not in the multivariate regressions presented later in the paper.

Source: own elaboration based on Facebook API extracts, HH survey and census data.

**Figure 4.** Number of people/users born in/previously living in selected origins according to different data sources. Selection of largest foreign-born communities in Argentina, Ecuador, Mexico, Paraguay and Uruguay, circa 2020




Note: Here, *Facebook* refers to all Meta social media platforms included in the analysis (Instagram, Facebook, and Messenger). These origins were selected as they represent the largest foreign-born group selected destination.
Source: own elaboration based on Facebook API extracts, HH survey and census data.

In the multivariate analysis, the estimate of Meta daily users emerges as a strong and consistent predictor of both census- and household-survey-based population measures across all model specifications, using the total population captured either by the census or by household surveys. The bivariate association between Meta users and census counts (Model 1) is highly significant and close to one-to-one ($\beta \approx 1.06$, $p < 0.001$, $R^2 = 0.99$). When adding controls for country and sex (Models 2–3), the coefficient attenuates but remains positive and marginally significant ($\beta \approx 0.39$ in M2), indicating that a large share of variance in census totals is explained by Meta data alone. Once Meta

app penetration is introduced (Model 4), the Meta coefficient becomes non-significant, suggesting that penetration captures the same underlying factor of population size and digital reach that Facebook users reflect. Despite this, model fit remains very high, showing that the model still reproduces almost all cross-country variation in total population. A similar pattern appears for household-survey totals (Models 5–8). The Meta coefficient starts high ($\beta \approx 0.95$, $p < 0.001$), weakens with the addition of country and sex controls, and regains moderate significance once penetration is included (M8 $\beta \approx 0.33$, $p < 0.05$).

For migrant populations, Meta users also significantly predict both census and survey migrant stocks, though with smaller coefficients and more realistic fit levels. The census-based models (M1–M5) incorporate destination- and origin-country effects, controlling for both the structure of receiving contexts and the composition of sending populations. Across these models, the Meta coefficient remains strongly significant ($\beta \approx 0.37$–$0.73$, $p < 0.001$), while adjusted $R^2$ rises from 0.55 to 0.76. The household-survey models (M6–M10) maintain a positive and significant Meta effect in all cases ($\beta \approx 0.62$–$1.29$, $p < 0.001$), with model fit between 0.40 and 0.60. Notably, these migrant-stock regressions include Meta penetration rates at both origin and destination, allowing separate tests of coverage and selection mechanisms. Neither variable undermines the Meta coefficient, indicating that the association between Meta users and enumerated migrant populations is not simply driven by general platform reach at either end of the migration corridor.

When comparing model performance across data sources, Meta user estimates align more tightly with census-based measures than with those derived from household surveys. In the total-population regressions, the census models display almost perfect fit (Adj. $R^2 \approx 0.99$–$1.00$) and coefficients close to unity, while the survey-based models show slightly lower explanatory power and more variation across specifications. The same pattern holds for migrant-stock analyses, where census models (Adj. $R^2 \approx 0.55$–$0.76$) outperform survey models (Adj. $R^2 \approx 0.40$–$0.60$). These differences suggest that Meta data track census outcomes more closely than household-survey outcomes, consistent with the census's comprehensive enumeration and the platform's population-level coverage.

**Table 4.** Model fit summary for Total Population

| Model | Dependent Variable | Controls Included | Meta Coef. (β) | Adj. $R^2$ | AIC | BIC |
|---|---|---|---|---|---|---|
| M1 | log (Census total population) | - | 1.061 *** | 0.986 | -5.54 | -4.94 |
| M2 | | Destination | 0.394 * | 1.000 | -39.38 | -37.56 |
| M3 | | + Sex | 0.082 | 1.000 | -44.21 | -42.09 |
| M4 | | + Meta penetration | 0.408 | 1.000 | -49.12 | -46.70 |
| M5 | log (HH survey total population) | - | 0.954 *** | 0.934 | 8.06 | 8.67 |
| M6 | | Destination | 0.281 | 0.999 | -35.91 | -34.09 |
| M7 | | + Sex | -0.138 | 1.000 | -43.81 | -41.69 |
| M8 | | + Meta penetration | 0.327 * | 1.000 | -78.24 | -75.82 |

Notes: All models estimated by OLS with logs of variables. *p < 0.05 *, **p < 0.01 **, ***p < 0.001 ***.

