



# EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS DEMOGRÁFICOS, URBANOS Y AMBIENTALES

ANÁLISIS DE LA VALIDEZ DE DATOS DE FACEBOOK PARA EL ESTUDIO  
DE LA INMIGRACIÓN EN MÉXICO

Tesis presentada por  
TANIA DANIELA VARONA MOLINA

Para optar por el grado de  
MAESTRA EN DEMOGRAFÍA

Directoras de tesis  
CLAUDIA PATRICIA MASFERRER LEÓN  
VICTORIA PRIETO ROSAS

CIUDAD DE MÉXICO, OCTUBRE DE 2022

*A Miguelito y Clau.*

## **AGRADECIMIENTOS**

Esta investigación fue posible gracias a la contribución de profesoras, profesores y de instituciones que me apoyaron de distintos modos. Estoy sumamente agradecida con todos y además quisiera dar un agradecimiento especial a:

Claudia Masferrer y Victoria Prieto, codirectoras de esta tesis. El apoyo riguroso y continuo fue fundamental en la elaboración de cada apartado. Además de haber tenido su excelente asesoría en términos académicos, tuve la fortuna de contar con su consejo, empatía y acompañamiento a nivel personal.

Jéssica Nájera, lectora de esta tesis. Las preguntas que me formuló para clarificar y mejorar mis argumentos; la lectura detallada y la orientación oportuna; y las discusiones distensionadas, sin duda, mejoraron cada apartado.

Martín Pedemonte, del Instituto de Computación de la Facultad de Ingeniería de la Universidad de la República (Uruguay), por su amable asesoría y por realizar los extractos de población total y población expatriada de los datos de Facebook en marzo de 2020, que son el insumo esencial para la realización de esta tesis.

Consejo Nacional de Ciencia y Tecnología (CONACYT) del Gobierno de México, por otorgarme una beca que me permitió dedicarme a tiempo completo a mis estudios de Maestría. Al Colegio de México y al Centro de Estudios Demográficos, Urbanos y Ambientales (CEDUA) por su acompañamiento.

También siento un profundo agradecimiento hacia mis amigas, amigos y familia, quienes son parte fundamental de mi vida. Así, quisiera agradecer con todo mi cariño a:

Anna, Ajpub, Julio, Itzel, Penélope, Mariana, Elder, Dayana, Dileri y Karen; amigas y amigos con quien tuve la fortuna de cursar la Maestría en Demografía 2020 – 2022. Personas maravillosas con quien compartí inolvidables experiencias que hicieron de estos dos años de maestría una experiencia digna de recordar con cariño.

Paola, Andrea J. y Luis, por su amistad sincera y cariñosa a través de tantos años. A mis cómplices Lina y Andrea G., por estar siempre presentes y ser inspiración. A Sharly por las conversaciones estimulantes y por apoyarme en mejorar la redacción de esta tesis. A Paco por ser luz y acompañarme en este camino.

Mi abuelo Carlos Arturo; a mis hermanas y hermano Mafe, Pipe y Pau; a mi sobrina Antonia y a mi sobrino Alejandro. A mi papá, Miguel y a mi mamá, Claudia, para quienes mi agradecimiento es inconmensurable, pues siempre me han impulsado a ser valiente, a afrontar la vida con paciencia y con amor y, a experimentar todo aquello que me hace feliz. A ellos, dedico esta tesis.

## RESUMEN

La medición de la migración internacional es de crucial importancia para comprender la dinámica demográfica de una población. No obstante, regularmente esta medición se enfrenta a la escasez de datos oportunos en el tiempo, por tanto surge la necesidad de aprovechar fuentes de información innovadoras que subsanen este déficit. Facebook Advertising es una fuente potencial de datos demográficos en tiempo real. Sin embargo, sus usuarios, no necesariamente son representativos de la población que reside en un país. Considerando esto, la presente investigación propone evaluar los sesgos y la validez de los datos extraídos de la API de Facebook para cuantificar la inmigración en México. Para este objetivo, se estimaron stocks de inmigración por sexo, edad y país de origen, y se contrastaron con seis definiciones de inmigración basadas en el Censo de Población y Vivienda 2020 de México, a través de 54 modelos de regresión.

Los resultados indican que, de las seis definiciones contrastadas, los datos de Facebook tienen mayor proximidad a la definición de inmigración basada en el país de residencia previa. Adicionalmente, la evaluación de sesgos por grupos etarios muestra que la población inmigrante de 25 a 54 años está subrepresentada, mientras que la de 55 a 65 años está sobrerrepresentada en Facebook. En cuanto a la dispersión de los grupos etarios, se tiene que el grupo de 25 a 54 años es el que muestra menor varianza, cuando se contrastan las dos fuentes de información empleadas. Asimismo, se identificaron sesgos de subestimación de población migrante proveniente de Cuba, El Salvador, Guatemala, Honduras y Venezuela. Con respecto a la variable sexo, no se encontró evidencia estadística suficiente para identificar sesgos. Sin embargo, al estimar modelos desagregados por esta variable, se encontró que los modelos calculados para los hombres indican un mayor poder de explicación en comparación con los de las mujeres.

Estos resultados evidencian la oportunidad de utilizar datos de Facebook para obtener stocks de inmigración desagregados por edad, sexo y país de origen en México, a reserva de sus sesgos intrínsecos. En este sentido, red social en línea podría aprovecharse como una fuente de información autónoma, que permite cuantificar stocks de inmigración en tiempo real. Esto sugiere una alternativa de medición en periodos intercensales o en coyunturas de fenómenos migratorios impulsados por situaciones imprevistas.

# CONTENIDO

Introducción .....	7
Capítulo I: Perspectiva de análisis y marco conceptual.....	11
1.1 Big data, demografía digital y migración internacional.....	12
1.2 Migración digital.....	15
1.3 Medición de la inmigración .....	18
1.3.1 Fuentes de información tradicional.....	18
1.3.2 Fuentes de información innovadoras .....	22
Capítulo II: Estrategia metodológica .....	26
2.1 Datos .....	28
2.1.1 Fuentes de información.....	28
2.2 Indicadores .....	30
2.2.1 Características demográficas y definiciones de inmigración.....	30
2.2.2 Proporción de Acceso a Internet .....	33
2.2.3 Razón de acceso a Facebook.....	34
2.3 Métodos.....	35
2.3.1 Formalización de modelos lineales para la evaluación de sesgos.....	35
2.3.2 Formalización de modelos lineales para la evaluación de sesgos desagregados por sexo .....	37
Capítulo III: Resultados .....	38
3.1 Estructura poblacional y contexto de la inmigración en México.....	41
3.2 Acceso a internet y a Facebook de la población inmigrante en México.....	47
3.3 Análisis de sesgos de los datos de Facebook .....	53
3.3.1 Evaluación gráfica .....	53
3.3.2 Evaluación analítica .....	60
Capítulo IV: Discusión .....	65
Conclusiones.....	72
Referencias.....	75
Apéndices.....	78
Apéndice 1. Razones de acceso a Facebook.....	79
Apéndice 2. Resumen de los modelos estadísticos.....	81

## INTRODUCCIÓN

La migración internacional es el componente demográfico más complejo de medir oportunamente en el tiempo. El número de migrantes internacionales presenta una tendencia de crecimiento a nivel global, con 153 millones en 1990 y 281 millones de migrantes internacionales en 2020 (OIM, 2022), esto representa un aumento de más del 80 % en 30 años. Sin embargo, pese a su gran volumen e importancia, a efectos del cambio en la estructura y composición demográfica de las poblaciones, es difícil cuantificar de forma oportuna su magnitud. Esto radica en que los censos de población y vivienda, que son los instrumentos por excelencia utilizados para cuantificar la inmigración en un país, se realizan en promedio cada diez años. Por consiguiente, surge la necesidad de solventar esta dificultad mediante mecanismos innovadores para la medición de la migración internacional en periodos intercensales.

El desarrollo de la Web 2.0 puso en el centro a los usuarios de Internet para promover la conexión social, el intercambio de información y la creación de contenido de los individuos, desde principios de este siglo (Wilson et al., 2011). Este enfoque, basado en los usuarios, hizo que las redes sociales en línea y los motores de búsqueda sean hoy en día los servicios más utilizados de Internet. En consecuencia, se ha generado una gran cantidad de datos que reflejan aspectos reales de las sociedades (Panagiotou et al., 2016). Estos datos se traducen en huellas digitales que permiten aproximarse a los fenómenos demográficos como el de la migración internacional; y a su vez permite que redes sociales en línea, como Facebook se conviertan en una fuente de información capaz de mitigar la escasez de datos actualizados y detallados para describir el volumen y características de las poblaciones migrantes en tiempo real. La necesidad de disponer de fuentes de información con dichas características se hizo particularmente evidente para el año 2020, cuando la pandemia por el SARS-COV-2 postergó la realización de la ronda de censos 2020 en la mayoría de los países de América Latina.

Facebook es la red social en línea más utilizada del mundo. De acuerdo con Statista (2021), la tasa de penetración de Facebook en Estados Unidos en 2021 fue de 71.3 %, en Europa de 62.4 % y en América Latina y el Caribe de 69.5 %. Estas cifras reflejan que su uso generalizado y el volumen de población que utiliza esta red social es tan vasto en la actualidad, que ha llevado a que algunos autores equiparen a Facebook con un Censo Digital (Cesare et al., 2018; Rampazzo et al., 2021; Zagheni et al., 2017), pues además permite conocer características demográficas básicas de

las personas. En Europa y Estados Unidos se han realizado estudios basados en la premisa de que Facebook se aproxima a un censo digital. Su objetivo ha sido medir los sesgos de esta red social con respecto a fuentes de información tradicionales, como las encuestas y estimaciones de población producidas por Naciones Unidas, desde aproximaciones estadísticas inferenciales y bayesianas (Alexander et al., 2020; Spyrtatos et al., 2019; Zagheni et al., 2017). Sin embargo, los estudios para América Latina son escasos.

La revisión de antecedentes de esta investigación identificó dos estudios seminales: el de Spyrtatos et al. (2019) y el de Palotti et al. (2020), los cuales se centran en estimar stocks de inmigración con datos de Facebook. De acuerdo con las estimaciones de los autores de estos dos estudios, el volumen de población venezolana es superior al que reportan las fuentes de información oficiales de Colombia como el Registro Administrativo de Migrantes Venezolanos en Colombia (RAMV) de 2018 o los registros de inmigración que publica periódicamente Migración Colombia, y pone de manifiesto la dificultad a la que se enfrentan los países de acogida para cuantificar la inmigración con las características particulares del proceso de emigración que se vive en Venezuela. Sin embargo, en ningún caso, los datos empleados para validar las cifras de usuarios de Facebook a los que se atribuye la condición de migrante provienen de censos de población, a pesar de que esta es la fuente que mejor permite captar la magnitud de las poblaciones migrantes. Una de las razones por las cuales no se ha hecho esta comparación, obedece a que la literatura centrada en estas comparaciones es bastante reciente, por lo que no se había tenido la coincidencia de obtener datos de un Censo y de Facebook en un mismo momento.

En marzo de 2020, México realizó el último Censo de Población y Vivienda, en el que empadronó un total de 126,014,024 millones de habitantes en el territorio mexicano. Es en virtud de la existencia de datos censales recientes y la disponibilidad de datos sobre usuarios de la red social Facebook para este mismo país y la misma fecha, que surge el interés de cotejar ambas fuentes de información y explorar la validez de esta segunda fuente para analizar la inmigración en México. Este contexto motivó la realización de esta tesis, cuyos objetivos de investigación son: primero, comprender en qué medida el uso de internet y Facebook son recursos generalizados en las poblaciones migrantes de México; y segundo, analizar la validez de la definición de la variable *expatriado* de Facebook y el alcance de esta red social para captar y medir poblaciones migrantes en tiempo real. Este segundo objetivo permitirá comprender en qué medida, en los periodos

intercensales, sería posible confiar en las cifras de expatriados como variable proxy de poblaciones migrantes. Asimismo, se busca comprender cuál es la definición de inmigración a la que más se aproximan los datos de Facebook; al mismo tiempo se evalúa la dirección y magnitud de los sesgos en términos de edad, sexo y país de origen. De este modo, se presenta por primera vez la posibilidad de comparar las estimaciones de una fuente de información *estándar de oro* para la medición de la migración internacional con datos de Facebook para el caso mexicano.

Para captar la migración por medio de Facebook, se emplea la variable expatriado que se define como aquella persona que habitualmente vivía en un país (por ejemplo, Venezuela) y ahora vive en el extranjero (por ejemplo, en México), lo que configura la variable “Vivió en Venezuela. Antiguos expatriados – Venezuela”. En sentido estricto, el concepto de expatriado no capta la migración internacional, pero es una variable proxy que permite acercarse a esta. De acuerdo con la literatura disponible hasta la fecha, se desconoce el método y los criterios usados por esta red para designar a los usuarios con esta etiqueta. Sin embargo, se deduce que hay dos elementos que desempeñan un papel clave en la identificación de los expatriados. El primero es la autodeclaración de la *ciudad actual* y la *ciudad natal* de los propios usuarios dentro del listado de *lugares donde has vivido* en el perfil personal. El segundo elemento es la estructura de la red de amistades; por ejemplo, tener al menos dos amigos de Facebook en el país de origen y dos amigos de Facebook en el país de destino (Spyratos et al., 2019; Zagheni et al., 2017).

Para entender a qué definición de inmigración se aproxima más la categoría expatriado, en este estudio se seleccionaron seis definiciones de inmigración, las cuales se configuran con base en las definiciones clásicas de la condición migratoria según *lugar de nacimiento* y *lugar de residencia anterior en una fecha de referencia*. Las cuatro definiciones adicionales se realizaron considerando algunas interacciones entre las dos primeras definiciones, de las que se obtienen: inmigrantes recientes por lugar de nacimiento, inmigrantes que en 2015 residían en el país de nacimiento, inmigrantes que en 2015 residían en un país distinto del de nacimiento y población mexicana retornada.

De acuerdo con la definición de expatriado de Facebook y la información del Censo de población, las características demográficas que se evaluaron de la población extranjera fueron: sexo, tres grupos etarios (13 a 24, 25 a 54, 66 a 65), y trece países de origen (Argentina, Brasil, Chile, Colombia, Cuba, El Salvador, Guatemala, Haití, Honduras, Nicaragua, Perú, República

Dominicana, y Venezuela). Con base en estas características, se analizó la proporción de expatriados en el total de usuarios de Facebook y la proporción de inmigrantes en el total de población en México. Posteriormente se estimaron modelos de regresión lineal bivariados y multivariados para evaluar los sesgos de las mediciones provenientes de datos de Facebook por sexo, edad y país de origen.

Los resultados de esta investigación, además de aportar una evaluación sistemática de los sesgos que se presentan en una red social como Facebook para estimar stocks de inmigración, ofrecen una serie de coeficientes que permitirán estimar en tiempo real el volumen de la población inmigrante en México con la información del número de usuarios de esta red social. Asimismo, estos resultados contribuyen con la escasa acumulación de estudios de migración digital desarrollados para América Latina, que permiten seguir discutiendo los beneficios y los retos de emplear fuentes de información de *big data* en la medición de los procesos demográficos.

**CAPÍTULO I:**  
**PERSPECTIVA DE ANÁLISIS Y MARCO CONCEPTUAL**

## 1.1 BIG DATA, DEMOGRAFÍA DIGITAL Y MIGRACIÓN INTERNACIONAL

El paso de la era análoga a la era digital representó oportunidades y retos sin precedentes en el campo de los datos y de la investigación en ciencias sociales, y en consecuencia en la demografía. En la actualidad, gran parte de las interacciones humanas que décadas atrás ocurrían principalmente en la realidad material, se dan también en el campo digital. Este hecho expande la investigación de los fenómenos demográficos y de las relaciones humanas al menos en un escenario adicional: la virtualidad. También permite la formulación o reformulación de preguntas de investigación según fuentes de información novedosas, basadas en desarrollos tecnológicos innovadores y masivos.

La era digital propició un avance vertiginoso en la producción de datos, así como en los formatos en los que esta información se captura, se analiza y se difunde. Hoy en día, mejor conocido como el *big data*<sup>1</sup>, este conjunto de datos complejos comprende una variedad de conceptos. La definición comúnmente aceptada está compuesta por las “3V’s”: volumen, variedad y velocidad (Chen et al., 2013; Salganik, 2018). En este sentido, *big data* es una fuente de información basada en datos masivos, cuyo volumen en ocasiones puede superar la capacidad de procesamiento computacional tradicional; en paralelo, presenta una gran variedad de formatos de información: imágenes satelitales, interacciones en Internet, y/o registros biométricos en aplicaciones móviles, por mencionar algunos. La velocidad es la frecuencia de mayor rapidez con la que se producen los datos, respecto de las fuentes de información tradicionales.

De acuerdo con Salganik (2018), el *big data* tiene tres características fundamentales que ayudan a la investigación en ciencias sociales. En primer lugar, estos datos son grandes (*big*), esto permite el estudio de eventos raros, la estimación de heterogeneidad, y la detección de pequeñas diferencias que se hacen visibles luego de trabajar con datos masivos. En segundo lugar, estos datos siempre están disponibles (*always-on*), por eso se pueden medir eventos inesperados en tiempo real (*nowcasting*). Gracias a que los sistemas de *big data* están constantemente recolectando información, de igual modo es posible obtener conjuntos de datos longitudinales, que

---

<sup>1</sup> Al considerar el amplio uso en las áreas a las que apunta esta investigación, se empleará la definición de *big data*, del Diccionario Panhispánico del Español Jurídico: “conjunto de técnicas que permiten analizar, procesar y gestionar conjuntos de datos extremadamente grandes que pueden ser analizados informáticamente para revelar patrones, tendencias y asociaciones, especialmente en relación con la conducta humana y las interacciones de los usuarios.” (Diccionario Panhispánico del Español Jurídico, 2022).

permiten estudios *ex ante*, *durante* y *ex post* de un evento. Finalmente, hay menor probabilidad de un cambio de comportamiento en los sujetos observados (*non reactive*) cuando se capte información sobre eventos de cierta sensibilidad social (como el racismo), ya que las personas no siempre son conscientes del tipo de información que comparten en redes sociales. Sin embargo, estas ventajas deben evaluarse a la luz de una serie de retos metodológicos para las ciencias sociales, entre ellos, según Salganik (2018):

1. Información incompleta: fuentes de información que no han sido creadas para fines académicos, por lo que dependerá de las preguntas de investigación planteadas, si es posible abordar conceptos teóricos con el *big data*. Al plantearse preguntas relacionadas con conceptos más abstractos como inteligencia, capital social o democracia, posiblemente los datos disponibles no sean lo suficientemente completos para abordar estos conceptos. Las soluciones que se ofrecen a estos casos son tres: primero, recolectar la información mediante encuestas; segundo, imputar datos para inferir características de la información disponibles y completarla; y tercero, combinar distintas fuentes de información.
2. Limitaciones en el acceso a la información: incidencia de acuerdos de confidencialidad en la mayoría de los datos, y que provienen de empresas privadas o gobiernos que deben mantener la información anonimizada.
3. Datos no representativos: insuficiencia numérica de la muestra debido a cómo se captura la información o se tiene acceso a esta. Si este es el caso, se debe tener precaución para no generalizar hallazgos o conclusiones con base en información no representativa.
4. Cambios en los datos sin planificación: variabilidad en las fuentes de *big data*; muchas están a la deriva porque suceden cambios en, al menos, tres vías: primero, cambios en *quién* usa los recursos web; segundo, cambios en *cómo* se usan estos recursos (cambios comportamentales); y tercero, cambios en los recursos web en sí mismos, como los cambios que ha implementado Facebook a lo largo del tiempo con cada actualización. En este sentido, la comparabilidad de la información en el tiempo puede ser difícil o incluso incorrecta.
5. Incertidumbre algorítmica: desconocimiento de los criterios que constituyen los algoritmos; como los sistemas que captan la información están diseñados para promover conductas específicas en los usuarios que interactúan en una determinada red social, la confiabilidad de los datos se pone en cuestión.

6. Información sucia: fuentes de *big data* que nos han sido procesadas adecuadamente según una perspectiva de análisis en particular.
7. Información sensible: datos personales y características específicas de las personas; en la mayoría de los casos las personas no son conscientes de la información que comparten, por esto quienes recolecten y procesen datos puedan acceder a información sensible. Si por algún motivo, esta no es anónima, es posible rastrear a personas específicas que pueden ser vulneradas en razón de sexo, religión, gustos y comportamientos específicos.

Una vez presentadas estas características, es necesario establecer estrategias para trabajar con este tipo de datos. De acuerdo con Salganik (2018), hay al menos tres formas de abordar preguntas de investigación pertinentes e interesantes de estos datos: (1) contrastar cuantitativamente dos teorías existentes gracias al tamaño de los datos; (2) hacer predicciones de corto plazo (*nowcasting*) con la información disponible de *big data*; y (3) realizar estimaciones sin la necesidad de experimentos en la realidad material. Así, esta investigación adopta la segunda estrategia en vista de su orientación al análisis del potencial de datos en una red social en línea, para hacer predicciones o estimaciones de corto plazo de poblaciones migrantes.

De la proliferación de *big data* ha surgido un campo de análisis dentro de la Demografía conocido como *demografía digital*; este emplea una variedad de datos digitales extraídos en su mayoría de la web, para responder antiguas y nuevas preguntas sobre demografía. Este campo de análisis permite hacer frente a algunos de los principales obstáculos en cuanto a la disponibilidad de información y su desagregación a nivel individual respecto de la edad, el sexo, la educación, las relaciones sociales, e intereses, entre otros (Weber & State, 2017). En paralelo, advierte y mide los sesgos de los datos digitales para producir estimaciones en tiempo real y con gran granularidad<sup>2</sup>.

Conforme a los objetivos de esta tesis, los antecedentes se enfocarán en investigaciones relevantes de la demografía digital relacionadas con la migración internacional, a sabiendas que este es uno de los campos de estudio que más se ha beneficiado con el uso de *big data*.

---

<sup>2</sup> Calco semántico del término *data granularity* en inglés, que hace referencia al detalle y la especificación de los datos. Un dato será más granular en la medida que esté más subdividido y sea más específico (Techopedia, 2022)

## 1.2 MIGRACIÓN DIGITAL

La migración internacional es uno de los componentes demográficos que presenta mayores dificultades para su medición. Esto se relaciona principalmente con tres dimensiones. La primera hace referencia al tipo de migración, es decir, si se trata de migración voluntaria o forzada, relativa a *los motivos de la migración*. La segunda dimensión se relaciona con la *temporalidad*, esto se traduce en migración reciente o de larga data. La última dimensión se refiere al *registro* de la migración, vinculado estrechamente con las dos dimensiones anteriores, pues depende del tipo de migración, así como de su temporalidad, la calidad del registro que se pueda obtener. Respecto del registro de la población migrante, sobresale el subregistro y/o la dificultad de contabilizar la inmigración irregular o forzada, cuando no se tiene la documentación necesaria para el ingreso legal; además, de la posible intención de la persona migrante de no ser visible ante los registros oficiales.

Así vemos que la medición del volumen o stock de inmigración se enfrenta a estas dificultades; además, según la coyuntura del país de destino y el fenómeno migratorio de estudio, la medición de los stocks con fuentes de información alternativas adquiere mayor o menor relevancia, pues son fuentes potencialmente fiables y en las que estas dificultades no aparecen.

De ello, un aspecto a resaltar es la disponibilidad de datos oportunos y desglosados por características de la población respecto de la migración de larga y corta duración a nivel mundial, pues esta sigue siendo muy limitada (Spyratos et al., 2019). En este contexto, por el creciente desarrollo de *big data* y su disponibilidad en Internet, la investigación sobre migración internacional se ha concentrado en estudiar fuentes de información digitales que amplían los recursos de análisis, para realizar estimaciones del fenómeno migratorio a modo complementario o alternativo (Zagheni et al., 2017).

Para subsanar la información faltante, en demografía digital se ha adoptado una estrategia de estudio de la migración internacional a través de huellas digitales. Desde 2010 algunos estudios empezaron a incorporar datos digitales; como es el caso de la investigación de Zagheni y Weber (2012) en la que se calcularon tasas de migración por edad y sexo utilizando datos extraídos de una muestra de mensajes de correo electrónico de Yahoo! En este trabajo pionero, los autores

estimaron las tasas específicas de migración internacional a través de la localización geográfica de los usuarios según sus direcciones IP<sup>3</sup>.

El estudio de Zagheni y Weber (2012) abrió un camino en la investigación demográfica para incorporar fuentes de información anteriormente no utilizadas, y promover el desarrollo y adaptación de métodos estadísticos destinados a analizar la validez de la información derivada de *big data*. Así, además de las direcciones IP, nuevas fuentes de datos se han empleado para estudiar la migración internacional, como: los registros de datos de llamadas (CDR)<sup>4</sup>, datos de tiendas de *retail* o plataformas de redes sociales en línea como Facebook o LinkedIn, y redes de microblogueo como Twitter (Sîrbu et al., 2021).

Las investigaciones en el campo de la migración internacional con datos derivados de Internet se han concentrado principalmente en dos vertientes: por un lado, aquellas que realizan *mediciones* de sesgos entre una fuente tradicional y una digital, con el propósito de comprender el nivel de confiabilidad de una fuente de información de *big data*, para estimar stocks, flujos y hacer proyecciones de migración. Por otro lado, están otras investigaciones con fuentes de información derivadas de redes sociales en línea, que permiten identificar, por ejemplo, procesos de asimilación cultural. En suma, estos dos campos se concentran en realizar estimaciones estadísticamente robustas, y en utilizar *big data* para responder preguntas sobre relaciones sociales, políticas y económicas en pro de estudiar relaciones demográficas.

Esta tesis se sitúa en la primera vertiente, la de la medición demográfica. En este conjunto de investigaciones, se han empleado métodos estadísticos inferenciales y bayesianos. Desde la estadística inferencial se usan técnicas como la de *diferencias en diferencias* (Zagheni & Weber, 2015), en las que se asume el supuesto de que existen condiciones homogéneas de acceso a internet en los países de origen; que hay una proporción relativamente alta de personas que acceden a redes sociales en línea; y que los sesgos y los errores para estimar un modelo de regresión son aditivos. También se han utilizado *regresiones log – log* para seleccionar ponderaciones óptimas de los coeficientes que evalúan los sesgos derivados de los datos de redes sociales (Spyratos et al., 2019), mediante un proceso de iteración de estas regresiones. Asimismo, las regresiones *logit* para el modelamiento de sesgos y la estimación de probabilidades de emigración (Hsiao et al., 2020); y

---

<sup>3</sup> Dirección de Protocolo de Internet, por sus siglas en inglés.

<sup>4</sup> Call Data Records, por sus siglas en inglés

las regresiones *lineales* o *log-lineales* para identificar sesgos específicos en características de las personas como la edad o el país de origen (Zagheni et al., 2017).

Del lado de la estadística bayesiana sobresalen dos trabajos: uno del contexto estadounidense (Alexander et al., 2020) y otro del contexto británico (Rampazzo et al., 2021). El objetivo principal de la investigación de Alexander es una proyección de corto plazo (*nowcasting*) de stocks de inmigración en Estados Unidos. Para esto, los autores estimaron los sesgos entre Facebook y la *American Community Survey* (ACS), y posteriormente aplicaron un modelo bayesiano jerárquico en el que integraron ambas fuentes de información para hacer las proyecciones. Resulta relevante que para realizar dichas proyecciones los autores utilizaron información agregada de varios años de la ACS como de los datos de Facebook.

Por su parte, en la investigación de Rampazzo el objetivo principal es comparar la estimación de stocks de inmigración de tres fuentes de información. Para esto, el stock de inmigrantes se estimó según datos de: (1) *Labour Force Survey* (LFS); (2) Facebook (teniendo en cuenta los cambios en el algoritmo que esta red social hizo para identificar expatriados entre 2018 y 2019); (3) un modelo basado en teorías migratorias, cuyas variables explicativas son: el tamaño de la población en el país de origen, la inmigración y emigración de países de Europa hacia Reino Unido, la tasa de crecimiento del PIB y la tasa de desempleo de los países de origen. Este proceder permitió a los autores comparar los stocks de población con base en tres fuentes de información y conocer en qué medida se asemejan las estimaciones obtenidas.

Las investigaciones presentadas hasta aquí ayudan a entender los métodos estadísticos necesarios para incorporar datos de redes sociales en línea que estiman stocks de inmigración. Asimismo, con base en las técnicas y fuentes de información tradicionales empleadas en estos estudios se observan medidas de ajuste globales significativas al comparar fuentes de información como Facebook y Twitter. Ello aporta evidencia empírica significativa para el desarrollo de medidas de migración según datos de redes sociales en línea.

### 1.3 MEDICIÓN DE LA INMIGRACIÓN

La medición de la migración internacional involucra criterios de tiempo, cruce de límites fronterizos y, de lugares de origen y destino. Según estos criterios, instrumentos de medición como los censos de población y las encuestas formulan las preguntas que captan a la población migrante a través de conceptos relacionados con el lugar de nacimiento y el lugar de residencia habitual. Con estas especificaciones se determina si una persona es inmigrante internacional desde la perspectiva del lugar de destino, pues es allí donde se visibiliza un esfuerzo de las autoridades por contabilizar la migración. Según el lugar de nacimiento y de residencia habitual, una persona inmigrante es quien “se traslada a un país distinto al de su residencia habitual durante un periodo de al menos un año” (UNDESA, 1998). Asimismo, el periodo de estancia determina si el inmigrante es de larga o de corta duración. Habitualmente, la inmigración de corta duración responde a periodos de un año o menos, y la inmigración de larga duración a periodos mayores a un año.

En lo que corresponde a las fuentes de información innovadoras o basadas en *big data*, el concepto de inmigración internacional tal como se mencionó arriba no existe, debido a que estos datos no están diseñados —hasta el momento— con el objetivo de captar a la población migrante. Sin embargo, debido a la información demográfica y de geolocalización que proporcionan los usuarios en redes sociales en línea, aplicaciones móviles y, en general, recursos digitales, es posible generar métricas que capturan procesos de movilidad y derivar de forma indirecta mediciones de migración internacional.

#### *1.3.1 Fuentes de información tradicional*

Las fuentes de información tradicionales para la medición de la inmigración son los censos de población y vivienda, las encuestas, y los registros administrativos producidos, gestionados y publicados por instituciones oficiales de los gobiernos nacional y local. También integran este grupo los datos y estimaciones que centralizan y difunden organismos internacionales como el Banco Mundial (BM), la Organización para la Cooperación y el Desarrollo Económicos (OCDE), la Comisión Económica para América Latina y el Caribe (CEPAL), ya que producen estimaciones o reúnen, armonizan y publican datos originalmente suministrados por los gobiernos de cada país.

### *Censos de población*

Los censos de población y vivienda son una fuente de información privilegiada para captar la migración, ya que mide el volumen de la población y sus características sociodemográficas. La población inmigrante se deriva de las preguntas que indagan por el lugar de nacimiento, lugar de residencia habitual de hace cinco años, lugar de residencia actual y nacionalidad; la recolección de información es periódica y se hace cada diez años (UNSTAT, 2017). Sobre la identificación de flujos migratorios, el censo resulta una fuente insuficiente, pues es difícil que se lleve un registro cotejado de los cambios de residencia respecto de la emisión y recepción de población migrante, esto limita la identificación del punto de partida y llegada como insumos mínimos para cuantificar el flujo migratorio.

No obstante, como cualquier fuente de información, los censos de población tienen fortalezas y debilidades para la medición de la migración internacional (UNSTAT, 2017). Entre las fortalezas, la principal radica en su cobertura universal, que además permite obtener información de población migrante sin documentación oficial. Asimismo, la desagregación por individuo: en términos demográficos (edad, sexo, lugar de residencia), en características socioeconómicas (educación, participación en el mercado de trabajo, entre otras), y en granularidad geográfica (gracias al registro poblacional a nivel municipal).

Por su parte, una potencial desventaja de los censos es su periodicidad de realización cada diez años. Esto da cabida a que no se registren, en un lapso de diez años, cambios relativos a los factores políticos, económicos y ambientales que motivan procesos de migración internacional. Adicionalmente, es factible que la población inmigrante no se capte con la misma probabilidad que la población nativa cuando se realizan cuestionarios ampliados del censo, cuya población objetivo se delimita por muestreo, debido a que, los migrantes suelen estar altamente concentrados en lugares específicos del territorio. Esto hace que no se logre profundizar en aspectos de interés de las poblaciones migrantes basados en una batería de preguntas más amplias contenidas en los cuestionarios ampliados.

### *Encuestas de hogares*

Las encuestas de hogares son instrumentos cuya unidad básica de medida es el hogar. De acuerdo con su propósito, estas pueden recabar información sobre dinámica demográfica, ocupación y empleo, uso de tecnología de información, y otros temas especializados, entre los que se encuentra la migración internacional. Su periodicidad varía en función del proyecto estadístico, por lo que podría ser mensual, trimestral, o sin una periodicidad definida. Las encuestas de hogares dirigidas directamente a población migrante permiten entender aspectos relacionados con las causas y consecuencias de la migración, así como la medición de flujos, entre otros aspectos relevantes. Asimismo, permiten tener información representativa de la población migrante.

No obstante, cuando las encuestas están dirigidas a una población más amplia y a temas especializados diversos, una de las principales dificultades de esta fuente de información, de acuerdo con Del Fava et al. (2019) y UNSTAT (2017), es la poca precisión estadística para alcanzar muestras representativas de población migrante, debido a que con frecuencia esta población es relativamente pequeña, y los tamaños de muestra no alcanzan a ser lo suficientemente grandes para captarla con precisión.

Una característica adicional (y potencial ventaja) en las encuestas de hogares, encuestas de personas y los censos de población, es que además de captar la población inmigrante que pasa periodos relativamente prolongados o indefinidos en el país de destino, estas fuentes de información pueden registrar población que no tenga autorización migratoria para su asentamiento en el país de destino (Castillo & Rojas Wiesner, 2020). Con estas fuentes de información, el objetivo de cuantificar y caracterizar la población inmigrante logra mayor alcance, en comparación a otras fuentes, como los registros administrativos y controles migratorios en frontera.

### *Registros administrativos*

Los registros administrativos comprenden la información que suministra la población inmigrante en instituciones oficiales del país de destino, entre las que se destacan oficinas de migración, centros de salud, centros educativos, o controles de paso en las fronteras. Esta información tiene un gran potencial para captar stocks de población y flujos migratorios, porque puede contener información sobre cambios de lugares de residencia de forma continua, a diferencia de los censos que son registros con una periodicidad más espaciada. Sin embargo, los registros hechos por parte de las autoridades migratorias al interior de un país dan cuenta exclusivamente de inmigración documentada o autorizada (Castillo & Rojas Wiesner, 2020). Así pues, en estos registros queda por fuera la población sin documentos migratorios que pueden estar asentadas en el país de destino. Adicionalmente, de acuerdo con Del Fava et al. (2019)), una potencial dificultad de estas fuentes de información es el subregistro de migrantes, principalmente en los procesos de emigración y retorno, puesto que los procesos inmigratorios suelen ser más rigurosos para permitir la estancia de población inmigrante, en comparación con la documentación que se diligencia en un proceso de emigración o retorno.

### 1.3.2 Fuentes de información innovadoras

Con la rápida expansión de Internet y de tecnologías de la información, hay mayor disponibilidad de fuentes *big data*. A diferencia de la información estadística tradicional, aquella derivada de plataformas digitales se genera pasivamente a través de *huellas digitales* de los usuarios. Eventualmente, depende de la plataforma si se puede obtener información respecto de procesos migratorios más oportunos en el tiempo y con mayor granularidad, que además facilite el diseño de políticas públicas y de la respuesta de emergencias en modos más eficientes (Bosco et al., 2022).

Entre las plataformas digitales de las que se puede obtener información de procesos migratorios o de refugio se encuentran: Facebook, Twitter, Instagram, Flickr, Reddit, LinkedIn, y Google. Adicionalmente, otras fuentes combinan información de Google con datos de pasajeros aéreos o servicios móviles de pago. También hay fuentes de información derivadas de los registros detallados de llamadas o de los registros extendidos de llamadas, y de imágenes satelitales (Bosco et al., 2022). Cada una de estas plataformas ofrece información distinta que, según el contexto, permitirá o no dar respuestas a las preguntas de investigación planteadas en torno a los procesos migratorios. En atención a los objetivos de esta tesis, se presentarán ahora las principales características de Facebook y cómo es posible identificar población migrante a través de esta red social. En ese sentido, la siguiente descripción no abunda en detalles sobre Facebook como red social, sino fuente de información para aproximarse a la migración internacional.

Así, Facebook Advertising<sup>5</sup> tiene como objetivo gestionar anuncios publicitarios. Para esto, la plataforma de marketing permite realizar segmentaciones de mercado a través de la definición de audiencias que cumplan con características específicas de: ubicación (país, entidad federativa, municipio, o una dirección específica por medio de coordenadas); edad (desde los 13 años y, hasta los 65 años y más); sexo; nivel educativo; industria o sector de trabajo; estado civil; intereses (deportes, tecnología, compras, comida y bebidas, entre otros); y variables de comportamiento relacionadas con temas de: movilidad internacional, compras, política, actividades digitales y viajes, entre otros. Es decir, esta plataforma contiene información

---

<sup>5</sup> La información presentada aquí sobre la segmentación de población que se puede hacer en Facebook se obtuvo de la exploración de la plataforma de Meta Business Suite, así como del sitio de internet de Meta for Developers, consultado en: <https://developers.facebook.com/docs/marketing-api/audiences/reference/basic-targeting>

demográfica de interés relacionada con edad, sexo y ubicación, y además permite identificar intereses de las personas que interactúan con el contenido de esta red social en línea.

Como se anticipó en la introducción, para captar la migración a través de Facebook, se emplea la variable *expatriado* definida como aquella persona que habitualmente vivía en un país (por ejemplo, Venezuela) y ahora vive en el extranjero (por ejemplo, México). Esto configura la variable “Vivió en Venezuela. Antiguos expatriados – Venezuela” según los datos de la plataforma. En sentido estricto, el concepto de expatriado no capta la migración internacional, pero es una variable proxy que permite acercarse a este análisis.

De acuerdo con la literatura disponible hasta la fecha, se desconocen el método y los criterios usados por esta red para designar a los usuarios con esta etiqueta. Sin embargo, es posible deducir tres elementos que desempeñan un papel clave en la identificación de los expatriados en Facebook. El primero es la autodeclaración de la *ciudad actual* y la *ciudad natal* de los propios usuarios dentro del listado de *lugares donde has vivido* en el perfil personal. El segundo elemento es la estructura de la red de amistades; por ejemplo, tener al menos dos amigos de Facebook en el país de origen y dos amigos de Facebook en el país de destino (Herdağdelen and Marelli 2017). El tercer elemento es la geolocalización de los usuarios. Esta se estima de acuerdo con una serie de factores como las direcciones IP de los usuarios y la autodeclaración de la ubicación geográfica (Rampazzo et al. 2021).

Una razón para utilizar Facebook en la medición de stocks de inmigración internacional responde a que sus tasas de penetración son relativamente altas. En América del Norte, esta tasa alcanza 71.3 %, en América Latina 69.5 % y en Europa 62.4 %<sup>6</sup>. Esto permite contar con un volumen de usuarios lo suficientemente amplio para aproximarse a la población total de un país o de una región; en ese sentido, Facebook podría aproximarse a un censo digital (Kashyap et al., 2020; Zagheni et al., 2017). Así, esta plataforma supondría una fuente de información idónea para analizar el volumen y la estructura de la población de forma oportuna en el tiempo, puesto que la información de expatriados y de los usuarios en general puede obtenerse diariamente.

Sin embargo, trabajar con datos de Facebook presenta algunas desventajas: el algoritmo que capta y configura la variable expatriado no es público, además de ser cambiante en el tiempo

---

<sup>6</sup> Consultado en: <https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/>

(Bosco et al., 2022; Spyrtos et al., 2019). Esto dificulta la comprensión a través del tiempo de las métricas que publica Facebook y supone una evaluación constante de los posibles criterios que se utilizan para obtener la variable expatriado. Además de los posibles sesgos que se generados del funcionamiento de Facebook, los usuarios también tienen comportamientos que producen sesgos. Rampazzo et al (2021), documentaron en su estudio que, de acuerdo con información publicada por Facebook, entre 2018 y 2019 se estimó que 11 % de las cuentas en esta plataforma estaban duplicadas y que 5 % eran falsas. La mayoría de estas anomalías se detectaron en el sudeste asiático.

### **Cuadro 1**

#### *Características de las fuentes de información tradicionales e innovadoras*

<b>Características</b>	<b>Tradicionales</b>	<b>Innovadoras</b>
Propósito de la recolección de los datos	Estadístico	Comercial y otros propósitos
Recolectados por	Profesionales entrenados	Dispositivos móviles, sensores, usuarios de aplicaciones de Internet
Metodología transparente	Siempre	Raramente
Heterogeneidad de los datos	Bajo	Alto
Representatividad demográfica	Alto	Puede ser bajo
Integridad espacial y temporal	Armonizados	Varía
Área de referencia	Áreas administrativas	Localización del usuario/dispositivo
Periodo de referencia	Definido	Indefinido, continuo, puede ser en tiempo real
Puntualidad	Bajo	Alto
Mecanismo de control de calidad	Presente	Desconocido
Costo	Alto	Varía pero puede ser bajo
Estabilidad/sostenibilidad de los datos fuente	Alto	Bajo

Nota: traducción propia del cuadro “Characteristics of Innovative and Traditional Data”. En Bosco, C., Grubanov-Boskovic, S., Iacus, S.M., Minora, U., Sermi, F., & Spyrtos, S. (2022). *Data Innovation in Demography, Migration and Human Mobility* (L. Publications Office of the European Union, Ed.). <https://doi.org/10.2760/027157>.

Los trabajos que han incorporado fuentes de información innovadoras para la medición de la migración internacional representan un avance en los estudios sobre demografía digital, realizados principalmente en Europa y Estados Unidos (Alexander et al., 2020; Palotti et al., 2020; Rampazzo et al., 2021; Spyratos et al., 2019; Zagheni et al., 2017), y algunos casos en países de América Latina, como Colombia (Palotti et al., 2020). Estos estudios además se han servido de fuentes de información tradicionales variadas como encuestas y estadísticas centralizadas por organismos internacionales. Respecto de la medición de la migración internacional, sin embargo, en América Latina no hay evidencia hasta ahora del uso de censos poblacionales para entender los posibles sesgos de una fuente de información innovadora como Facebook.

De la exploración de los estudios sobre demografía digital resulta evidente la necesidad de elaborar más estudios sobre países del sur global; y, en consecuencia profundizar en la interacción de poblaciones migrantes en Internet y redes sociales, con el fin de sentar precedentes en la comparación de un caso específico de América Latina respecto de los estudios desarrollados sobre Estados Unidos y Europa.

En este sentido, esta investigación espera brindar evidencia empírica sobre la utilización de Facebook como una fuente innovadora para la medición de stocks de inmigración internacional. Considerando los sesgos intrínsecos de esta plataforma, se espera encontrar diferencias cuantitativas en la desagregación por edad, sexo y país de origen de los stocks de inmigración que resultan de contrastar el Censo de Población y Vivienda 2020 de México y los datos de expatriados de esta red social en línea. Asimismo, se espera encontrar diferencias cuantificables en las proporciones de acceso a internet y a Facebook de la población inmigrante en México. Como resultado de estas mediciones se prevé obtener información suficiente para determinar si Facebook puede desempeñarse como un censo digital para realizar estimaciones de inmigración independientes de una fuente de información tradicional, teniendo en consideración la cuantificación de sesgos en las variables demográficas.

**CAPÍTULO II:**  
**ESTRATEGIA METODOLÓGICA**

El propósito de esta sección es presentar los datos, los indicadores y los métodos empleados para alcanzar los objetivos de esta tesis. La estrategia metodológica se basó en el trabajo de Zagheni et al. (2017), en el que los autores evaluaron datos de expatriados de Facebook a la luz de fuentes de información tradicionales como la ACS y las estimaciones de población migrante del Banco Mundial. En la figura 1 se muestra la esquematización de esta sección.

**Figura 1.**  
*Esquematización de la estrategia metodológica*



Fuente: elaboración propia.

## 2.1 DATOS

### 2.1.1 Fuentes de información

La fuente de información *tradicional* empleada fue el Censo de Población y Vivienda 2020 (Censo 2020), realizado del 2 al 27 de marzo de 2020 en México. Este censo tiene como objetivo producir la información sobre la dimensión, estructura y distribución espacial de la población, así como de sus principales características socioeconómicas y culturales, además de obtener el número de viviendas y sus principales características (INEGI, 2021). Los censos en México son *de jure*, esto implica que la población se contabiliza en el lugar donde reside habitualmente. Los datos que se emplean aquí provienen de la muestra del cuestionario ampliado del censo publicados por el INEGI<sup>7</sup>. Para la construcción de los indicadores y los modelos que se estimaron en esta investigación, se utilizó la muestra de aproximadamente 11.9 % que publica el INEGI en forma de microdatos con base al cuestionario ampliado del Censo 2020.