**Table 5.** Model fit summary for Migrant Stocks

| Model | Dependent Variable | Controls Included | Meta Coef. (β) | Adj. $R^2$ | AIC | BIC |
|---|---|---|---|---|---|---|
| M1 | log (Census migrant population) | - | 0.727 *** | 0.55 | 492.0 | 497.7 |
| M2 | | Destination | 0.659 *** | 0.60 | 481.0 | 498.1 |
| M3 | | + Origin | 0.368 *** | 0.75 | 430.4 | 481.7 |
| M4 | | + Sex | 0.369 *** | 0.75 | 432.3 | 486.5 |
| M5 | | + Meta penetration (origin + dest.) | 0.709 *** | 0.76 | 321.4 | 367.9 |
| M6 | log (HH survey migrant population) | - | 0.993 *** | 0.40 | 649.9 | 655.6 |
| M7 | | Destination | 0.920 *** | 0.48 | 635.1 | 652.2 |
| M8 | | + Origin | 0.625 *** | 0.56 | 623.8 | 675.1 |
| M9 | | + Sex | 0.622 *** | 0.56 | 625.4 | 679.6 |
| M10 | | + Meta penetration (origin + dest.) | 1.290 *** | 0.60 | 468.5 | 515.1 |

Notes: All models estimated by OLS with logs of variables. *p < 0.05 *, **p < 0.01 **, ***p < 0.001 ***.

**Conclusions and further work**

Overall, the analysis demonstrates that Meta data provide a reliable adjustment for measuring migration, performing comparably—and in some cases better—than traditional data sources in capturing migrant stocks. The alignment with census figures is consistently stronger than with household surveys, both for total populations and for migrant populations, reflecting the platform's broad coverage and the census's more exhaustive enumeration. While household surveys show greater divergence, especially for smaller origin groups, Meta estimates reproduce major population patterns with notable precision. This also suggests that Meta data could serve as a better input for monitoring the magnitude of migrant populations than household surveys—which, although not originally designed for that purpose, have often been used for migration estimates in the region during the inter-census period due to the lack of alternative sources. Importantly, the results underscore that even censuses and surveys, traditionally treated as "gold standards," display growing limitations: higher omission rates, inconsistencies across instruments, and increased reliance on imputation procedures. In this context, Meta data cannot replace traditional statistics but can complement them by providing timely, flexible, and scalable indicators of migration trends—particularly valuable in regions where conventional data collection continues to face significant challenges.

Further work will include a more detailed analysis of cross-country differences by origin and destination, as well as a disaggregated examination of coefficients by sex and Meta penetration rates to deepen understanding of the underlying demographic and digital factors shaping these relationships.

**References**

Del Popolo, F. (2024). Breve panorama de los censos de población y vivienda 2020 en América Latina y el Caribe y principales desafíos de cara a la ronda. Presentation, Aug 27 2024. Available at 2030https://www.cepal.org/sites/default/files/presentations/panorama_censos_2020_desafios_ronda-2030_cepal-celade_28ag.pdf

Gutiérrez, A., Mancero, X., Fuentes, A., López, F., & Molina, F. (2020). Criterios de calidad en la estimación de indicadores a partir de encuestas de hogares: una aplicación a la migración internacional. Serie Estudios Estadísticos, CEPAL, no. 101.

Montiel, C. (2024). Facebook versus encuestas de hogares: oportunidades y límites de los datos de Facebook para el estudio de la migración internacional en América Latina. Tesis de Maestría en Demografía y Estudios de Población.

Palotti J, Adler N, Morales-Guzman A, Villaveces J, Sekara V, Garcia Herranz M, Al-Asad M, Weber I. (2020). Monitoring of the Venezuelan exodus through Facebook's advertising platform. PLoS One. 2020 Feb 21;15(2): e0229175.