La fuente de información *innovadora* o de *big data* que se empleó fue Facebook Advertising. Para esta investigación, se utilizaron datos del Proyecto "Using internet-based data to quantify and sample international migrants. Applications to examine recent immigration to Uruguay" financiado por la Agencia Nacional de Investigación e Innovación (Uruguay) y el Max Planck Institute in Demographic Research (Alemania). Este proyecto comenzó en diciembre de 2019 y desde entonces ha descargado datos de la población total de usuarios de Facebook, y de expatriados en México y de otros países latinoamericanos, incluyendo extracciones realizadas en marzo de 2020. Para el propósito de esta investigación se utilizaron los usuarios activos por mes (MAU, por sus siglas en inglés) máximos; y el mecanismo de acceso a esta información fue a través de la Interfaz de programación de aplicaciones de Meta (API, por sus siglas en inglés). La segmentación demográfica se hizo con base en las variables demográficas y de comportamiento de usuarios en Facebook disponibles en *Meta for developers*<sup>8</sup>.

Se emplearon los datos del Proyecto mencionado anteriormente, porque los datos extraídos de la API de Facebook no se pueden extraer de forma retrospectiva, sino mediante la descarga diaria de los DAU, o mensual de los MAU; y que estos datos, si bien corresponden a usuarios de

---

<sup>7</sup> Los microdatos del CNPV 2020 se pueden consultar en:

<https://www.inegi.org.mx/programas/ccpv/2020/#Microdatos>

<sup>8</sup> Para ampliar esta información consultar: <https://developers.facebook.com/docs/marketing-api/audiences>

Facebook, se presentan de modo agregado, por lo tanto, son datos anonimizados. Asimismo, los cambios en los algoritmos de Facebook para definir sus audiencias publicitarias y la variable expatriado pueden cambiar sin previo aviso (Rampazzo et al., 2021; Zagheni et al., 2017), por lo tanto, el escenario ideal para contrastar una fuente de información tradicional con datos de Facebook se da cuando el levantamiento de información de ambas fuentes se hace al mismo tiempo. De este modo, también se espera que el volumen de las poblaciones objeto de estudio coincidan en tiempo y forma, puesto que el dinamismo en los recursos digitales es mucho más acelerado. Precisamente, esta investigación se sustenta en esta coincidencia, ya que los datos del Censo 2020 y los de usuarios de Facebook se obtuvieron en marzo de 2020. Adicionalmente, en marzo de 2020 se declaró la pandemia por el SARS-COV-2 (OMS, 2021), y este evento limitó significativamente la movilidad internacional. Por lo tanto, se esperaría que Facebook pudiese captar de modo más completo, con respecto a fechas anteriores, la población inmigrante en México en marzo de 2020.

## 2.2 INDICADORES

En esta investigación se emplean tres indicadores como mediciones preliminares a la estimación de los modelos de regresión; los cuales proporcionan el contexto de dos condiciones previas para analizar la inmigración en México con los datos de Facebook y el Censo 2020, a saber: que los inmigrantes tengan acceso a internet y que tengan acceso a Facebook. A continuación, las características demográficas sobre las que se desarrollan los resultados de esta investigación; las definiciones de inmigración empleadas; la formalización de la Proporción de Acceso a Internet (PAI); y la Razón de Acceso a Facebook (RAF).

### *2.2.1 Características demográficas y definiciones de inmigración*

Las características demográficas que se comparan en esta investigación son la edad, el sexo y el país de origen. En vista de que Facebook registra edades desde los 13 años y hasta los 65 años, se realizó el análisis con base en tres grupos etarios: de 13 a 24 años, de 25 a 54 años y de 55 a 65 años. Asimismo, se incorporó la variable sexo (hombres y mujeres); y la información del volumen de expatriados de trece países de América Latina: Argentina, Brasil, Chile, Colombia, Cuba, El Salvador, Guatemala, Haití, Honduras, Nicaragua, Perú, República Dominicana y Venezuela. Con estas características demográficas, se obtuvieron 78 combinaciones posibles para el análisis de expatriados. Esta información proviene de la base de datos de *personas* del Cuestionario ampliado del Censo 2020, que se publica en forma de microdatos por el INEGI.

En función de estas características, se adecuó la información de inmigrantes derivada del Censo 2020 con el propósito de realizar mediciones comparables. En este sentido, aunque para México la inmigración proveniente de Estados Unidos, Canadá y España es importante en términos de volumen y características, estos países estarán excluidos del análisis que se hace con los datos de Facebook, ya que la información de expatriados obtenida en marzo de 2020 se focalizó en las poblaciones de países de América Latina.

Con fundamento en el Censo 2020, se establecieron seis posibles definiciones de inmigración con la que la variable expatriado de Facebook sería comparable. El propósito de delimitar estas definiciones es tener un conjunto de procesos migratorios distintos, en los que se iteran criterios de lugar de nacimiento y residencia, para entender mejor la forma en la que Facebook configura la variable expatriado. Las definiciones de inmigración descritas aquí se

basaron en las respuestas a las preguntas sobre el país de residencia cinco años atrás y el país de nacimiento de la población extranjera; así como las respuestas sobre la residencia en un país extranjero de la población mexicana en 2015. Estas seis definiciones son:

1. Inmigrantes por país de nacimiento. Personas que nacieron en un país distinto a México, y que eran residentes habituales de este país en el momento de la aplicación del cuestionario del censo. Esta definición se hizo con base en la pregunta *¿En qué estado de la República Mexicana o en qué país nació?*, del Censo 2020.
2. Inmigrantes por lugar de residencia cinco años antes. Población que era residente habitual en un país distinto a México en 2015, con una edad mayor o igual a cinco años; comprende además de los extranjeros, a los mexicanos que residían en el exterior en 2015 (población de retorno). Esta definición se hizo con base en la pregunta *Hace 5 años, en marzo de 2015, ¿en qué estado de la República o en qué país vivía?*, del Censo 2020.
3. Inmigrantes recientes por lugar de nacimiento. Población de cinco años o más que nacieron en un país distinto a México (exclusivamente extranjeros).
4. Inmigrantes que en 2015 residían en el mismo país que en el que nacieron. Población de cinco años o más cuyo país de residencia habitual en 2015 es el mismo país donde nacieron. Un caso hipotético estaría representado por una persona que nació en Venezuela y residía en este mismo país en 2015.
5. Inmigrantes que en 2015 residían en un país distinto que en el que nacieron. Población de cinco años o más cuyo país de residencia habitual en 2015 era distinto al país en donde nacieron. Un caso hipotético estaría representado por una persona que nació en Venezuela, pero residía en Colombia en 2015.
6. Población mexicana de retorno. Población de cinco años o más que nació en México y que en 2015 vivía en el exterior. Un caso hipotético estaría representado por una persona que nació en México y residía en Estados Unidos en 2015.

Cabe precisar que las definiciones basadas en la pregunta sobre el lugar de residencia cinco años antes, muestran información sobre un periodo de cinco años, y no de un año en particular (Masferrer, 2021). En este sentido, las definiciones de la dos a la seis representan la inmigración del quinquenio 2015 – 2020. Asimismo, se debe tener en cuenta que, en razón del planteamiento

de la pregunta (refiriéndose a marzo de 2015), es probable que la persona migrante haya pasado por países distintos al que reportó en el periodo 2015 – 2020 (Welti Chanes, 1997). Podría plantearse hipotéticamente el caso de una persona migrante que respondió en el Censo 2020 que en el año 2015 residía en Chile, pero que, luego en 2018 residió en Colombia. En los datos del censo esa historia migratoria quedaría truncada y, se definiría que esta persona es ahora inmigrante en México que proviene de Chile.

### 2.2.2 Proporción de Acceso a Internet

La Proporción de Acceso a Internet (PAI) relaciona en su numerador el conjunto de inmigrantes con acceso al servicio de internet y en su denominador el total de la población (inmigrantes y no inmigrantes) con acceso y sin acceso al servicio de internet. Tanto el numerador como el denominador están desagregados por grupo etario, sexo y país de origen. Para su cálculo, se utilizaron datos del Censo 2020. Específicamente se utilizó la base de datos de *viviendas* donde se pregunta sobre el acceso a internet de los hogares; esta se cruzó con las características demográficas de la definiciones de inmigración (edad, sexo y país de origen) de la base de datos de *personas*. Los datos utilizados provienen del Cuestionario ampliado del Censo 2020 que se publican en forma de microdatos por el INEGI.

La Proporción de Acceso a Internet capta el porcentaje de acceso a internet de las poblaciones y visibiliza las posibles diferencias en el acceso, considerando las características demográficas señaladas anteriormente. La ecuación 1, muestra la formalización de las PAI específicas:

#### Ecuación 1.

$$PAI_{e,s,o} = \frac{\text{Personas con acceso a internet}_{e,s,o}}{(\text{Personas con acceso a internet} + \text{Personas sin acceso a internet})_{e,s,o}} \times 100$$

Donde e = grupo etario; s = sexo; o = origen

### 2.2.3 Razón de acceso a Facebook

La Razón de Acceso a Facebook (RAF) relaciona la población total de Facebook con la población total de un país derivada de un censo. El numerador de este indicador representa datos de la población total extraídas de la API de Facebook, que están desagregados por grupo etario y sexo; mientras que para el denominador se utiliza la base de datos de *personas* del Cuestionario ampliado del Censo 2020, la cual se publica en forma de microdatos por el INEGI. En ese sentido, esta medida visibiliza la penetración que tiene Facebook en la población que reside en México según edad y sexo; esto significa que se incluye la población que nació en México y la población nacida en el extranjero que reside en territorio mexicano. La ecuación 2 muestra la formalización de las RAF específicas:

#### **Ecuación 2.**

$$RAF_{e,s} = \frac{pob\ Facebook_{e,s}}{pob\ Censo_{e,s}}$$

Donde  $e$  = grupo etario;  $s$  = sexo

## 2.3 MÉTODOS

Para esta investigación se estimaron modelos de regresión que permitieron evaluar los sesgos de los datos de Facebook en comparación con el Censo 2020. La estrategia empleada consistió en estimar modelos de regresión lineal aditivos en los que la primera variable independiente que se incluyó fue la proporción específica de expatriados en escala logarítmica. Luego se incorporaron progresivamente, como variables independientes, la edad, el sexo y el país de origen. Posteriormente, se estimó un set de modelos separados por sexo, en los que se siguió la misma lógica de adición de variables independientes de forma progresiva.

Este ejercicio tuvo como resultado 24 modelos generales (4 modelos x 6 definiciones de inmigración) y 36 modelos separados por sexo (3 modelos x 6 definiciones de inmigración x 2 sexos). Con los coeficientes estimados, se obtuvo información sobre la dirección de los sesgos en cuanto a edad, sexo y país de origen. Los coeficientes con signo negativo indican una sobrestimación, mientras que los coeficientes positivos indican una subestimación en los stocks de inmigración captados por el censo en México. A continuación, se presenta la formalización de estos modelos.

### 2.3.1 Formalización de modelos lineales para la evaluación de sesgos

#### Modelo 1.

$$\log(\text{PropInmg}_{e,s,o}) = \beta_0 + \beta_1 \log(\text{PropExpat}_{e,s,o}) + \varepsilon_{e,s,o}$$

#### Modelo 2.

$$\log(\text{PropInmg}_{e,s,o}) = \beta_0 + \beta_1 \log(\text{PropExpat}_{e,s,o}) + \beta_2 \text{sexo}_m + \varepsilon_{e,s,o}$$

#### Modelo 3.

$$\begin{aligned} \log(\text{PropInmg}_{e,s,o}) \\ = \beta_0 + \beta_1 \log(\text{PropExpat}_{e,s,o}) + \beta_2 \text{sexo}_m + \beta_3 (\text{edad}_1) + \beta_4 (\text{edad}_2) \\ + \varepsilon_{e,s,o} \end{aligned}$$

**Modelo 4.**

$$\begin{aligned} \log(\text{PropInmg}_{e,s,o}) \\ = \beta_0 + \beta_1 \log(\text{PropExpate}_{e,s,o}) + \beta_2 \text{sexo}_m + \beta_3(\text{edad}_1) + \beta_4(\text{edad}_2) \\ + \beta_5(\text{origen}_1) + \dots + \beta_{17}(\text{origen}_{12}) + \varepsilon_{e,s,o} \end{aligned}$$

De estos modelos,  $\text{PropInmg}_{e,s,o}$  es la proporción de inmigración específica (grupo etarios, sexo, país de origen), que se calculó con base en las seis definiciones de inmigración presentadas anteriormente;  $\text{PropExpate}_{e,s,o}$  es la proporción de expatriados específica;  $\text{sexo}_m$  es una variable dicotómica, donde la variable de referencia son los hombres ( $\text{sexo}_h$ );  $\text{edad}_1$  y  $\text{edad}_2$  indican los grupos etarios de 25 a 54 años y de 55 a 65 años, respectivamente y en el que el grupo etario de 13 a 24 años ( $\text{edad}_0$ ) queda como categoría de referencia;  $\text{origen}_1$  hasta  $\text{origen}_{12}$  indican los países de origen, donde Argentina ( $\text{origen}_0$ ) queda como categoría de referencia.

### 2.3.2 Formalización de modelos lineales para la evaluación de sesgos desagregados por sexo

#### Modelo 1.

$$\log(\text{PropInmg}_{e,s,o}) = \beta_0 + \beta_1 \log(\text{PropExpat}_{e,s,o}) + \varepsilon_{e,s,o}$$

#### Modelo 2.

$$\log(\text{PropInmg}_{e,s,o}) = \beta_0 + \beta_1 \log(\text{PropExpat}_{e,s,o}) + \beta_2(\text{edad}_1) + \beta_3(\text{edad}_2) + \varepsilon_{e,s,o}$$

#### Modelo 3.

$$\begin{aligned} \log(\text{PropInmg}_{e,s,o}) \\ = \beta_0 + \beta_1 \log(\text{PropExpat}_{e,s,o}) + \beta_2(\text{edad}_1) + \beta_3(\text{edad}_2) + \beta_4(\text{origen}_1) \\ + \dots + \beta_{16}(\text{origen}_{12}) + \varepsilon_{e,s,o} \end{aligned}$$

Las variables de los modelos desagregados por sexo están definidas de forma análoga a los modelos generales, donde  $\text{PropInmg}_{e,s,o}$  es la proporción de inmigración específica (grupo etarios, sexo, país de origen), que se calculó con base en las seis definiciones de inmigración presentadas anteriormente;  $\text{PropExpat}_{e,s,o}$  es la proporción de expatriados específica;  $\text{edad}_1$  y  $\text{edad}_2$  indican los grupos etarios de 25 a 54 años y de 55 a 65 años, respectivamente y en el que el grupo etario de 13 a 24 años ( $\text{edad}_0$ ) queda como categoría de referencia;  $\text{origen}_1$  hasta  $\text{origen}_{12}$  indican los países de origen, donde Argentina ( $\text{origen}_0$ ) queda como categoría de referencia. Cuando el subíndice  $s$  sea igual a *hombres* se obtendrán los modelos específicos para los hombres y cuando el subíndice  $s$  sea igual a *mujeres* se obtendrán los modelos específicos para las mujeres.

**CAPÍTULO III:  
RESULTADOS**

Para iniciar la presentación de resultados de esta tesis es conveniente recordar sus objetivos, a saber: (1) comprender en qué medida internet y Facebook son recursos generalizados en las poblaciones migrantes de México; y (2) analizar la validez de la definición de *expatriado* de Facebook y el alcance de esta red social para captar y medir poblaciones migrantes en tiempo real. La motivación que impulsó la elaboración este trabajo radica en contribuir a los esfuerzos de la academia por entender la forma en la que las redes sociales en línea como Facebook pueden utilizarse como un instrumento de medición de migración internacional en contextos de escasez de datos basados en fuentes tradicionales. La necesidad de disponer de fuentes de información alternativas e innovadoras se hizo particularmente evidente para el año 2020, cuando la pandemia por SARS-COV-2 postergó en la mayoría de los países de América Latina, la ronda de censos 2020. Sin embargo, en México, el levantamiento del Censo 2020 se llevó a cabo con éxito, lo que permitió tener un insumo fundamental para evaluar los datos de Facebook.

Como se ha mencionado antes, la definición de expatriados de Facebook carece de los elementos suficientes para constituirse en una medición de migración internacional, tal como está comúnmente aceptada (Spyratos et al., 2019). No obstante, su definición se asemeja a la connotación de inmigración por país de residencia previa que se deriva de los censos; sin embargo, los estudios que se han publicado hasta la fecha (Alexander et al., 2020; Palotti et al., 2020; Spyratos et al., 2019; Zagheni et al., 2017) han utilizado la definición de inmigración por país de nacimiento, debido a que en el momento de elaboración de sus estudios no contaban con un censo de población con fechas de levantamiento de información cercanas a la extracción de datos de Facebook, que les permitiera contrastar distintas definiciones de inmigración.

Un aporte fundamental de esta investigación es contrastar seis definiciones de inmigración derivadas del Censo 2020 a las que la variable expatriado se podría aproximar. Debido a que Facebook registra información sobre lugar de nacimiento, cambios de residencia (interna e internacional), y georreferenciación por medio del registro de sitios visitados, entre otros aspectos, resulta plausible suponer la relación entre la variable expatriado, con casi cualquier definición de inmigración que simultáneamente implique la interacción entre los datos sobre el país de nacimiento y los cambios de residencia internacional o registro de visitas en lugares en el extranjero por largos periodos.

En esta investigación, el proceso de contraste de las seis definiciones de inmigración se hizo de modo sistemático, no obstante, para simplificar la presentación de los resultados, la exposición se centrará en las definiciones uno y dos derivadas del censo, en comparación con la variable expatriado, es decir, inmigración por país de nacimiento y migración por país de residencia en 2015. Se decidió proceder de esta forma por dos razones: primero, para mantener la discusión con la literatura especializada sobre el tema, en la que preponderantemente aparece la definición de inmigración por país de nacimiento; y dos, porque la definición de inmigración por país de procedencia es la que presenta un mejor ajuste en términos estadísticos para el caso de México. Esta última razón se desarrollará con mayor detenimiento en la sección 7.4 sobre el análisis de sesgos de los datos de Facebook. De igual modo, resulta ilustrativo comprender la estructura de las poblaciones inmigrantes en México debido a que estas dan cuenta de procesos de asentamiento reciente o inmigración permanente. Esto se relaciona sustancialmente con la discusión sobre los patrones demográficos en términos de edad, sexo y país de origen, así como sobre la discusión de movilidad y migración que se elabora en esta tesis.

### 3.1 ESTRUCTURA POBLACIONAL Y CONTEXTO DE LA INMIGRACIÓN EN MÉXICO

México es un país con 126,014,024 habitantes, de los cuales 1,169,883 nacieron en el extranjero, lo que representa 0.93 % del total de la población en México. De esta población, los cinco primeros países con las proporciones más altas por país de origen son: Estados Unidos (64.2 %), Guatemala (5.1 %), Venezuela (4.8 %), Colombia (3 %) y Honduras (2.9 %). En relación con la inmigración por país de procedencia, en los cinco primeros lugares se mantienen los mismos países mencionados antes, pero en un orden distinto. Estados Unidos es el principal país de procedencia de inmigración y representa 67.2 %; seguidamente está Venezuela con 7.8 %, posteriormente está Colombia (2.5 %), Guatemala (2.4 %) y Honduras (2.3 %). Esto muestra el volumen y procedencia de las definiciones uno y dos de inmigración. En las cuatro definiciones de inmigración adicionales el orden de los países es semejante y Estados Unidos se posiciona invariablemente en el primer lugar. En las definiciones 3, 4 y 5, Venezuela es el primer país de nacimiento y procedencia de la población migrante en México. En cuanto a la población retornada, los tres principales países de donde proviene la población mexicana son Estados Unidos (88.3 %), Canadá (1.7 %) y España (1.3 %).

La magnitud de estas proporciones de inmigración está asociada en algunos casos con contextos históricos, y en otros con contextos coyunturales, que motivó la emigración de estas poblaciones hacia México. A continuación, se presenta un contexto breve sobre la inmigración de países, cuyos procesos de llegada a México se podrían denominar *históricos* o de más larga duración; y seguidamente, se presenta un contexto breve de tres casos de procesos de inmigración más recientes en México o basados en procesos *coyunturales* recientes en los países de salida. Este abordaje contextual de los orígenes de la inmigración en México permite una aproximación cualitativa para entender mejor el contraste de información que se desarrolla con base en el Censo de Población y Vivienda 2020 y los datos de Facebook.

Castillo y Rojas Wiesner (2020), señalan que la inmigración española se dio entre 1930 y 1940 como consecuencia de la Guerra Civil, cuando los republicanos exiliados encontraron en México un país de acogida; la inmigración de países como Chile y Argentina se produjo en condiciones similares a la española en la década de 1970; para el caso de Nicaragua, Guatemala y El Salvador, los motivos de la emigración inicial guardaron relación con la inestabilidad política, de inseguridad y de conflicto interno en sus países en la década de 1980. Para estos tres casos de

inmigración, México recibió estas nacionalidades en el marco de las relaciones político–diplomáticas, por lo que adquirió importancia a nivel internacional (Castillo & Rojas Wiesner, 2020). De esta descripción, se tiene que la estructura poblacional de la inmigración de España muestra una pirámide más homogénea a lo largo de los grupos etarios, con edades sobresalientes en la población económicamente activa (PEA) y edades de 70 a 79 años, esto muestra una estructura poblacional relativamente envejecida y con patrones de asentamiento de más larga duración (Gráfico 1). Por su parte, la inmigración de Argentina, Chile, El Salvador, Guatemala y Nicaragua muestra un patrón de población preponderantemente adulta, que se concentran en edades económicamente activas; esto da como resultado una pirámide poblacional con menor concentración en sus extremos.

Los motivos de la inmigración de Estados Unidos están relacionados principalmente con procesos de retorno voluntario o involuntario de población mexicana, cuyas causas son múltiples y se pueden asociar con motivaciones económicas, familiares o de curso de vida (Masferrer, 2020). Es de notar que la mayor proporción de población nacida en Estados Unidos presente en México se encuentra en edades tempranas, lo que hace plantear la hipótesis de que esta inmigración se asocia, principalmente, a migración familiar (Castillo, 2014; Castillo & Rojas Wiesner, 2020) pues la mayoría reside con madre o padre mexicano (Masferrer et al., 2019). Este argumento se refuerza al observar las pirámides poblacionales del Gráfico 1, donde se destaca la concentración en los primeros grupos etarios (19 años y menos).

En la inmigración de menor tradición histórica destacan, en el Censo 2020, países como Venezuela, Haití y Colombia cuyo volumen y proporción han aumentado en comparación con el Censo de 2010. De acuerdo con Gandini et al., (2020), la inmigración venezolana reciente en México y otros países de América Latina encuentra sus causas en el deterioro de la estructura política, económica y social de Venezuela. Estos mismos autores, señalan que la migración venezolana en México tiene un grado mayor de selectividad, en términos de educación, comparado con países más cercanos geográficamente a este país. Otro aspecto destacado es el creciente número de solicitudes de refugio que ha realizado la población venezolana en México, considerando que, en 2021 el número de solicitudes fueron aproximadamente 17 veces mayor que en 2016, con 6,134 y 361 respectivamente (COMAR, 2018, 2022). De igual modo, se observa que en la estructura poblacional de la inmigración por país de nacimiento y país de residencia en 2015,

hay una mayor concentración en edades entre 20 y 50 años, esto da señales de una migración en una etapa relativamente reciente, debido a su concentración en edades económicamente activas (Grafico 1).

Del lado de migración haitiana, Moreno Mena (2019), señala que las motivaciones surgen de procesos de inestabilidad política y violencia; así como de crisis humanitarias a causa de desastres de origen natural, como los huracanes de 2004, 2008 y 2010. Estos escenarios de vulnerabilidad, aunados a la debilidad de la respuesta institucional del gobierno de Haití, actuaron como factores expulsores para un gran número de personas de este país. En los últimos veinte años, los motivos de llegada de población haitiana a México se relacionan con un proyecto migratorio hacia Estados Unidos; luego de experimentar dificultades para el ingreso o la integración en países como Chile, Brasil y Argentina. En el tránsito hacia Estados Unidos, esta población ha buscado en México opciones de asentamiento y refugio. Sin embargo, en 2021 la tasa de reconocidos como refugiados fue de 23 % para la población haitiana, mientras que para los venezolanos fue de 97% (COMAR, 2022). En relación con la estructura poblacional de la inmigración por país de nacimiento y país de residencia en 2015, se observa una concentración preponderante en las edades económicamente activas, asimismo, hay una proporción mayor de hombres de 30 a 39 años (Grafico 1).

El proceso de emigración de la población colombiana ha estado vinculado a diversos factores sociopolíticos, económicos y de conflicto armado desde la década de 1970. De acuerdo con Polo Alvis et al., (2019), el proceso de inmigración en México no se vincula exclusivamente con procesos de asilo y refugio; sino que toma importancia la oferta educativa y laboral a la que acceden los colombianos en este país, debido a que, en promedio, tienen un nivel educativo mayor que las personas que no migran en Colombia. Además, son personas dueñas de medianas empresas, y en general de clase media que identifican factores restrictivos para su desarrollo humano en el país de origen (Valenzuela Montoya, 2011). Por consiguiente, existen factores diferenciados que han dado forma al asentamiento de los colombianos en México que, a su vez, se reflejan en su pirámide poblacional, donde la mayor parte de la población se concentra en la PEA de 20 a 49 años (Gráfico 1).

Finalmente, es preciso resaltar la importancia de analizar las definiciones de inmigración a luz de los procesos históricos y coyunturales de un país. Para el caso mexicano, donde la

inmigración representa menos del uno por ciento, y en donde la migración de retorno de Estados Unidos representa un componente cualitativo significativo, cobra relevancia testear mediciones que capten la realidad de los procesos de inmigración de larga o reciente duración en el país de destino; por esto cobra relevancia disponer de un instrumento de medición como el Censo de población que permite desarrollar definiciones variadas de inmigración.

**Cuadro 2**

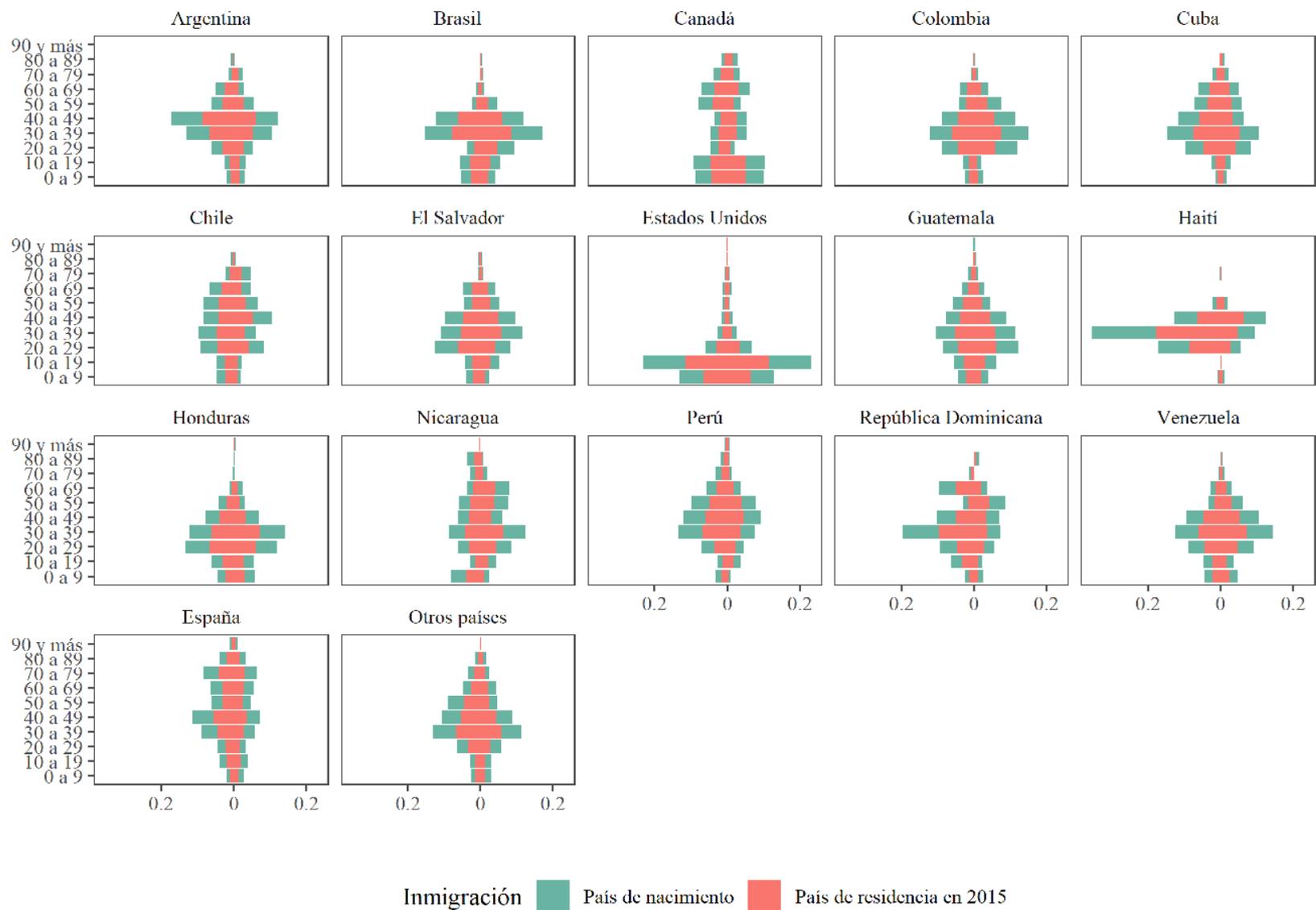
*Distribución absoluta y relativa de la población inmigrante en México de acuerdo a seis definiciones de inmigración del Censo 2020*

País	Inmigrantes absolutos		Inmigrantes recientes llegados entre 2015 y 2020									
	Inmigrantes por país de nacimiento		Por lugar de residencia en 2015		Por lugar de nacimiento		Nacidos en el mismo país de residencia en 2015, según país de nacimiento		Nacidos en un país distinto del de residencia en 2015, según país de residencia		Población mexicana retornada según país de residencia en 2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Argentina	18,908	1.6%	9,016	1.6%	6,444	2.8%	5,762	2.6%	682	4.2%	2,883	0.9%
Brasil	9,315	0.8%	6,803	1.2%	4,362	1.9%	3,926	1.8%	436	2.7%	1,462	0.4%
Canadá	10,936	0.9%	8,858	1.6%	2,928	1.3%	2,613	1.2%	315	1.9%	5,615	1.7%
Chile	7,928	0.7%	3,656	0.6%	2,348	1.0%	2,234	1.0%	114	0.7%	686	0.2%
Colombia	34,754	3.0%	13,940	2.5%	13,865	5.9%	12,376	5.7%	1,489	9.2%	1,180	0.4%
Cuba	25,998	2.2%	12,671	2.2%	12,734	5.4%	11,983	5.5%	751	4.6%	645	0.2%
El Salvador	19,481	1.7%	6,585	1.2%	6,783	2.9%	5,957	2.7%	826	5.1%	399	0.1%
España	21,995	1.9%	8,847	1.6%	4,442	1.9%	3,846	1.8%	596	3.7%	4,179	1.3%
Estados Unidos	751,363	64.2%	381,464	67.2%	83,245	35.6%	81,841	37.6%	1,404	8.7%	294,516	88.3%
Guatemala	59,866	5.1%	13,791	2.4%	12,717	5.4%	11,947	5.5%	770	4.8%	1,283	0.4%
Haití	8,821	0.8%	3,893	0.7%	4,777	2.0%	3,763	1.7%	1,014	6.3%	128	0.0%
Honduras	33,675	2.9%	13,089	2.3%	13,358	5.7%	12,558	5.8%	800	4.9%	443	0.1%
Nicaragua	4,706	0.4%	1,984	0.3%	1,483	0.6%	1,429	0.7%	54	0.3%	404	0.1%
Perú	9,840	0.8%	3,229	0.6%	2,648	1.1%	2,206	1.0%	442	2.7%	762	0.2%
República Dominicana	2,532	0.2%	1,151	0.2%	748	0.3%	535	0.2%	213	1.3%	338	0.1%
Venezuela	56,198	4.8%	33,826	6.0%	34,364	14.7%	31,709	14.6%	2,655	16.4%	1,093	0.3%
Otros países	93,567	8.0%	44,502	7.8%	26,439	11.3%	22,815	10.5%	3,624	22.4%	17,381	5.2%
Total de inmigrantes	1,169,883	100.0%	567,305	100.0%	233,685	100%	217,500	100.0%	16,185	100.0%	333,397	100.0%

Fuente: elaboración propia con datos del Censo 2020.

### Gráfico 1.

Distribución relativa de la población inmigrante en México según país de nacimiento y país de residencia en 2015, 2020



Fuente: elaboración propia con datos del Censo 2020

### **3.2 ACCESO A INTERNET Y A FACEBOOK DE LA POBLACIÓN INMIGRANTE EN MÉXICO**

Los medios digitales son un recurso de alto valor para la población migrante en cada de una de las etapas del proyecto migratorio —antes, durante y después— (Moran, 2022); principalmente, para mantener o adquirir nuevo capital social (Dekker & Engbersen, 2014). Tanto Internet como las redes sociales, se revisten de importancia para obtener información relevante sobre rutas y medios para la emigración; asimismo, se convierten en el medio para adquirir conocimientos sobre el país de destino y establecer o mantener redes de apoyo que faciliten el proceso de inmigración. Una vez en el lugar de destino, estos recursos digitales cobran importancia para que los inmigrantes mantengan contacto con sus familiares o personas cercanas en sus países de origen; y, simultáneamente, para que establezcan nuevas redes de apoyo en el país de destino. Adicionalmente, el acceso a información local y la interacción con poblaciones de acogida facilita los procesos de integración (Dekker & Engbersen, 2014; Maria Pesando et al., 2020). Gracias a la interacción de la población migrante en Internet y en redes sociales en línea como Facebook, se producen las huellas digitales que permiten hacer una medición aproximada de los fenómenos de migración internacional, como es el caso específico de la inmigración en México.

Considerando la interacción de la población migrante con los recursos digitales, esta investigación parte del supuesto de que el acceso a internet es una condición necesaria para que una persona se pueda conectar a Facebook. De esta premisa, se desprende la hipótesis de que no todas las personas tienen un acceso homogéneo a internet sino que, por el contrario, en el acceso se pueden ver diferencias de acuerdo con la edad, el sexo, y en el caso de la población inmigrante, con su país de nacimiento o procedencia. Asimismo, el acceso a internet depende de los dispositivos con los que se cuenta para lograr la conexión, los cuales pueden ser dispositivos móviles como celulares o fijos como computadores.

Para el propósito de esta investigación, a continuación, se presentan las proporciones de acceso a internet de la población objeto de estudio. Desde de ahora, la exposición sobre los países de nacimiento o procedencia de los inmigrantes se enfocará en los trece países de América Latina en los que se concentra esta investigación<sup>9</sup>.

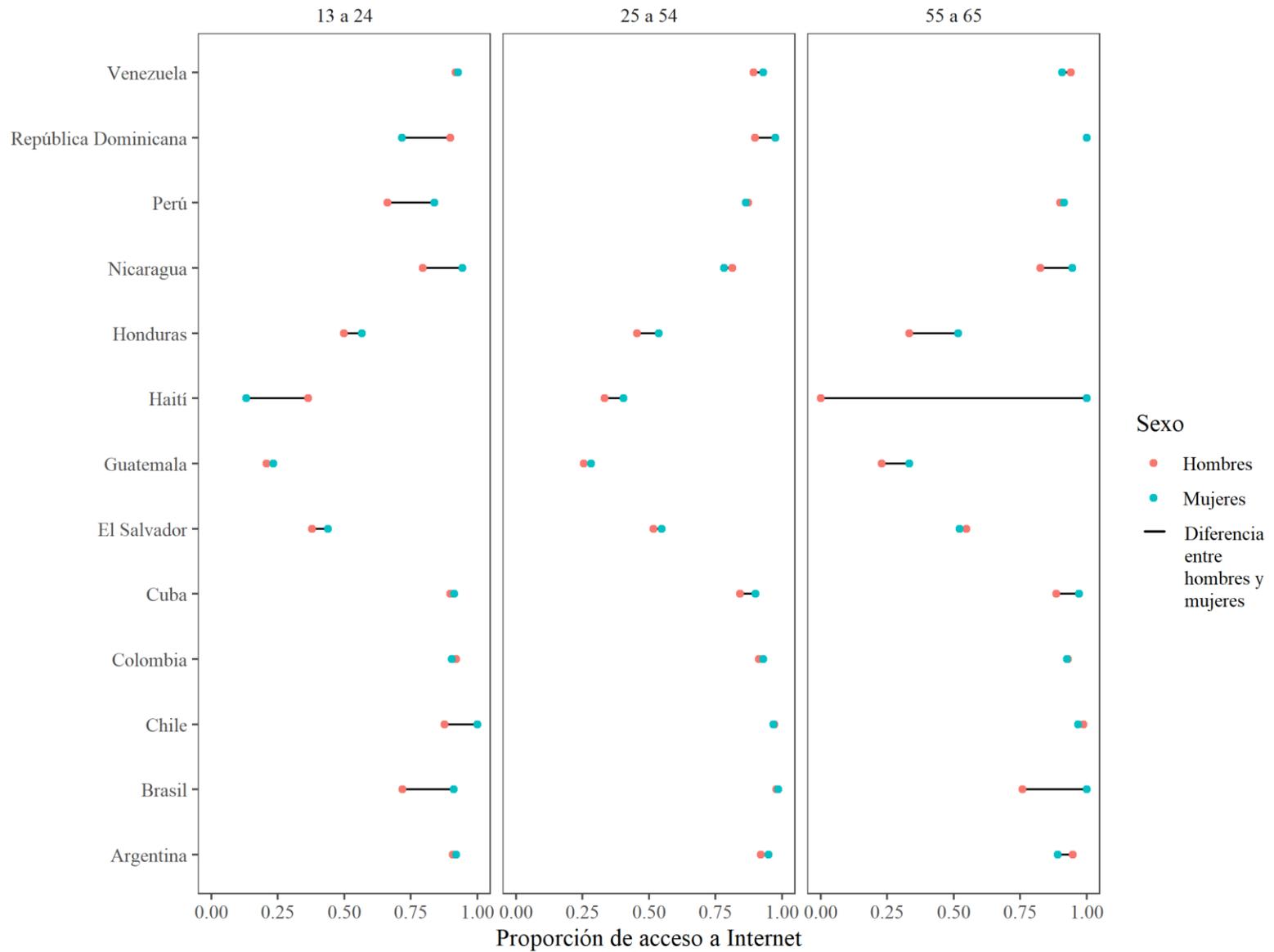
---

<sup>9</sup> Argentina, Brasil, Chile, Colombia, Cuba, El Salvador, Guatemala, Haití, Honduras, Nicaragua, Perú, República Dominicana, y Venezuela.

Para estimar la proporción de acceso a internet la población inmigrante, se utilizaron los datos del módulo de Vivienda del Censo 2020, en el que se pregunta si la vivienda cuenta con servicio de internet. Producto de estos datos y las características de las personas, en los Gráficos 2 y 3 se presentan las proporciones de acceso a internet para inmigrantes por país de nacimiento e inmigrantes por país de residencia en 2015. Es preciso recordar que el análisis del acceso a internet se hace a nivel de personas. Esto implica que las respuestas que se dan en relación con el acceso a internet se hacen extensivas a los residentes habituales de cada vivienda, independientemente de si en esta reside población extranjera, población mexicana o ambas. Este indicador tiene la limitación de contemplar un único ámbito donde los inmigrantes acceden a internet (la vivienda), pues no se consideran otros espacios donde eventualmente se puede acceder a internet, como en las escuelas, lugares de trabajo o espacios públicos donde hay Wifi libre.

## Gráfico 2

Proporción de acceso a internet de la población nacida en el exterior en México según país de nacimiento, 2020

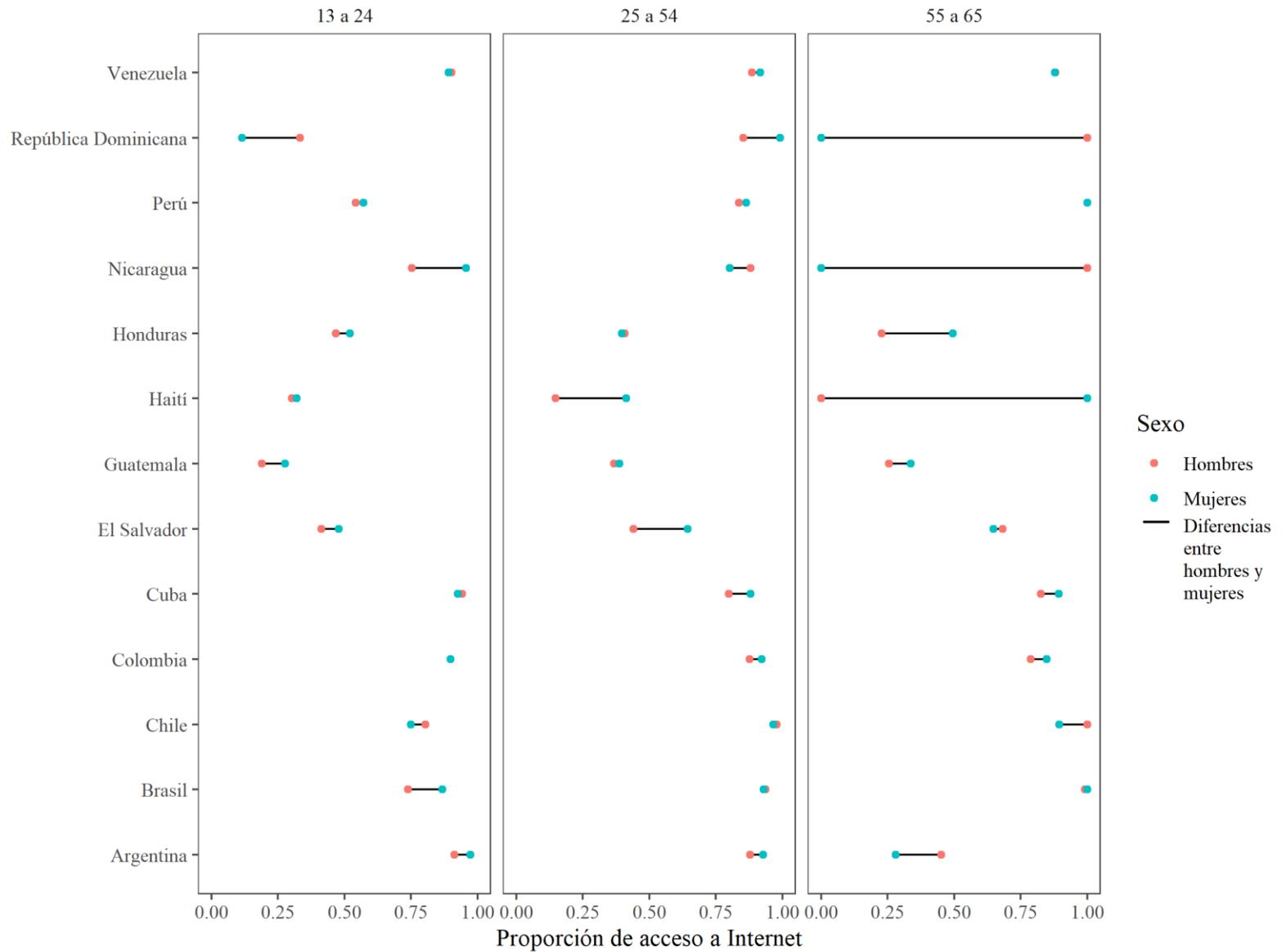


Fuente: elaboración propia con datos del Censo 2020

Para el caso de México, las proporciones de acceso a internet de la población inmigrante por país de nacimiento presenta diferencias de acuerdo con el grupo etario y el sexo. Los países que tienen proporciones de acceso a internet más altas son Chile, Venezuela, Argentina y Colombia; mientras que los países que tienen proporciones más bajas son Haití, Guatemala y El Salvador. El grupo etario en el que hay diferencias menores entre hombres y mujeres es el de 25 a 54 años; mientras que el grupo etario con diferencias más amplias es el de 55 a 65 años. Asimismo, un patrón adicional que se observa en el Gráfico 2 es que, en promedio, las mujeres migrantes tienen proporciones de acceso a internet más altas que los hombres migrantes. En lo que concierne al comportamiento de la inmigración por país de procedencia, se observan patrones similares; con la salvedad de que los inmigrantes recientes tienen menor acceso que los extranjeros en su conjunto, pues estos últimos, incluyen a población más establecida en el territorio mexicano.