Spyratos S, Vespe M, Natale F, Weber I, Zagheni E, Rango M. (2019). Quantifying international human mobility patterns using Facebook Network data. PLoS One. 2019 Oct 24;14(10): e0224134.

Varona, T., Masferrer, C., Prieto Rosas, V., & Pedemonte, M. (2024). Which definition of migration better fits Facebook 'expats'? A response using Mexican census data. *Demographic Research,* 50(39), 1171-1184.

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(4), 721-734.

## Appendix

**Table A1.** Overview of Census Design, Type, and Data Collection Methods by Country

| Country | Census Design | Census Type | Data collection method | | | |
|---|---|---|---|---|---|---|
| | | | Digital census (CAWI) | Manual data entry (Paper) | Manual data entry (CAPI device) | Manual data entry (CATI device) |
| Argentina | Traditional census | De jure census | ✓ (50.3%) | ✓ | – | – |
| Ecuador | Traditional census | De jure census | ✓ (13.6%) | ✓ | ✓ | |
| Mexico | Traditional census | De jure census | ✓ (0.3%)* | ✓ (Alternative because COVID-19 pandemic; uptake was minimal) | ✓ | ✓ |
| Paraguay | Traditional census | De facto census | – | ✓ | – | – |
| Uruguay | Combined census | De jure census | ✓ (60%) | – | ✓ | ✓ |

Notes: 1) Definition of Census design: Traditional census: Data are collected directly from the population through field enumeration (face-to-face or self-administered) and Combined census: A hybrid approach in which field enumeration is complemented by administrative records to achieve full population coverage. 2) CAWI = Computer-Assisted Web Interviewing; CAPI = Computer-Assisted Personal Interviewing; CATI = Computer-Assisted Telephone Interviewing. 3) Figures in parentheses indicate the officially reported percentage of persons who completed the census using the CAWI modality in each country. For México is own estimation in base of information on: https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825197537.pdf

## Table A2. Census

| Country | e-Census Modey and Implementation Details | Administrative data Implementation Details | Census total Undercount (%) |
|---|---|---|---|
| Argentina | The online self-enumeration was available from March 16 to May 15, 2022, closing just before the in-person enumeration phase, which took place on May 18 of the same year (INDEC, 2022) | – | No official data available |
| Ecuador | The web self-enumeration application was used for approximately 10% of households, primarily targeting public servants and officials from various institutions. | Administrative records were integrated to validate and improve census data quality by analyzing consistency and reducing non-response. Sources included the Civil Registry, Ecuadorian Social Security Institute, Ministry of Public Health, Secretariat of Higher Education, Science, Technology and Innovation, and Ministry of Economy and Finance. These processes supported the correction of coverage gaps and the enhancement of demographic and geographic information. | 4.2% |
| Mexico | Administrative records were used to address underreporting of young children—a known issue in population censuses. A consistency analysis between the census and administrative sources led to the imputation of | To address the known undercount of young children, administrative records were used to impute children under the age of seven in households where women of reproductive age reported surviving children not listed as usual residents. | No official data available |

| | | | |
|---|---|---|---|
| | children under seven years of age in households where women of reproductive age reported surviving children who were not listed as usual residents. | Additionally, official documentation reports the imputation of data in pending households, accounting for approximately 5% of the total population. | |
| Paraguay | – | – | No official data available |
| Uruguay | The web self-enumeration was implemented in two phases: an initial two-month period allowing only digital self-enumeration, followed by a second phase with census enumerators visiting households using mobile devices. Due to coverage issues, the digital census period was reopened in September 2023, enabling the web enumeration rate to reach the final reported value (INE, 2023). | The resulting microdata combines information collected in the field with administrative data, which were processed and validated through the Integrated System of Statistical Registers and Surveys (SIREE). This approach incorporated real individuals from administrative sources to fill gaps left by direct enumeration, avoiding synthetic imputation. | 10.3% |