### Gráfico 3

Proporción de acceso a internet de la población nacida en el exterior en México según país de residencia en 2015, 2020



Fuente: elaboración propia con datos del Censo 2020

Otra condición que favoreció la elaboración de esta investigación fue la alta penetración de Facebook en México. Como se ha mencionado antes, las métricas que se emplean para identificar la audiencia de Facebook por país son: usuarios activos mensuales (MAU) y usuarios activos diarios (DAU,); y derivado de estas métricas Facebook delimita un subconjunto llamado audiencia publicitaria (Kemp, 2022). El modo en el que Facebook delimita su audiencia no es publicado, pero es probable que esté relacionado con el comportamiento de los usuarios relacionado con las páginas en las que indicaron “me gusta”, y la información autodeclarada en cada perfil; anuncios en los que se hizo clic; lugares que se indicó haber visitado y registrado en Facebook<sup>10</sup>. De acuerdo con Kemp (2022) en 2022, México es el quinto país con más usuarios en Facebook con un total de 91.4 millones, después de Brasil, Indonesia, Estados Unidos e India. Adicionalmente, del total de usuarios activos en Facebook, el 69.6 % conformaba la audiencia publicitaria en México.

En esta investigación, la medida que se emplea es la de MAU *máximos* de 13 a 65 años tanto de población total, como de expatriados en México, porque la medición de usuarios mensuales es más estable que la medición diaria (Rampazzo et al. 2021). Para medir la penetración de Facebook, con los datos del Censo 2020 se segmentaron los grupos poblacionales conforme a los tres grupos etarios de interés y la separación por sexo que se describió en la sección de indicadores y se estimaron las Razones de Acceso a Facebook (RAF). De aquí se obtuvo que la RAF en México es de 98.3 % para la población en general; y que las RAF específicas tanto para hombres como para mujeres inmigrantes es mayor en el grupo etario más joven (109.7 % hombres y 117.8 % mujeres), y va disminuyendo en los siguientes dos grupos etarios. Un aspecto para resaltar en el grupo etario de 13 a 24 años es que la RAF es mayor a 100 %, esto se podría explicarse por la posibilidad de que un usuario de Facebook tenga más de una cuenta activa. Asimismo, solo en el grupo etario más joven, las mujeres tienen una RAF mayor que la de los hombres<sup>11</sup>.

---

<sup>10</sup> Para profundizar en el modo en el que Facebook presenta la publicidad a sus usuarios, consultar: [https://www.facebook.com/help/794535777607370/?helpref=uf\\_share](https://www.facebook.com/help/794535777607370/?helpref=uf_share)

<sup>11</sup> Las proporciones específicas de acceso a Facebook de la población total y de la población migrante según país de nacimiento y residencia en 2015, se presentan en el Apéndice 1.

### 3.3 ANÁLISIS DE SESGOS DE LOS DATOS DE FACEBOOK

#### 3.3.1 Evaluación gráfica

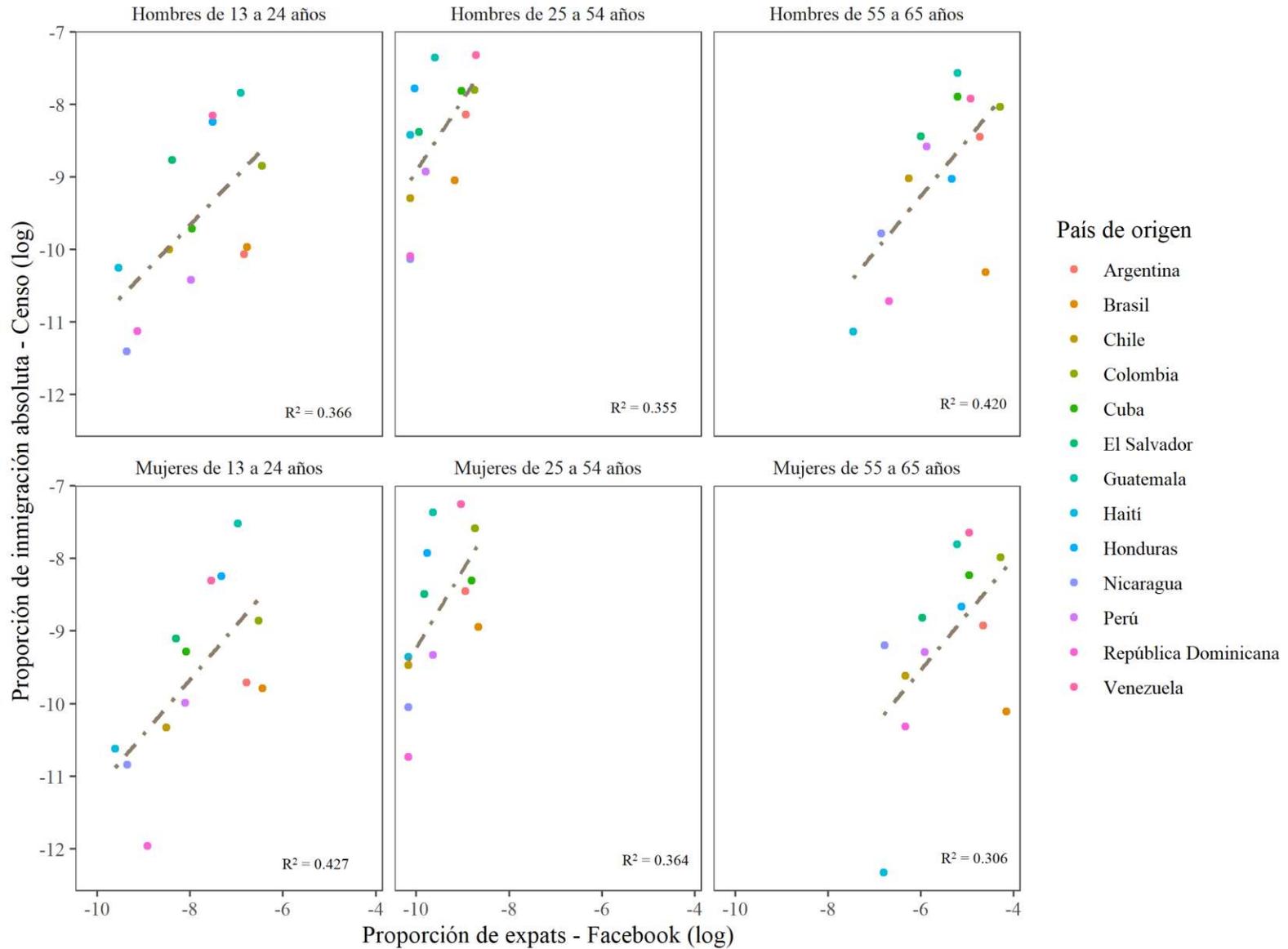
La comparación de los datos de Facebook con el Censo 2020 se hizo relacionando la proporción de inmigración de ambas fuentes de información en escala logarítmica con el fin de identificar los posibles sesgos derivados de los datos de Facebook. Los gráficos 4 y 5 relacionan los grupos por sexo y edad de los expatriados de Facebook con las definiciones de inmigración por país de nacimiento y procedencia del Censo 2020, respectivamente. De aquí se observa que la definición de inmigración que muestra una dispersión menor respecto a la diagonal de 45 grados es la *de país de procedencia*. Esta relación se da así porque la definición de expatriado de Facebook se basa en el lugar de residencia previa. Asimismo, los datos de estos gráficos presentan un comportamiento similar a los resultados de Zagheni et al., (2017), donde hay sesgos sistemáticos en la mayoría de los datos, dado que, se sitúan por encima de la línea de 45 grados, lo que implica que Facebook tiende a subestimar los stocks de migración en México.

Este análisis gráfico muestra que las dos definiciones de inmigración derivadas del censo presentan comportamientos distintos respecto a los grupos etarios. En la definición por país de nacimiento no hay un patrón constante en los grupos etarios respecto al ajuste de los modelos lineales. Por una parte, en el Gráfico 4 se observa que en el caso de los hombres, el modelo del grupo etario mayor (de 55 a 65 años) es el que mejor se ajusta al modelo lineal con un  $R^2 = 42\%$ ; mientras que en el caso de las mujeres, el mejor ajuste se observa en el grupo etario menor (de 13 a 24 años) con un  $R^2 = 42.7\%$ . Por otra parte, en el gráfico 5 se muestra un patrón: el ajuste de los modelos tiene mayor poder de explicación en el grupo etario del medio con un  $R^2 = 59.9\%$  para los hombres y de  $R^2 = 54.9\%$  para las mujeres. Estos resultados se presentan en escala logarítmica con el fin de poder observar que la distribución de los sesgos es bastante lineal. En consecuencia, en la siguiente sección se estiman modelos de regresión que adoptan una forma lineal para la evaluación analítica de sesgos.

Esta evaluación de sesgos pone de manifiesto la importancia de contrastar los datos de expatriados de Facebook con definiciones de inmigración distintas, pues dependiendo del concepto empleado, los resultados son diferentes. Asimismo, resulta relevante considerar combinaciones de grupos etarios distintos y análisis separados por sexo para dimensionar detalladamente el comportamiento de los sesgos.

### Gráfico 4

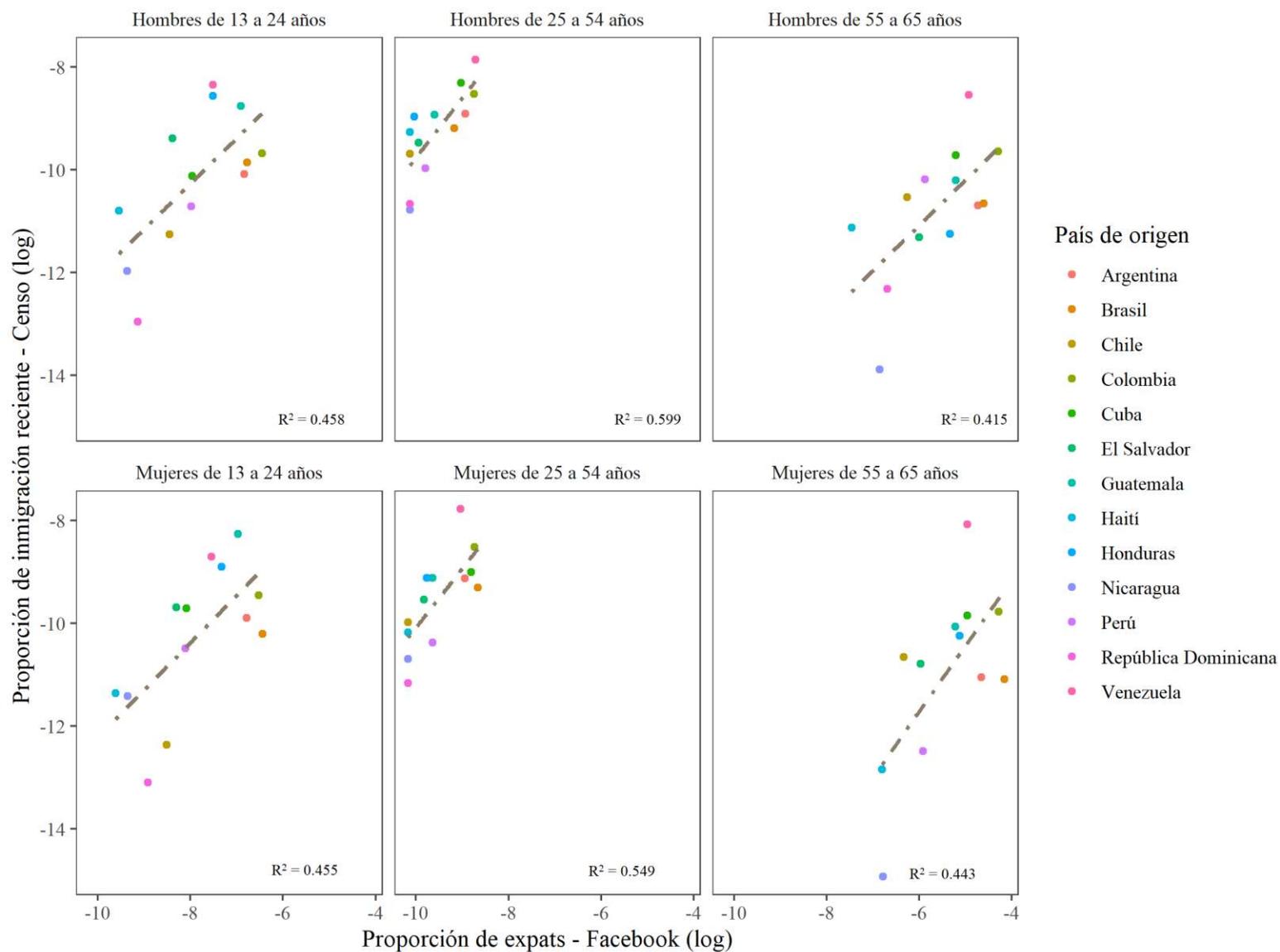
Relación entre la proporción de inmigración por país de nacimiento del Censo 2020 con la proporción de expatriados de Facebook 2020 en escala logarítmica



Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020

### Gráfico 5

Relación entre la proporción de inmigración por país de procedencia del Censo 2020 con la proporción de expatriados de Facebook 2020 en escala logarítmica

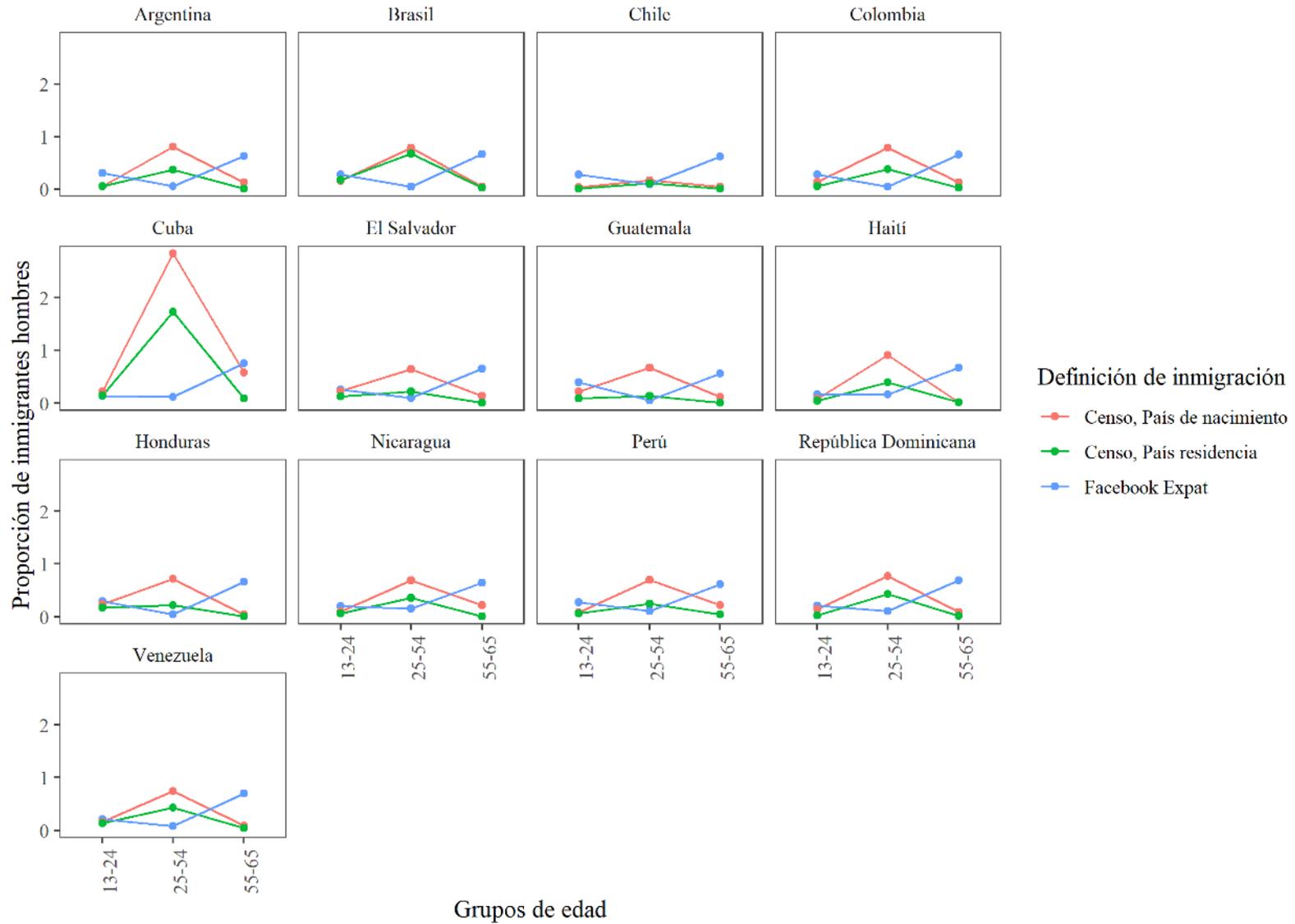


Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020

Luego de hacer una evaluación inicial de los sesgos en los stocks de inmigración, los gráficos 6 y 7 muestran una forma intuitiva de identificar sus diferencias. Estos gráficos, a diferencia de los anteriores, muestran los stocks de inmigración en términos proporcionales y no en escala logarítmica; además muestra en un solo gráfico las dos definiciones de inmigración del Censo 2020 y la de Facebook. De aquí se observa que hay un patrón tanto en hombres como en mujeres, de una brecha relativamente menor entre la inmigración por país de procedencia y los datos de expatriados de Facebook, que entre la definición de inmigración por país de nacimiento y los datos de esta red social. Este caso se hace particularmente evidente en el grupo etario de 25 a 54 años en países como Chile, El Salvador, Guatemala y Perú, donde el stock de inmigración por país de procedencia (línea verde) se une con el stock de expatriados de Facebook (línea azul). Otro aspecto sobresaliente de estos gráficos es que el stock de expatriados del grupo etario de 55 a 65 años es mucho mayor a las definiciones de inmigración del Censo 2020. Este hallazgo difiere del estudio de Zagheni et al (2017), donde la proporción de Facebook en edades más avanzadas es sistemáticamente menor; sin embargo, es preciso decir que los grupos etarios empleados en dicho estudio son distintos a los que se consideran en esta tesis.

## Gráfico 6

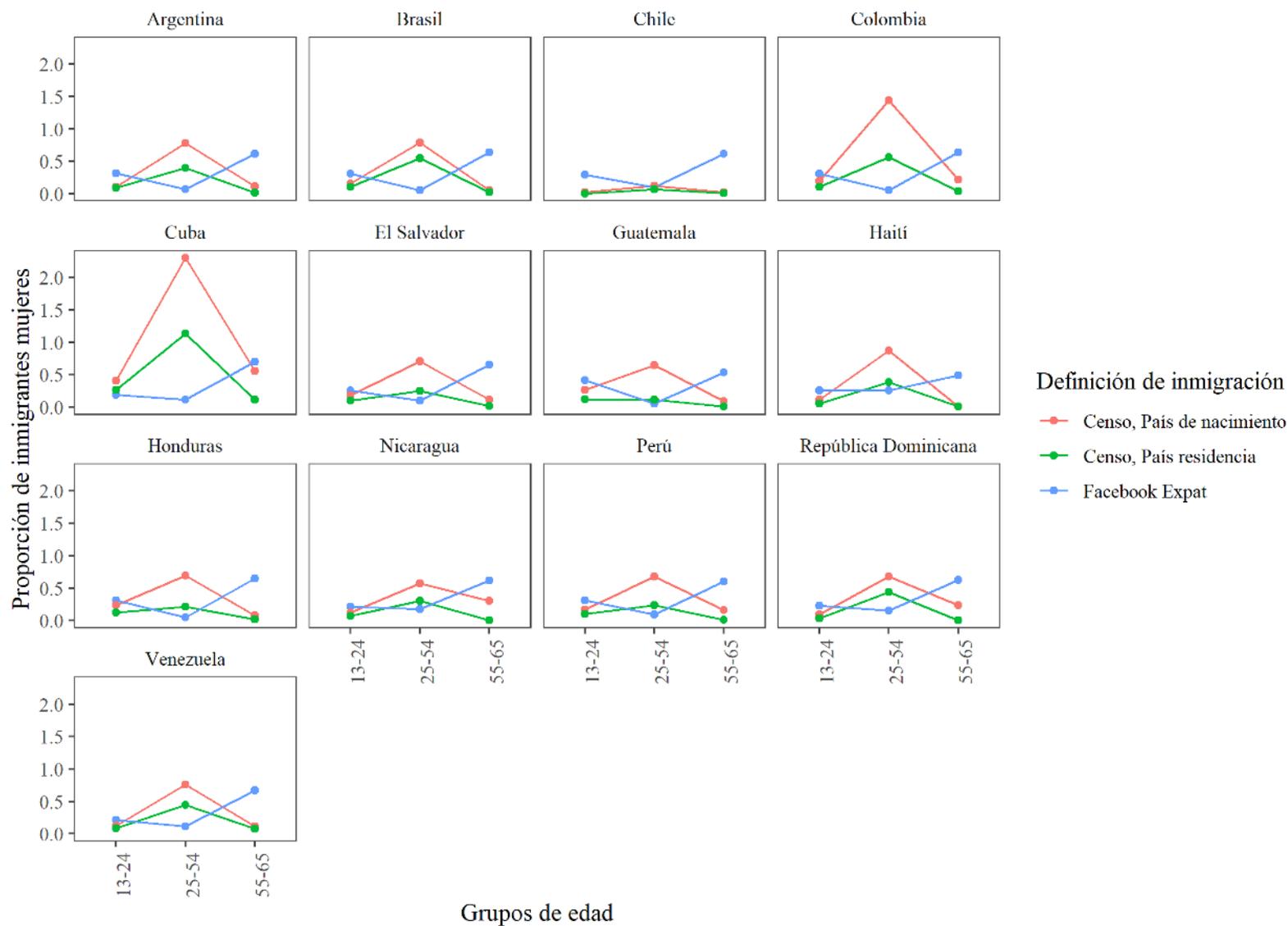
Proporción de inmigración de hombres en México según datos del Censo 2020 y Facebook 2020



Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020

## Gráfico 7

Proporción de inmigración de mujeres en México según datos del Censo 2020 y Facebook 2020



Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020

El sesgo de selección es la razón principal detrás de las diferencias en las proporciones de los stocks de inmigración de Facebook y del Censo 2020. De acuerdo con Zagheni & Weber (2015), este tipo de sesgo es la principal limitación que se tiene al trabajar con datos de Internet y redes sociales en línea; por lo tanto, si no se resuelve el sesgo de selección, es probable que las investigaciones conduzcan a resultados poco fiables. Como consecuencia de este tipo de sesgo, es más probable que en países como México las mujeres usen Facebook, a diferencia de otros países como Arabia Saudita (Kashyap et al., 2020). Adicionalmente, están los sesgos que se derivan del funcionamiento intrínseco de Facebook, y de los cuales no hay información pública, como los cambios que efectúan en sus algoritmos para alcanzar las audiencias de interés publicitario o definir los criterios con los que se crea la variable expatriado (Rampazzo et al., 2021).

Otras razones que pueden influir en los sesgos de los datos de Facebook pueden estar relacionados con el comportamiento de sus usuarios. Es probable que un usuario tenga más de una cuenta activa (Rampazzo et al., 2021) o que no registre información relevante en Facebook respecto a sus características demográficas, lugar de nacimiento o los cambios de residencia a países distintos al de nacimiento (Zagheni et al., 2017).

A pesar de la existencia de estos sesgos, se podría esperar una correlación relativamente alta entre las audiencias de Facebook y los datos de fuentes tradicionales como el Censo 2020, puesto que en el modelo de negocio de Facebook uno de los principales activos es tener información lo más detallada y real de sus usuarios para realizar segmentaciones de mercado que sean de utilidad para las empresas que buscan hacer publicidad por medio de esta red social en línea. De este modo, es de esperar que Facebook tenga el incentivo de captar de la forma más precisa las características demográficas y de comportamiento de sus usuarios.

### 3.3.2 Evaluación analítica

Para finalizar el análisis de sesgos, a continuación, se presentan los resultados de los modelos donde se compara la definición de inmigración por país de residencia en 2015 con los datos de expatriado de Facebook. Como se mencionó antes, estos modelos están inspirados en el trabajo de Zagheni et al., (2017), del que se replica la metodología de estimación que se presenta aquí. No obstante, en esta investigación se propone incluir la variable sexo como variable independiente y posteriormente se presentan modelos desagregados por sexo. La finalidad de estos modelos es ofrecer una evaluación sistemática y analítica de los sesgos que resultan al comparar los datos de ambas fuentes de información; gracias a que los coeficientes que resultan de estos modelos indican la dirección en la que se presentan los sesgos.

Como se anticipó en la sección de Métodos, para identificar la definición de inmigración a la que mejor se ajustan los datos de Facebook se estimaron 24 modelos generales y 36 modelos separados por sexo. Estas estimaciones se presentan en el Apéndice 2. El criterio para seleccionar la definición de inmigración a la que mejor se ajustan los datos de expatriados de Facebook fue la del R-cuadrado ajustado. De aquí que la definición de inmigración escogida haya sido la de migración por país de residencia en 2015, con un R-cuadrado ajustado de 80.6 %. Por tanto, en esta sección se presentan cuatro modelos generales (Cuadro 6) y seis modelos separados por sexo (Cuadro 7), con base en la definición de inmigración seleccionada. La evaluación analítica se corresponde con la evaluación gráfica que se presentó en la sección anterior, en la cual se muestra que los datos de expatriados se aproximan más a los datos de inmigración por país de residencia previa.

Las variables independientes que se emplearon son: proporción de expatriados en escala logarítmica, sexo, edad y país de origen. Estas variables se añadieron una a una conforme se estimaron los modelos, de aquí, que el Modelo 1 tenga una variable independiente y el Modelo 4 tenga cuatro. Con base en el criterio de información de Akaike, se tiene que el Modelo 4 tiene una bondad del ajuste relativamente mejor a los otros tres. Por consiguiente, se infiere que añadir características de las personas como sexo, edad y el país del que proceden los migrantes incrementa la capacidad de explicación del modelo, del que se obtiene un R-cuadrado ajustado de 80.6 %.

Para entender la forma en la que se comportan los sesgos derivados de los datos de Facebook, es preciso observar los coeficientes del Modelo 4 (Cuadro 3). Los coeficientes con signo negativo indican una sobrestimación, mientras que los coeficientes positivos indican una subestimación en los stocks de inmigración en México. Por consiguiente, de acuerdo con estos resultados, en Facebook hay una sobrestimación en el grupo etario de 55 a 65 años. Análogamente, los coeficientes positivos y significativos de los países de origen indican una subestimación de los datos de Facebook; tal es el caso de Cuba, El Salvador, Guatemala, Honduras y Venezuela. Los resultados de los grupos etarios difieren del caso de Estados Unidos que presenta Zagheni et al., (2017), donde los coeficientes de los grupos de edad mayores (55 a 54 y 55 a 59 años) son positivos, lo que implica una subestimación de estas poblaciones en Facebook. Sin embargo, debido a que el coeficiente de la variable sexo no resulta significativo, no es posible concluir la forma en la que operan los sesgos. Esto motivó la estimación de los modelos desagregados por sexo.

Además de que este modelo da información sobre los sesgos de los datos de expatriados de Facebook cuando se evalúan a luz de los datos de inmigración reciente del Censo 2020, proporciona los coeficientes necesarios para estimar stocks de inmigración específicos con base en datos exclusivamente de Facebook. Para desarrollar esta idea, resulta ilustrativo plantearse el caso hipotético en el que se requiera conocer el stock de inmigración de mujeres de 25 a 54 años que anteriormente residían en Venezuela y que en agosto de 2022 residen en México ( $y_{e,s,o}$ ). Para esto, se deberá contar con extractos de esa misma fecha para usuarias mensuales de la red social que sean mujeres expatriadas que anteriormente vivían en Venezuela de 25 a 54 años ( $x_{e,s,o}$ ) e introducir este valor en la siguiente ecuación:

$$y_{e,s,o} = -4.856 + 0.712x_{e,s,o} - 0.085 + 2.004$$

De esta estimación, se obtendrá un stock de inmigración específico fuertemente relacionada con los datos de inmigración reciente del Censo 2020, debido a que el modelo presenta un R-cuadrado ajustado de 80.6 %. De este modo, se pueden obtener stocks de inmigración en tiempo real, basados en MAU máximos de expatriados de Facebook, los cuales se pueden extraer diariamente por medio de su API.

La segunda parte de la evaluación analítica consistió en estimar seis modelos desagregados por sexo, con el propósito de entender la forma en la que se comportan los sesgos en hombres y mujeres por separado (Cuadro 4). Para esto, se siguió la misma lógica de los modelos generales, en los que se fueron sumando variables independientes. De acuerdo con el criterio de Akaike, tanto para hombres como para mujeres, el modelo que presentó una bondad del ajuste relativamente mejor a los otros, fueron los Modelos 3 y 6 que se muestran en el Cuadro 4. Asimismo, la dirección de los signos en los coeficientes de los grupos etarios se mantiene como en el modelo general. Por esta razón, tanto en hombres como mujeres inmigrantes en México están subrepresentados en Facebook.

No obstante, respecto a los países de origen se tiene que los coeficientes significativos únicamente son positivos. Los países que muestran una subestimación de Facebook respecto al Censo 2020, en el caso de los hombres son Guatemala y Venezuela; y, en el caso de las mujeres, Guatemala, Honduras y Venezuela. En consecuencia, el modelo de los hombres presenta una capacidad de explicación relativamente mayor que el de las mujeres, con un R-cuadrado ajustado de 79.1 %, respecto a un R-cuadrado ajustado de 76.1 % en el caso de las mujeres.

Los resultados de estos modelos si bien aportan información específica de la dirección de los sesgos en la variable edad y país de origen, no aporta información estadísticamente significativa para el coeficiente de expatriados; por tanto, los coeficientes del modelo en su conjunto no aportan información robusta para hacer estimaciones de stocks de población inmigrante independientes a una fuente de información tradicional como el Censo 2020; en contraposición a los stocks en tiempo real que sí se pueden estimar con el modelo general.

**Cuadro 3**

*Estimación de la proporción de inmigración por lugar de residencia en 2015 (log), considerando como variables independientes la proporción de expatriados de Facebook (log), sexo, edad, y país de origen*

	Variable dependiente: Proporción de inmigración por país de residencia en 2015 (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	-0.073 (0.087)	-0.077 (0.087)	1.016*** (0.129)	0.712*** (0.231)
Mujer		-0.187 (0.322)	0.013 (0.217)	-0.085 (0.150)
25 a 54 años			2.505*** (0.338)	2.004*** (0.417)
55 a 65 años			-3.056*** (0.404)	-2.392*** (0.567)
Brasil				-0.216 (0.359)
Chile				0.277 (0.496)
Colombia				0.476 (0.364)
Cuba				0.885** (0.377)
El Salvador				0.825* (0.460)
Guatemala				1.055*** (0.371)
Haití				0.555 (0.610)
Honduras				0.953** (0.392)
Nicaragua				-0.923 (0.576)
Perú				0.020 (0.434)
República Dominicana				-1.046* (0.543)
Venezuela				1.960*** (0.363)
Constante	-10.727*** (0.684)	-10.474*** (0.814)	-2.227** (1.048)	-4.856*** (1.590)
Observaciones	77	77	77	77
R <sup>2</sup>	0.009	0.014	0.567	0.846
R <sup>2</sup> ajustado	-0.004	-0.013	0.543	0.806
AIC	274.7123	276.3619	217.0173	161.1651
BIC	281.7437	285.7371	231.0801	203.3536
Error estándar residual	1.404 (df = 75)	1.410 (df = 74)	0.947 (df = 72)	0.618 (df = 60)
Estadístico F	0.717 (df = 1; 75)	0.524 (df = 2; 74)	23.558*** (df = 4; 72)	20.673*** (df = 16; 60)

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020.

**Cuadro 4.**

*Estimación de la proporción de inmigración por lugar de residencia en 2015 (log), considerando como variables independientes la proporción de expatriados de Facebook (log), edad, y país de origen*

	Variable dependiente: Proporción de inmigración por país de residencia en 2015 (log)					
	(1)	Hombres		(4)	Mujeres	
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	-0.109 (0.113)	0.888*** (0.176)	0.668* (0.347)	-0.041 (0.136)	1.153*** (0.193)	0.768* (0.388)
25 a 54 años		2.394*** (0.464)	2.029*** (0.623)		2.626*** (0.503)	1.997*** (0.695)
55 a 65 años		-2.741*** (0.559)	-2.206** (0.877)		-3.368*** (0.598)	-2.610*** (0.913)
Brasil			-0.255 (0.510)			-0.160 (0.596)
Chile			0.434 (0.728)			0.133 (0.818)
Colombia			0.427 (0.503)			0.518 (0.610)
Cuba			0.841 (0.522)			0.938 (0.635)
El Salvador			0.667 (0.654)			0.999 (0.774)
Guatemala			0.922* (0.521)			1.189* (0.616)
Haití			0.881 (0.871)			0.261 (1.044)
Honduras			0.711 (0.536)			1.213* (0.670)
Nicaragua			-0.998 (0.844)			-0.826 (0.961)
Perú			0.336 (0.622)			-0.286 (0.721)
República Dominicana			-0.966 (0.763)			-1.173 (0.959)
Venezuela			1.905*** (0.511)			2.009*** (0.602)
Constant	-10.904*** (0.878)	-3.255** (1.396)	-5.352** (2.452)	-10.569*** (1.088)	-1.080 (1.569)	-4.521 (2.795)
Observaciones	39	39	39	38	38	38
R <sup>2</sup>	0.025	0.552	0.873	0.003	0.589	0.858
Adjusted R <sup>2</sup>	-0.002	0.514	0.791	-0.025	0.553	0.761
AIC	136.39075	110.02807	84.76393	143.13631	113.39840	97.13743
BIC	141.3814	118.3459	113.0445	148.0491	121.5863	124.9764
Residual Std. Error	1.322 (df = 37)	0.921 (df = 35)	0.604 (df = 23)	1.511 (df = 36)	0.997 (df = 34)	0.730 (df = 22)
F Statistic	0.935 (df = 1; 37)	14.389*** (df = 3; 35)	10.578*** (df = 15; 23)	0.091 (df = 1; 36)	16.275*** (df = 3; 34)	8.841*** (df = 15; 22)

Nota:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020.

**CAPÍTULO IV:**  
**DISCUSIÓN**

En el estudio de la migración internacional, desde un enfoque cuantitativo, persisten limitaciones en las fuentes de información para contar con una perspectiva actualizada y detallada de las poblaciones migrantes. Las limitaciones se pueden ver al menos en dos sentidos. Por una parte, están las limitaciones relacionadas con el tiempo o la periodicidad en la que los datos sobre migración son recabados, pues para dimensionar este fenómeno con detalle, transcurren diez años entre cada censo de población. Además, estos instrumentos de medición, que son por excelencia el estándar de oro para obtener información sobre la población migrante en un país, capturan información universal de la inmigración; pero no capturan en su totalidad la emigración como fenómeno, puesto que en el caso mexicano, las preguntas sobre emigración se incluyen en el cuestionario ampliado que se le aplica únicamente a una muestra de la población. De aquí que dimensionar en su totalidad el fenómeno migratorio sea complejo.

Por otra parte, están las limitaciones relacionadas con los motivos de la migración. Cuando la migración es irregular dimensionar movilidad es aún más desafiante como consecuencia del subregistro de movimientos que ocurren en esta condición. Asimismo, cuando se está ante situaciones de migración forzada por causas de desplazamiento o desastres de origen natural, existen obstáculos relacionados con la logística y la operación en terreno para cuantificar el fenómeno. Las limitaciones en función de la frecuencia en el levantamiento de información y de las motivaciones de la migración, redundan en agregar complejidad a la medición de la migración oportunamente, en comparación con otros fenómenos demográficos.

Ahora, este contexto particular está intrínsecamente relacionado con las fuentes de información tradicionales, que están diseñadas específicamente para obtener información demográfica basada en criterios de calidad y gestionada, tanto en su recolección de datos como en su análisis, por capital humano calificado. Por ende, es natural que el acceso a esta información no se pueda hacer de forma inmediata y continua en el tiempo. Aunque la información de las fuentes de información tradicionales mantiene altos estándares de calidad para el estudio de los fenómenos demográficos, no deja de existir la necesidad de contar con recursos de información más sensibles en tiempo real; de ahí que las fuentes de información derivadas de las interacciones que ocurren en el espacio web se posicionen como recursos viables para el estudio de la migración internacional.

Para incluir fuentes de información innovadoras como las redes sociales en línea en el estudio de la migración internacional, es necesario, en primer lugar, contar con fuentes de información tradicionales que operen como base para contrastar los datos de las redes sociales; y en segundo lugar, realizar un análisis de sesgos de las características de las personas, para asegurar la producción de estimaciones y proyecciones de población robustas desde el punto de vista estadístico. En consecuencia, los objetivos de esta tesis se encaminaron en este sentido; aprovechando el más reciente Censo de Población y Vivienda que tiene México, y las cifras de usuarios de Facebook de este mismo país recogidas en 2020, en esta investigación se analizó la validez de los datos de Facebook para el estudio de la inmigración en México.

La estrategia adoptada para atender a este objetivo consistió en comprender en qué medida el uso de internet y Facebook son recursos generalizados en las poblaciones migrantes de México. Esto en el entendido del valor que tienen ambos recursos en el proyecto migratorio (Moran, 2022). Los resultados para el caso mexicano muestran que las proporciones de acceso a internet difieren en función del sexo, la edad y el país de procedencia. En este sentido, se encontró que en promedio las mujeres migrantes tienen mayor acceso a internet que los hombres; que el grupo etario con proporciones de acceso a internet más altas es el de 25 a 54 años; y que las poblaciones migrantes procedentes de Chile, Venezuela, Argentina y Colombia son, en este orden, las que muestran los niveles más altos de acceso a internet; mientras que las poblaciones migrantes procedentes de Haití, Guatemala y El Salvador son las que ostentan los menores niveles de acceso. Asimismo, se encontró que, en promedio, las personas migrantes que llevan más tiempo en México tienen proporciones de acceso a internet más altas que los inmigrantes recientes.

En lo que concierne a la penetración de Facebook en México para la población total, los resultados son consistentes con la literatura antecedente donde, en promedio, los hombres tienen mayor acceso a Facebook en comparación con las mujeres (Garcia et al., 2018; Kashyap et al., 2020). Sin embargo, para el caso de México, este patrón se invierte cuando se trata de la población inmigrante, pues, son las mujeres quienes tienen razones de acceso a Facebook más altas. Los argumentos detrás de la brecha en el acceso a Facebook entre hombres y mujeres radican en códigos culturales (Kashyap et al., 2020) y aspectos estructurales de inequidad en términos de educación, salud, participación política y oportunidades económicas (Garcia et al., 2018). Si bien esta investigación no ahonda en estos temas, se reconoce la relevancia de trazar proyectos de

investigación encaminados a entender las razones estructurales en el acceso y representación en Internet de hombres y mujeres migrantes en los países de destino.

En segundo lugar, gracias a que en esta investigación se utilizó el Censo 2020, se pudieron delimitar seis definiciones de inmigración basadas en interacciones entre la definición de inmigración por país de nacimiento (absoluta) y la migración por país de procedencia (reciente). El definir seis posibles definiciones de inmigración para comparar la variable expatriado de Facebook solo pudo ser posible al tener una fuente de información tradicional como un censo de población. Además permitió captar de manera amplia y desagrega las variables de sexo, edad y país de origen, que sirvieron para establecer las seis definiciones que se compararon con los datos de Facebook que, por su parte, únicamente cuenta con una posible definición de inmigración, basado en la variable de expatriado.

Adicionalmente, el Censo 2020 representa, como se ha mencionado antes, un estándar de oro respecto a la medición de la migración internacional; y es la primera vez —al menos hasta donde pudo llegar la revisión de literatura de este estudio— que se realiza un análisis de la validez de la población de expatriados de Facebook con base en un censo de población. Otros estudios han empleado encuestas, registros administrativos o información armonizada sobre migración de organismos internacionales (Alexander et al., 2020; Palotti et al., 2020; Rampazzo et al., 2021; Spyrtos et al., 2019; Zagheni et al., 2017).

En Facebook una persona expatriada se define como aquella persona que habitualmente vivía en un país (por ejemplo, Venezuela) y ahora vive en el extranjero (por ejemplo, en México), lo que configura la variable “Vivió en Venezuela. Antiguos expatriados – Venezuela”. El análisis de esta variable de Facebook, respecto a las seis definiciones de inmigración, arrojó que la inmigración por lugar de residencia previa derivada del Censo 2020 es la que muestra un mejor ajuste en términos estadísticos. Este resultado abona la discusión en torno a qué tipo de movimientos internacionales capta Facebook, pues si bien, la intuición aproximaba la variable expatriado a la definición de inmigración reciente, esta hipótesis no se había contrastado. Como alternativa, en los estudios de Alexander et al. (2020); Palotti et al. (2020); Rampazzo et al. (2021); Spyrtos et al. (2019); Zagheni et al. (2017), se empleó la definición de inmigración por país de nacimiento, pues era el indicador del que podían disponer los autores.

En tercer lugar, para comprender en qué medida los datos de Facebook pueden ser utilizados como insumo para el estudio de la migración internacional, en la literatura especializada se han empleado técnicas de estadística inferencial (Palotti et al., 2020; Spyrtos et al., 2019; Zagheni et al., 2017) y bayesiana (Alexander et al., 2020; Rampazzo et al., 2021). La estadística inferencial se ha utilizado con el objetivo de contrastar los datos de Facebook con una fuente de información tradicional y encontrar sesgos específicos en variables demográficas; mientras que la estadística bayesiana se ha empleado para combinar fuentes de información tradicionales e innovadoras para completar un conjunto de datos, con el propósito de estimar y proyectar stocks de inmigración en un país. Teniendo en cuenta estos dos marcos estadísticos, la presente investigación se basó en la estadística inferencial; puesto que el análisis consistió en contrastar dos fuentes de información; y así realizar estimaciones independientes de coeficientes de stocks de inmigración con base en datos de Facebook.

En los hallazgos en términos de patrones demográficos de esta investigación, se encontró que los coeficientes de los stocks de inmigración de hombres y mujeres en edades de 55 a 64 años presentan una sobrestimación. Los signos negativos en los coeficientes de este grupo etario difieren de los resultados de Zagheni et al. (2017), donde los stocks de inmigración en las edades de 35 a 39 años y 55 a 59 años, presentaban coeficientes positivos en el caso de Estados Unidos. Un posible argumento para explicar estas diferencias puede estar vinculado el sesgo de selección (Zagheni & Weber, 2015), que, entre otros aspectos, refleja los códigos culturales y sociales de una población en la esfera digital.

En relación con los coeficientes significativos de los países de origen, se tiene que Cuba, El Salvador, Guatemala, Honduras y Venezuela muestran una subestimación de los datos de Facebook respecto al Censo 2020. Esto podría explicarse por la cantidad de población que hay en cada grupo etario. Las edades del medio (25 a 54 años), que agrupan 29 años, tienen menor volumen de población de expatriados en comparación con los inmigrantes del Censo 2020. De aquí que en los gráficos 6 y 7 la línea azul que representa a Facebook adopte forma de “V”. Sin embargo, los grupos etarios de los extremos (13 a 24 y 55 a 65 años), agrupan once y diez años respectivamente, con mayor volumen de expatriados.

Esto implica que solo en el grupo del medio Facebook está captando menos expatriados que el Censo 2020. Esto podría representar un sesgo en la forma que Facebook distribuye la

población de expatriados por grupos etarios, puesto que, para el caso de México captó menor volumen de población en una agrupación de edades más amplia en contraste con agrupaciones de edades más acotadas.

Un aspecto diferenciador de esta investigación es haber estimado, por una parte, modelos generales y, por otra, modelos desagregados por sexo. Esto permitió entender que el modelo general tiene un poder de explicación de 80.6 % basado en su R-cuadrado ajustado; y que el poder de explicación del modelo estimado para los hombres es mayor que el de las mujeres (R-cuadrado ajustado de 79.1 % contra 76.1 %, respectivamente). Adicionalmente, haber estimado un set de modelos generales y otro de modelos separados por sexo, permitió encontrar que, para el caso mexicano, los coeficientes destinados a la estimación de stocks de inmigración en tiempo real se deben hacer con base en el modelo agregado, donde la variable sexo se incluye como variable independiente.

Las limitaciones de esta investigación se presentan en dos vías. Por un lado, la literatura sobre demografía digital es relativamente reciente, esto hace que sea escasa, particularmente en investigaciones para países de América Latina. Asimismo, los estudios enfocados en la medición de la migración internacional basados en *big data* emplean fuentes de información diversas, desde redes sociales en línea hasta registros de compradores en tiendas de *retail*; por lo tanto, surge la dificultad de comparar los resultados de esta investigación con estudios en los que se utilice específicamente Facebook para la evaluación de sesgos en variables demográficas. Por otro lado, una dificultad adicional se relaciona con el hecho de que a la información de Facebook no se puede acceder de forma retrospectiva; esto implicó que no se contara con datos de expatriados de Estados Unidos (que representan la mayor proporción de inmigración en México), dado que, la concepción de esta tesis es posterior a las extracciones de datos de Facebook realizadas en marzo de 2020. Asimismo, la información con la que se contaba de 2020 no permitió realizar nuevos grupos etarios.

Para enfrentar dichas dificultades en futuras investigaciones, se introducen dos recomendaciones encaminadas a reducir —al menos en parte— las brechas que en la actualidad existen en los estudios sobre demografía digital en América Latina. En primer lugar, se anima a que la investigación en demografía aúne esfuerzos con la investigación en ciencia de datos y

computacionales enfocándose en los principales retos que enfrenta hoy en día el estudio de la dinámica poblacional; para esto se requiere un acercamiento técnico de los investigadores en demografía a temas relacionados con *big data*. Estas alianzas de investigación académica, también se verían enriquecidas con alianzas con empresas que recaban y producen la información digital. En segundo lugar, se recomienda impulsar una cultura de *big data* que permita entender y extraer información digital de forma constante. Esto con el propósito de construir conjuntos de datos basados en entender mejor las características de las personas en tiempo real y hacer proyecciones de población migrante de corto plazo (*nowcasting*).

Como aspectos intrínsecos a la estrategia metodológica y los resultados de esta tesis, se recomienda que, en una réplica o ampliación, se profundice en la *desagregación de las variables demográficas*. Por ejemplo, que en el caso de la edad se opte por grupos de edades quinquenales; y que se amplíe la evaluación de los sesgos al nivel estatal o municipal en miras al censo de población que se realizará en México en 2025. En este sentido, es preciso adelantar extractos de datos de población expatriada que permitan calcular stocks y proporciones de inmigración, así como la extracción de la población total de Facebook en línea con las características demográficas específicas a evaluar (edad, sexo, país de origen, estado o municipio). Asimismo, es preciso mantener un análisis del contexto migratorio de México o del país de destino donde se replique esta investigación, para captar la población migrante. Adicionalmente, se debe hacer un análisis constante de los datos que se extraen de Facebook, con el fin de identificar si en el futuro surgen cambios en el algoritmo que identifica a los expatriados.

## CONCLUSIONES

Esta investigación analizó la validez de los datos de Facebook para estudiar la inmigración en México. Para esto se contrastaron datos de Facebook con datos del Censo de Población y Vivienda 2020 de México con el propósito de desarrollar dos objetivos: primero, comprender en qué medida el uso de internet y Facebook son recursos generalizados en las poblaciones migrantes de México; y segundo, analizar la validez de la definición de expatriado de Facebook y el alcance de esta red social para captar y medir poblaciones migrantes en tiempo real. En consecuencia, se esperaba encontrar diferencias en las mediciones desagregadas por edad, sexo y país de origen en los datos del censo y en las cifras de usuarios de la mencionada red social; y, con base en esta información, entender en qué medida Facebook podría utilizarse como un censo digital para hacer estimaciones de stocks de inmigración.

En primer lugar, resulta valioso decir que, efectivamente, se evidencian diferencias en los indicadores de acceso a internet, penetración de Facebook y en los coeficientes estimados de los stocks de inmigración en función de la edad, el sexo y el país de origen de la población objeto de estudio. A pesar de las diferencias observadas, se encontró que el uso de internet y Facebook resultan ser recursos generalizados entre la población migrante en México. Este resultado proporcionó las condiciones necesarias para desarrollar la segunda parte de la investigación, pues, se partió del supuesto de que el espacio digital es un ámbito relevante para las poblaciones migrantes.

Con los datos censales se consideraron seis definiciones de inmigración para evaluar las cifras de usuarios de Facebook etiquetados como expatriados, de las que se obtuvo que los datos de Facebook se asemejan más a la definición de inmigración por país de residencia previa que a la definición de migración por país de nacimiento. Este hallazgo aporta evidencia estadística concluyente respecto a la hipótesis presente en la literatura sobre migración digital que anticipaba que los usuarios etiquetados como expatriados en Facebook se aproximarían más a una definición de inmigración reciente basada en el país de residencia previa. Sin embargo, hasta la fecha no se había comprobado esta hipótesis mediante el contraste de esta y otras definiciones de inmigración con base en un censo de población.

En segundo lugar, esta investigación se propuso evaluar la existencia de sesgos en características como la edad, el sexo y el país de origen de las personas migrantes. Con base en el modelo general, se encontró que hay una sobrestimación, tanto en hombres como en mujeres en el grupo etario más avanzado (de 55 a 65 años). En general los signos positivos de los coeficientes significativos indican una subestimación de la inmigración, como es el caso de las poblaciones procedentes de Cuba, El Salvador, Guatemala, Honduras y Venezuela; mientras que el coeficiente negativo y significativo de República Dominicana indica una sobrestimación en los stocks de expatriados de Facebook con respecto a los stocks de inmigrantes del Censo 2020.

Para comprender el efecto de la variable sexo, se estimaron modelos separados para hombres y mujeres, concluyéndose que existen efectos distintos en esta variable. En el modelo de los hombres, los coeficientes positivos y significativos que indican una sobrestimación en los países de procedencia son Guatemala y Venezuela; mientras que en el caso de las mujeres la sobrestimación de esta variable se observó en Guatemala, Honduras y Venezuela. El comportamiento de la variable edad fue equivalente al del modelo general, en el que los coeficientes mostraron una sobrestimación (signo negativo) en el grupo etario mayor (de 55 a 65 años). En conjunto, las variables independientes del modelo de los hombres tienen un poder de explicación mayor que el de las mujeres, con un R-cuadrado ajustado de 79.1 %, respecto a un R-cuadrado ajustado de 76.1 %.

Luego de la evaluación de sesgos en el conjunto de modelos agregados y el conjunto de modelos desagregados por sexo, se concluyó que para obtener stocks de inmigración en tiempo real con base en los datos de expatriados de Facebook, se debe utilizar el modelo general que tiene como variables independientes la proporción de expatriados (en escala logarítmica), el sexo, la edad y los países de origen; puesto que, en este modelo, a diferencia de los modelos separados por sexo, la variable que capta los expatriados de Facebook sí resulta significativa. De este modo, se obtienen los coeficientes de una función lineal que permiten estimar el stock de inmigración en tiempo real, basado en un R-cuadrado ajustado de 80.6 %.

Como reflexión final, resulta útil indicar los aportes específicos de esta tesis a los estudios de demografía en América Latina. En primer lugar, se retoma la discusión sobre la calidad y disponibilidad de las fuentes de información. Esto condujo a realizar una evaluación sistemática

de uno de los recursos en línea más utilizados hoy en día como lo es Facebook, con el propósito de abonar la discusión sobre la calidad y la disponibilidad de las fuentes de información en demografía. Mediante dicha evaluación se despejaron interrogantes respecto a la validez del uso de esta red social para el estudio de la inmigración en México, la forma en la que operan los sesgos en las variables demográficas y la posibilidad de contar con información en tiempo real de comunidades migrantes. Resulta razonable decir entonces que, en la actualidad la demografía se puede apoyar en los datos públicos de esta red social en línea para responder preguntas clásicas — y potencialmente nuevas—. De este modo, se minimiza una de las principales dificultades para el estudio de migración internacional: la disponibilidad constante de datos.

En segundo lugar, se presenta una estrategia metodológica de evaluación de sesgos para producir coeficientes de stocks de inmigración que, guardando las proporciones de trabajar directamente con un censo de población, permiten producir panoramas actualizados de la migración internacional en México en periodos intercensales. Adicionalmente, esta metodología presenta una alternativa para abordar el contexto de países que experimentan la ausencia de fuentes de información tradicionales (como censos de población o encuestas de hogares), pues pone en el panorama de los datos una fuente de información *constantemente disponible* como son, hasta el momento, los datos de Facebook.

La demografía es una disciplina que tiene como uno de sus principales pilares el análisis cuantitativo. En ese sentido, mantener una apertura hacia las nuevas fuentes de información y datos disponibles es vital. La discusión sobre qué significa *big data* para la demografía se seguirá perfilando en el tiempo. Sin embargo, hasta este momento, se puede decir que esta constituye una fuente de información que permite obtener datos con mayor frecuencia, gracias a su amplia disponibilidad en Internet, principalmente. Esto permite analizar fenómenos demográficos que con datos tradicionales son más complejos de realizar, ya que se requiere esperar la producción de datos de las oficinas de estadística de cada país; o implementar instrumentos de medición directamente por quienes investigan. Lo anterior supone una inversión significativa de recursos en términos de tiempo, dinero y capital humano. Si bien *big data* no se constituye en un sustituto perfecto de las fuentes tradicionales, sí amplía el espectro de los datos y de las fuentes de información que se emplean para el estudio de las ciencias sociales.

## REFERENCIAS

- Alexander, M., Polimis, K., & Zagheni, E. (2020). Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States. *Population Research and Policy Review*, 41(1), 1–28. <https://doi.org/10.1007/s11113-020-09599-3>
- Bosco, C., Grubanov-Boskovic, S., Iacus, S.M., Minora, U., Sermi, F., & Spyrtatos, S. (2022). *Data Innovation in Demography, Migration and Human Mobility* (L. Publications Office of the European Union, Ed.). <https://doi.org/10.2760/027157>
- Castillo, M. Á. (2014). La inmigración, 2000-2010. En C. Rabell Romero (Ed.), *Los mexicanos. Un balance del cambio demográfico* (pp. 513–560). Fondo de Cultura Económica.
- Castillo, M. Á., & Rojas Wiesner, M. L. (2020). Un balance de la inmigración y de la migración de tránsito en México. En S. Giorguli & J. Sobrino (Eds.), *Dinámica demográfica de México en el siglo XXI* (Vol. 2, pp. 131–168). El Colegio de México.
- Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). *Promises and Pitfalls of Using Digital Traces for Demographic Research*. <https://doi.org/10.1007/s13524-018-0715-2>
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: A data management perspective. *Frontiers of Computer Science*, 7(2), 157–164. <https://doi.org/10.1007/s11704-013-3903-7>
- Diccionario Panhispánico del Español Jurídico. (2022, May). *Big data*. <https://Dpej.Rae.Es/Lema/Big-Data>.
- COMAR. (2018). *Boletín Estadístico de Solicitantes de Refugio en México*. chrome-extension://efaidnbmninnibpcajpcglclefindmkaj/https://www.gob.mx/cms/uploads/attachment/file/413015/COMAR\_2016.pdf
- COMAR. (2022, Julio 1). *Gobierno de México*. La COMAR En Números Junio 2022. <https://www.gob.mx/comar/articulos/la-comar-en-numeros-306838?idiom=es>
- Dekker, R., & Engbersen, G. (2014). How social media transform migrant networks and facilitate migration. En *Global Networks* (Vol. 14).
- Del Fava, E., Wiśniowski, A., & Zagheni, E. (2019). *Modelling International Migration Flows by Integrating Multiple Data Sources*. <https://doi.org/10.31235/osf.io/cma5h>
- Gandini, L., Prieto Rosas, V., & Lozano-Ascencio, F. (2020). Nuevas movilidades en América Latina: la migración venezolana en contextos de crisis y las respuestas en la región. *Cuadernos Geográficos*, 59(3), 102–120. <https://doi.org/10.30827/cuadgeo.v59i3.9294>
- Garcia, D., Kassa, Y. M., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., & Cuevas, R. (2018). Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27), 6958–6963. <https://doi.org/10.1073/pnas.1717781115>
- Herdağdelen, A., & Marelli, M. (2017). Social media and language processing: How Facebook and Twitter provide the best frequency estimates for studying word recognition. *Cognitive Science*, 41(4), 976–995. <https://doi.org/10.1111/cogs.12392>

- Hsiao, Y., Fiorio, L., Wakefield, J., & Zagheni, E. (2020). *Modeling the bias of digital data: an approach to combining digital and survey data to estimate and predict migration trends*. 49(April).
- INEGI. (2021). *Marco Conceptual Censo de Población y Vivienda 2020*. [www.inegi.org.mx](http://www.inegi.org.mx)
- Kashyap, R., Fatehkia, M., al Tamime, R., & Weber, I. (2020). Monitoring global digital gender inequality using the online populations of Facebook and Google. *Demographic Research*, 43(September), 779–816. <https://doi.org/10.4054/DEMRES.2020.43.27>
- Kemp, S. (2022, May 11). *Facebook statistics and trends*. <https://Datareportal.Com/Essential-Facebook-Stats>.
- Maria Pesando, L., Rotondi, V., Stranges, M., Kashyap, R., Billari, F. C., & Maria, L. (2020). *The Internetization of International Migration*.
- Masferrer, C. (2020). Cambios demográficos y geográficos del retorno de los migrantes mexicanos desde Estados Unidos antes y después de la Gran Recesión. En S. Giorguli & J. Sobrino (Eds.), *Dinámica demográfica en México en el siglo XXI* (Vol. 2, pp. 169–222). El Colegio de México.
- Masferrer, C. (2021). *Atlas de migración de retorno de Estados Unidos a México* (Primera edición) [Libro].
- Masferrer, C., Hamilton, E. R., & Denier, N. (2019). Immigrants in Their Parental Homeland: Half a Million U.S.-born Minors Settle Throughout Mexico. *Demography*, 56(4), 1453–1461. <https://doi.org/10.1007/s13524-019-00788-0>
- Moran, C. (2022). The ‘connected migrant’: A scoping review. *Convergence*. <https://doi.org/10.1177/13548565221090480>
- Moreno Mena, J. A. (2019). Migración haitiana hacia la frontera norte de México. *Espacio Abierto*, 28, 67–85. <https://www.redalyc.org/articulo.oa?id=12262976004>
- OIM. (2022). *Informe sobre las migraciones en el mundo 2022*. [www.iom.int](http://www.iom.int)
- OMS. (2021, Enero 29). *Cronología de la respuesta de la OMS a la COVID-19*. <https://www.who.int/es/news/item/20-04-2020-joint-statement-by-wto-director-general-roberto-azev%C3%A1do-and-who-director-general-tedros-adhanom-ghebreyesus>.
- Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Herranz, M. G., Al-Asad, M., & Weber, I. (2020). Monitoring of the Venezuelan exodus through Facebook’s advertising platform. *PLoS ONE*, 15(2), 1–15. <https://doi.org/10.1371/journal.pone.0229175>
- Panagiotou, N., Katakis, I., & Gunopulos, D. (2016). *Solving Large Scale Learning Tasks. Challenges and Algorithms* (S. Michaelis, N. Piatkowski, & M. Stolpe, Eds.; Vol. 9580, pp. 42–84). Springer International Publishing. <https://doi.org/10.1007/978-3-319-41706-6>
- Polo Alvis, S., Serrano López, E., & Bernal-Marquez, C. (2019). Análisis de la diáspora colombiana en México: historia, estadísticas y lecturas de causalidades. *Díkaion*, 28(1), 90–116. <https://doi.org/10.5294/dika.2019.28.1.4>

- Rampazzo, F., Bijak, J., Vitali, A., Weber, I., & Zagheni, E. (2021). A framework for estimating migrant stocks using digital traces and survey data: An application in the United Kingdom. *Demography*, 58(6), 2193–2218. <https://doi.org/10.1215/00703370-9578562>
- Salganik, M. J. (2018). Bit by bit. Social research in the digital age. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).
- Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., Guidotti, R., Bertoli, S., Kim, J., Muntean, C. I., Pappalardo, L., Passarella, A., Pedreschi, D., Pollacci, L., Pratesi, F., & Sharma, R. (2021). Human migration: the big data perspective. *International Journal of Data Science and Analytics*, 11(4), 341–360. <https://doi.org/10.1007/s41060-020-00213-5>
- Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. *PLoS ONE*, 14(10), 1–22. <https://doi.org/10.1371/journal.pone.0224134>
- Techopedia. (2022). *Granular Data*. <https://www.techopedia.com/definition/31722/Granular-Data>.
- UNSTAT. (2017). *Handbook on Measuring International Migration through Population Censuses*.
- Valenzuela Montoya, C. (2011). Migración internacional de colombianas en la ciudad de México: Transnacionalismo y globalización. In H. J. Godoy, R. González Arana, & G. Orozco Restrepo (Eds.), *Construyendo lo global. Aporte al debate de Relaciones Internacionales* (pp. 61–91). Universidad del Norte.
- Weber, I., & State, B. (2017). Digital demography. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, 935–939. <https://doi.org/10.1145/3041021.3051104>
- Welti Chanes, Carlos. (1997). *Demografía* (C. Welti Chanes, Ed.) [Libro]. México: Programa Latinoamericano de Actividades de Población: Instituto de Investigaciones Sociales, Universidad Nacional Autónoma de México, 1997.
- Wilson, D. W., Lin, X., Longstreet, P., & Sarker, S. (2011). *Web 2.0: A Definition, Literature Review, and Directions for Future Research*. [http://aisel.aisnet.org/amcis2011\\_submissions/368](http://aisel.aisnet.org/amcis2011_submissions/368)
- Zagheni, E., & Weber, I. (2012). You are where you E-mail: Using E-mail data to estimate international migration rates. *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'12*, 348–357. <https://doi.org/10.1145/2380718.2380764>
- Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1), 13–25. <https://doi.org/10.1108/IJM-12-2014-0261>
- 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. <https://doi.org/10.1111/padr.12102>

## **APÉNDICES**

## APÉNDICE 1. RAZONES DE ACCESO A FACEBOOK

### Cuadro 1.

*Proporción de acceso a Facebook en México 2020*

Grupos poblacionales	Población Facebook (1)	Población Censo (2)	RAF (1) / (2)
Hombres 13 a 24	14,000,000	12,761,521	109.7%
Hombres 25 a 54	25,000,000	24,632,851	101.5%
Hombres 55 a 65	3,300,000	5,390,386	61.2%
Mujeres 13 a 24	15,000,000	12,731,007	117.8%
Mujeres 25 a 54	26,000,000	26,865,097	96.8%
Mujeres 55 a 65	3,700,000	6,092,533	60.7%
Total	87,000,000	88,473,395	98.3%

Nota: En este cuadro no se muestra el total de hombres y mujeres porque los extractos de Facebook que se hicieron en marzo de 2020 solo contemplaron la segmentación presentada aquí. Con esto, se pone de manifiesto que las segmentaciones hechas en Facebook no son sumables.

Fuente: Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020

**Cuadro 2.**

*Razón de acceso a Facebook de la población inmigrante en México según país de nacimiento y país de residencia previa, 2020*

Razón de Acceso a Facebook	País de origen	Edad - Hombres			Edad - Mujeres		
		13 a 24	25 a 54	25 a 55	13 a 24	25 a 54	25 a 55
RAF según Inmigración por país de nacimiento	Argentina	3142%	47%	3025%	1940%	58%	3563%
	Brasil	4020%	155%	32402%	2244%	74%	13360%
	Chile	164%	10%	377%	165%	7%	304%
	Colombia	2846%	42%	2539%	1861%	60%	2781%
	Cuba	796%	172%	3994%	1178%	145%	4433%
	El Salvador	186%	25%	817%	225%	22%	909%
	Guatemala	278%	11%	718%	202%	10%	722%
	Haití	222%	18%	5190%	322%	43%	7037%
	Honduras	295%	15%	3411%	226%	11%	1525%
	Nicaragua	915%	102%	1377%	484%	86%	566%
	Perú	1181%	52%	991%	821%	59%	1658%
	República Dominicana	1064%	98%	5500%	1852%	172%	2040%
Venezuela	216%	19%	1333%	241%	21%	824%	
RAF según Inmigración por país de residencia en 2015	Argentina	3195%	103%	28689%	2355%	113%	30208%
	Brasil	3582%	180%	45669%	3419%	106%	35484%
	Colombia	377%	20%	1886%	303%	19%	1816%
	Cuba	4288%	69%	15789%	2850%	122%	14063%
	Chile	2805%	255%	18056%	9074%	243%	12587%
	El Salvador	346%	74%	14394%	408%	62%	6560%
	Guatemala	702%	52%	10000%	426%	58%	6977%
	Haití	383%	43%	5190%	680%	98%	11875%
	Honduras	407%	48%	31429%	440%	37%	7407%
	Nicaragua	1605%	195%	84000%	857%	165%	175000%
	Perú	1585%	148%	4926%	1356%	168%	40435%
	República Dominicana	6667%	175%	27500%	5769%	263%	-
Venezuela	262%	32%	2479%	361%	36%	1273%	

Fuente: elaboración propia con datos de Facebook 2020 y del Censo 2020

## APÉNDICE 2. RESUMEN DE LOS MODELOS ESTADÍSTICOS

### Cuadro 5.

*Modelos generales para la definición de inmigración 1: inmigración por país de nacimiento*

	Variable dependiente:			
	Proporción de inmigración por país de nacimiento (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	0.093 (0.072)	0.092 (0.072)	0.787*** (0.120)	0.334* (0.191)
Mujeres		-0.096 (0.266)	0.061 (0.204)	-0.041 (0.124)
25 a 54 años			2.280*** (0.317)	1.536*** (0.345)
55 a 65 años			-1.307*** (0.373)	-0.256 (0.466)
Brasil				-0.800*** (0.298)
Chile				-0.167 (0.411)
Colombia				0.667** (0.302)
Cuba				0.592* (0.313)
El Salvador				0.710* (0.381)
Guatemala				1.531*** (0.308)
Haití				-0.680 (0.504)
Honduras				0.875*** (0.325)
Nicaragua				-0.624 (0.477)
Perú				-0.110 (0.360)
República Dominicana				-1.285*** (0.446)
Venezuela				1.290*** (0.302)
Constant	-8.372*** (0.567)	-8.240*** (0.677)	-3.469*** (0.973)	-7.041*** (1.311)
Observaciones	78	78	78	78
R <sup>2</sup>	0.022	0.023	0.447	0.848
Adjusted R <sup>2</sup>	0.009	-0.003	0.417	0.808
AIC	249.2782	251.1419	210.7668	133.9918
BIC	256.3483	260.5687	224.9071	176.4125
Error estándar residual	1.166 (df = 76)	1.173 (df = 75)	0.894 (df = 73)	0.513 (df = 61)
Estadístico F	1.692 (df = 1; 76)	0.902 (df = 2; 75)	14.761*** (df = 4; 73)	21.290*** (df = 16; 61)

Nota:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 6.**

*Modelos generales para la definición de inmigración 2: inmigración por país de residencia en 2015*

	Variable dependiente:			
	Proporción de migración por país de residencia en 2015 (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	-0.073 (0.087)	-0.077 (0.087)	1.016*** (0.129)	0.712*** (0.231)
Mujer		-0.187 (0.322)	0.013 (0.217)	-0.085 (0.150)
25 a 54 años			2.505*** (0.338)	2.004*** (0.417)
55 a 65 años			-3.056*** (0.404)	-2.392*** (0.567)
Brasil				-0.216 (0.359)
Chile				0.277 (0.496)
Colombia				0.476 (0.364)
Cuba				0.885** (0.377)
El Salvador				0.825* (0.460)
Guatemala				1.055*** (0.371)
Haití				0.555 (0.610)
Honduras				0.953** (0.392)
Nicaragua				-0.923 (0.576)
Perú				0.020 (0.434)
República Dominicana				-1.046* (0.543)
Venezuela				1.960*** (0.363)
Constant	-10.727*** (0.684)	-10.474*** (0.814)	-2.227** (1.048)	-4.856*** (1.590)
Observaciones	77	77	77	77
R <sup>2</sup>	0.009	0.014	0.567	0.846
Adjusted R <sup>2</sup>	-0.004	-0.013	0.543	0.806
AIC	274.7123	276.3619	217.0173	161.1651
BIC	281.7437	285.7371	231.0801	203.3536
Error estándar residual	1.404 (df = 75)	1.410 (df = 74)	0.947 (df = 72)	0.618 (df = 60)
Estadístico F	0.717 (df = 1; 75)	0.524 (df = 2; 74)	23.558*** (df = 4; 72)	20.673*** (df = 16; 60)

Nota:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 7.**

*Modelos generales para la definición de inmigración 3: inmigración reciente por lugar de nacimiento*

	Variable dependiente:			
	Proporción de inmigración 3 (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	-0.061 (0.090)	-0.063 (0.090)	0.838*** (0.165)	0.520* (0.278)
Mujer		-0.161 (0.337)	-0.011 (0.275)	-0.106 (0.181)
25 a 54 años			2.245*** (0.425)	1.723*** (0.502)
55 a 65 años			-2.413*** (0.528)	-1.849*** (0.685)
Brasil				-0.758* (0.431)
Chile				-0.148 (0.596)
Colombia				1.039** (0.436)
Cuba				1.299*** (0.452)
El Salvador				1.161** (0.552)
Guatemala				1.276*** (0.445)
Haití				0.685 (0.732)
Honduras				1.285*** (0.470)
Nicaragua				-0.760 (0.698)
Perú				0.073 (0.521)
República Dominicana				-1.420** (0.679)
Venezuela				2.282*** (0.436)
Constant	-10.719*** (0.712)	-10.500*** (0.850)	-3.814*** (1.341)	-6.641*** (1.917)
Observaciones	75	75	75	75
R <sup>2</sup>	0.006	0.009	0.365	0.793
Adjusted R <sup>2</sup>	-0.007	-0.018	0.328	0.736
AIC	272.3015	274.0633	244.7624	184.5261
BIC	279.2539	283.3333	258.6673	226.2409
Error estándar residual	1.448 (df = 73)	1.455 (df = 72)	1.182 (df = 70)	0.741 (df = 58)
Estadístico F	0.456 (df = 1; 73)	0.340 (df = 2; 72)	10.039*** (df = 4; 70)	13.914*** (df = 16; 58)

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 8.**

*Modelos generales para la definición de inmigración 4: inmigración de la población nacida en un país distinto del de residencia en 2015 según país de residencia*

	Variable dependiente:			
	Proporción de inmigración 4 (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	-0.088 (0.090)	-0.090 (0.091)	0.764*** (0.176)	0.365 (0.297)
Mujer		-0.109 (0.341)	0.074 (0.290)	-0.011 (0.195)
25 a 54 años			2.055*** (0.449)	1.399** (0.537)
55 a 65 años			-2.378*** (0.570)	-1.662** (0.737)
Brasil				-1.038** (0.460)
Chile				-0.417 (0.637)
Colombia				0.966** (0.466)
Cuba				1.160** (0.483)
El Salvador				0.636 (0.590)
Guatemala				1.220** (0.475)
Haití				0.274 (0.783)
Honduras				1.173** (0.502)
Nicaragua				-0.809 (0.760)
Perú				-0.179 (0.556)
República Dominicana				-1.846** (0.725)
Venezuela				2.255*** (0.465)
Constant	-11.042*** (0.719)	-10.891*** (0.863)	-4.588*** (1.420)	-7.866*** (2.050)
Observaciones	74	74	74	74
R <sup>2</sup>	0.013	0.014	0.315	0.769
Adjusted R <sup>2</sup>	-0.001	-0.013	0.275	0.704
AIC	269.4351	271.3291	248.4175	191.9274
BIC	276.3473	280.5454	262.2419	233.4006
Error estándar residual	1.455 (df = 72)	1.464 (df = 71)	1.238 (df = 69)	0.791 (df = 57)
Estadístico F	0.949 (df = 1; 72)	0.520 (df = 2; 71)	7.929*** (df = 4; 69)	11.868*** (df = 16; 57)

Nota:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 9.**

*Modelos generales para la definición de inmigración 5: inmigración reciente por lugar de nacimiento*

	Variable dependiente:			
	Proporción de inmigración 5 (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	0.069 (0.089)	0.061 (0.090)	0.529** (0.205)	-0.226 (0.397)
Mujer		-0.270 (0.352)	-0.331 (0.313)	-0.658** (0.259)
25 a 54 años			2.087*** (0.518)	1.025 (0.738)
55 a 65 años			-0.300 (0.672)	1.605 (1.011)
Brasil				0.107 (0.709)
Chile				-1.259 (0.933)
Colombia				1.290* (0.696)
Cuba				0.310 (0.704)
El Salvador				0.101 (0.812)
Guatemala				-0.416 (0.687)
Haití				0.304 (1.038)
Honduras				-0.082 (0.713)
Nicaragua				-3.074*** (0.983)
Perú				-0.509 (0.812)
República Dominicana				-1.021 (1.004)
Venezuela				0.899 (0.679)
Constant	-11.795*** (0.721)	-11.469*** (0.839)	-8.541*** (1.656)	-13.936*** (2.845)
Observaciones	60	60	60	60
R <sup>2</sup>	0.010	0.020	0.253	0.637
Adjusted R <sup>2</sup>	-0.007	-0.014	0.198	0.502
AIC	209.1646	210.5485	198.2912	178.9115
BIC	215.4477	218.9259	210.8573	216.6097
Error estándar residual	1.338 (df = 58)	1.343 (df = 57)	1.194 (df = 55)	0.940 (df = 43)
Estadístico F	0.595 (df = 1; 58)	0.589 (df = 2; 57)	4.652*** (df = 4; 55)	4.724*** (df = 16; 43)

Nota:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 10.***Modelos generales para la definición de inmigración 6: inmigración de retorno*

	Variable dependiente:			
	Proporción de inmigración 6 (log)			
	(1)	(2)	(3)	(4)
Proporción de expatriados (log)	-0.170* (0.086)	-0.174* (0.087)	0.751*** (0.195)	0.221 (0.540)
Sexo		-0.159 (0.332)	0.057 (0.283)	-0.051 (0.316)
25 a 54 años			2.210*** (0.504)	1.328 (1.049)
55 a 65 años			-2.411*** (0.580)	-1.238 (1.261)
Brasil				0.078 (0.669)
Chile				-0.400 (1.043)
Colombia				0.004 (0.713)
Cuba				-0.720 (0.760)
El Salvador				-0.767 (0.971)
Guatemala				0.175 (0.704)
Haití				-1.796 (1.259)
Honduras				-1.255 (0.762)
Nicaragua				-1.098 (1.201)
Perú				-0.765 (0.899)
República Dominicana				-1.229 (1.188)
Venezuela				-0.013 (0.682)
Constant	-13.584*** (0.686)	-13.383*** (0.808)	-6.772*** (1.485)	-10.188*** (3.631)
Observaciones	64	64	64	64
R <sup>2</sup>	0.059	0.062	0.353	0.453
Adjusted R <sup>2</sup>	0.044	0.032	0.309	0.267
AIC	220.0943	221.8523	202.0975	215.3812
BIC	226.5710	230.4878	215.0508	254.2411
Error estándar residual	1.309 (df = 62)	1.318 (df = 61)	1.113 (df = 59)	1.147 (df = 47)
Estadístico F	3.870* (df = 1; 62)	2.027 (df = 2; 61)	8.050*** (df = 4; 59)	2.431*** (df = 16; 47)

Nota:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 11.**

*Modelos separados por sexo para la definición de inmigración 1: inmigración por país de nacimiento*

	Variable dependiente: Proporción de inmigración 1 (log)					
	Hombres			Mujeres		
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	0.058 (0.099)	0.659*** (0.177)	0.284 (0.326)	0.130 (0.107)	0.918*** (0.167)	0.409 (0.286)
25 a 54 años		2.195*** (0.468)	1.573** (0.584)		2.369*** (0.439)	1.537*** (0.513)
55 a 65 años		-0.993* (0.564)	-0.081 (0.823)		-1.603*** (0.506)	-0.474 (0.668)
Brasil			-0.998** (0.479)			-0.582 (0.440)
Chile			-0.115 (0.683)			-0.188 (0.604)
Colombia			0.577 (0.472)			0.745 (0.451)
Cuba			0.551 (0.489)			0.651 (0.469)
El Salvador			0.708 (0.614)			0.745 (0.571)
Guatemala			1.436*** (0.489)			1.631*** (0.455)
Haití			-0.462 (0.817)			-0.838 (0.770)
Honduras			0.707 (0.503)			1.073** (0.495)
Nicaragua			-0.994 (0.792)			-0.207 (0.709)
Perú			-0.114 (0.584)			-0.083 (0.532)
República Dominicana			-1.282* (0.716)			-1.237* (0.681)
Venezuela			1.194** (0.479)			1.383*** (0.445)
Constant	-8.595*** (0.769)	-4.449*** (1.407)	-7.473*** (2.300)	-8.138*** (0.858)	-2.254 (1.354)	-6.555*** (2.059)
Observaciones	39	39	39	39	39	39
R <sup>2</sup>	0.009	0.397	0.852	0.038	0.511	0.879
Adjusted R <sup>2</sup>	-0.018	0.345	0.756	0.012	0.469	0.800
AIC	126.02936	110.64882	79.77621	128.80553	106.38517	75.92560
BIC	131.0200	118.9666	108.0568	133.7962	114.7030	104.2061
Error estándar residual	1.157 (df = 37)	0.928 (df = 35)	0.567 (df = 23)	1.199 (df = 37)	0.879 (df = 35)	0.539 (df = 23)
Estadístico F	0.340 (df = 1; 37)	7.686*** (df = 3; 35)	8.853*** (df = 15; 23)	1.458 (df = 1; 37)	12.208*** (df = 3; 35)	11.145*** (df = 15; 23)

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 12.**

*Modelos separados por sexo para la definición de inmigración 2: inmigración por país de residencia en 2015*

	Variable dependiente: Proporción de migración residencia (log):					
		Hombres			Mujeres	
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	-0.109 (0.113)	0.888*** (0.176)	0.668* (0.347)	-0.041 (0.136)	1.153*** (0.193)	0.768* (0.388)
25 a 54 años		2.394*** (0.464)	2.029*** (0.623)		2.626*** (0.503)	1.997*** (0.695)
55 a 65 años		-2.741*** (0.559)	-2.206** (0.877)		-3.368*** (0.598)	-2.610*** (0.913)
Brasil			-0.255 (0.510)			-0.160 (0.596)
Chile			0.434 (0.728)			0.133 (0.818)
Colombia			0.427 (0.503)			0.518 (0.610)
Cuba			0.841 (0.522)			0.938 (0.635)
El Salvador			0.667 (0.654)			0.999 (0.774)
Guatemala			0.922* (0.521)			1.189* (0.616)
Haití			0.881 (0.871)			0.261 (1.044)
Honduras			0.711 (0.536)			1.213* (0.670)
Nicaragua			-0.998 (0.844)			-0.826 (0.961)
Perú			0.336 (0.622)			-0.286 (0.721)
República Dominicana			-0.966 (0.763)			-1.173 (0.959)
Venezuela			1.905*** (0.511)			2.009*** (0.602)
Constant	-10.904*** (0.878)	-3.255** (1.396)	-5.352** (2.452)	-10.569*** (1.088)	-1.080 (1.569)	-4.521 (2.795)
AIC	136.39075	110.02807	84.76393	143.13631	113.39840	97.13743
BIC	141.3814	118.3459	113.0445	148.0491	121.5863	124.9764
Observations	39	39	39	38	38	38
R <sup>2</sup>	0.025	0.552	0.873	0.003	0.589	0.858
Adjusted R <sup>2</sup>	-0.002	0.514	0.791	-0.025	0.553	0.761
Residual Std. Error	1.322 (df = 37)	0.921 (df = 35)	0.604 (df = 23)	1.511 (df = 36)	0.997 (df = 34)	0.730 (df = 22)
F Statistic	0.935 (df = 1; 37)	14.389*** (df = 3; 35)	10.578*** (df = 15; 23)	0.091 (df = 1; 36)	16.275*** (df = 3; 34)	8.841*** (df = 15; 22)

Nota:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 13.**

*Modelos separados por sexo para la definición de inmigración 3: inmigración reciente por lugar de nacimiento*

	Variable dependiente: Proporción de inmigración 3 (log)					
	Hombres			Mujeres		
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	-0.098 (0.120)	0.739*** (0.222)	0.501 (0.398)	-0.026 (0.138)	0.947*** (0.253)	0.473 (0.477)
25 a 54 años		2.233*** (0.582)	1.839** (0.714)		2.273*** (0.643)	1.499* (0.854)
55 a 65 años		-2.154*** (0.721)	-1.710 (1.007)		-2.686*** (0.799)	-1.791 (1.135)
Brasil			-0.544 (0.585)			-0.965 (0.729)
Chile			0.147 (0.834)			-0.539 (1.003)
Colombia			1.017* (0.576)			1.082 (0.747)
Cuba			1.387** (0.598)			1.176 (0.777)
El Salvador			1.190 (0.750)			1.048 (0.949)
Guatemala			1.288** (0.597)			1.236 (0.754)
Haití			1.239 (0.998)			-0.012 (1.282)
Honduras			1.158* (0.615)			1.364 (0.821)
Nicaragua			-1.012 (0.967)			-0.451 (1.213)
Perú			0.586 (0.713)			-0.510 (0.883)
República Dominicana			-1.208 (0.924)			-1.731 (1.175)
Venezuela			2.334*** (0.586)			2.212*** (0.737)
Constant	-10.920*** (0.937)	-4.653** (1.763)	-7.091** (2.810)	-10.528*** (1.110)	-2.904 (2.045)	-6.990* (3.439)
Observaciones	38	38	38	37	37	37
R <sup>2</sup>	0.018	0.372	0.852	0.001	0.366	0.796
Adjusted R <sup>2</sup>	-0.009	0.317	0.752	-0.028	0.308	0.650
AIC	137.09604	124.08697	93.10663	140.48385	127.66397	109.70242
BIC	142.0088	132.2749	120.9456	145.3166	135.7186	137.0880
Error estándar residual	1.395 (df = 36)	1.148 (df = 34)	0.692 (df = 22)	1.531 (df = 35)	1.256 (df = 33)	0.893 (df = 21)
Estadístico F	0.664 (df = 1; 36)	6.725*** (df = 3; 34)	8.465*** (df = 15; 22)	0.035 (df = 1; 35)	6.348*** (df = 3; 33)	5.463*** (df = 15; 21)

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 14.**

*Modelos separados por sexo para la definición de inmigración 4: inmigración de la población nacida en un país distinto del de residencia en 2015 según país de residencia*

	Variable dependiente: Proporción de inmigración 4 (log)					
	Hombres			Mujeres		
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	-0.145 (0.122)	0.602** (0.253)	0.408 (0.502)	-0.029 (0.137)	0.936*** (0.251)	0.398 (0.463)
25 a 54 años		1.950*** (0.649)	1.628* (0.900)		2.174*** (0.638)	1.295 (0.829)
55 a 65 años		-2.013** (0.841)	-1.802 (1.283)		-2.736*** (0.793)	-1.702 (1.101)
Brasil			-1.060 (0.734)			-1.032 (0.708)
Chile			-0.076 (1.050)			-0.646 (0.973)
Colombia			0.966 (0.723)			0.942 (0.725)
Cuba			1.298* (0.750)			1.061 (0.754)
El Salvador			0.541 (0.943)			0.826 (0.921)
Guatemala			1.262 (0.749)			1.211 (0.732)
Haití			0.975 (1.257)			-0.267 (1.243)
Honduras			1.157 (0.772)			1.242 (0.797)
Nicaragua			-0.897 (1.242)			-0.579 (1.177)
Perú			0.326 (0.896)			-0.603 (0.857)
República Dominicana			-1.565 (1.161)			-1.998* (1.139)
Venezuela			2.352*** (0.735)			2.181*** (0.715)
Constant	-11.422*** (0.962)	-5.815*** (2.009)	-7.737** (3.549)	-10.633*** (1.100)	-3.026 (2.030)	-7.498** (3.335)
Observaciones	37	37	37	37	37	37
R <sup>2</sup>	0.039	0.288	0.785	0.001	0.363	0.804
Adjusted R <sup>2</sup>	0.011	0.224	0.631	-0.027	0.305	0.665
AIC	134.9824	127.8525	107.6085	107.6085	127.1309	107.4399
BIC	139.8152	135.9071	134.9941	144.6013	135.1855	134.8255
Error estándar residual	1.422 (df = 35)	1.260 (df = 33)	0.868 (df = 21)	1.517 (df = 35)	1.247 (df = 33)	0.866 (df = 21)
Estadístico F	1.406 (df = 1; 35)	4.458*** (df = 3; 33)	5.104*** (df = 15; 21)	0.045 (df = 1; 35)	6.268*** (df = 3; 33)	5.758*** (df = 15; 21)

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 15.**

*Modelos separados por sexo para la definición de inmigración 5: inmigración reciente por lugar de nacimiento*

	Variable dependiente: Proporción de inmigración 5 (log)					
	Hombres			Mujeres		
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	0.021 (0.113)	0.380 (0.238)	0.175 (0.579)	0.118 (0.149)	0.755* (0.382)	0.340 (0.570)
25 a 54 años		2.131*** (0.598)	1.824* (1.042)		2.168** (0.978)	2.125* (1.129)
55 a 65 años		0.148 (0.810)	0.797 (1.503)		-0.895 (1.177)	0.084 (1.343)
Brasil			-0.459 (0.941)			0.688 (1.006)
Chile			-0.516 (1.252)			-0.820 (1.324)
Colombia			0.963 (0.925)			1.277 (0.991)
Cuba			0.279 (0.914)			0.897 (1.019)
El Salvador			0.506 (1.095)			0.880 (1.168)
Guatemala			0.366 (0.913)			-0.811 (0.974)
Haití			1.642 (1.468)			0.778 (1.452)
Honduras			-0.098 (0.932)			0.589 (1.036)
Nicaragua			-1.480 (1.384)			-4.712*** (1.324)
Perú			0.192 (1.048)			-0.715 (1.236)
República Dominicana			0.066 (1.429)			-0.974 (1.324)
Venezuela			1.177 (0.903)			0.864 (0.961)
Constant	-12.042*** (0.895)	-10.159*** (1.894)	-11.995** (4.203)	-11.550*** (1.235)	-7.280** (2.971)	-11.061** (4.164)
Observaciones	34	34	34	26	26	26
R <sup>2</sup>	0.001	0.326	0.686	0.025	0.204	0.871
Adjusted R <sup>2</sup>	-0.030	0.258	0.425	-0.015	0.095	0.677
AIC	118.01301	108.64621	106.62906	96.03301	94.78670	71.53585
BIC	122.59209	116.27801	132.57719	99.80730	101.07719	92.92349
Error estándar residual	1.295 (df = 32)	1.099 (df = 30)	0.968 (df = 18)	1.423 (df = 24)	1.343 (df = 22)	0.803 (df = 10)
Estadístico F	0.036 (df = 1; 32)	4.833*** (df = 3; 30)	2.626** (df = 15; 18)	0.627 (df = 1; 24)	1.874 (df = 3; 22)	4.486** (df = 15; 10)

Nota: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.

**Cuadro 12.***Modelos separados por sexo para la definición de inmigración 6: inmigración de retorno*

	Variable dependiente: Proporción de inmigración 6 (log)					
	Hombres			Mujeres		
	(1)	(2)	(3)	(4)	(5)	(6)
Proporción de expatriados (log)	-0.201 (0.126)	0.627* (0.322)	-0.361 (1.328)	-0.140 (0.122)	0.877*** (0.223)	0.068 (0.549)
25 a 54 años		2.416*** (0.841)	0.568 (2.642)		1.995*** (0.572)	0.659 (1.058)
55 a 65 años		-1.774* (0.972)	0.437 (3.103)		-3.035*** (0.653)	-1.265 (1.248)
Brasil			0.986 (1.309)			-0.614 (0.689)
Chile			-0.357 (2.379)			-1.563 (1.051)
Colombia			0.790 (1.266)			-0.770 (0.795)
Cuba			0.155 (1.379)			-1.723** (0.802)
El Salvador			-1.133 (2.042)			-1.379 (1.058)
Guatemala			-0.230 (1.369)			0.237 (0.724)
Haití			-2.314 (2.517)			-2.524* (1.335)
Honduras			-1.428 (1.454)			-1.557* (0.815)
Nicaragua			-1.052 (2.517)			-1.890 (1.220)
Perú			-1.431 (1.888)			-0.757 (0.961)
República Dominicana			-1.795 (2.626)			-1.674 (1.242)
Venezuela			-0.505 (1.313)			0.222 (0.700)
Constant	-13.745*** (0.979)	-7.923*** (2.442)	-14.803 (9.133)	-13.432*** (0.990)	-5.414*** (1.745)	-10.774** (3.893)
Observaciones	33	33	33	31	31	31
R <sup>2</sup>	0.076	0.290	0.448	0.044	0.510	0.764
Adjusted R <sup>2</sup>	0.046	0.217	-0.039	0.011	0.456	0.528
AIC	120.67564	115.98758	131.68447	104.31992	87.58928	88.97826
BIC	125.16516	123.47012	157.12510	108.62188	94.75922	113.35604
Error estándar residual	1.419 (df = 31)	1.286 (df = 29)	1.481 (df = 17)	1.222 (df = 29)	0.906 (df = 27)	0.844 (df = 15)
Estadístico F	2.553 (df = 1; 31)	3.947** (df = 3; 29)	0.919 (df = 15; 17)	1.328 (df = 1; 29)	9.370*** (df = 3; 27)	3.233** (df = 15; 15)

Nota: \*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Estimaciones propias basadas en datos del Censo 2020 y de Facebook 2020.