DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 50, ARTICLE 10, PAGES 291–324
PUBLISHED 13 FEBRUARY 2024
https://www.demographic-research.org/Volumes/Vol50/10/
DOI: 10.4054/DemRes.2024.50.10

Research Article

Measuring short-term mobility patterns in North America using Facebook advertising data, with an application to adjusting COVID-19 mortality rates

Lindsay Katz

Michael Chong

Monica Alexander

This publication is part of the Special Collection on Social Media and Demographic Research, organized by Guest Editor Emilio Zagheni.

Contents

Demographic Research: Volume 50, Article 10
Research Article

Measuring short-term mobility patterns in North America using Facebook advertising data, with an application to adjusting COVID-19 mortality rates

Lindsay Katz[1]

Michael Chong[2]

Monica Alexander[3]

Abstract

BACKGROUND
Patterns in short-term population mobility are important to understand, but the data required to measure such movements are often not available from traditional sources.

OBJECTIVE
To investigate patterns in short-term population mobility in all states and provinces in the United States and Canada using data collected from Facebook's advertising platform.

METHODS
We collected daily traveler data from Facebook's advertising platform, summarized the main characteristic patterns observed across geographic regions, and also used the traveler rates to adjust COVID-19 mortality rates over the period July 2020 to July 2021.

RESULTS
Rates of short-term travel vary substantially by geographic area but also by age and sex, with the highest rates of travel generally for males. Strong seasonal patterns are apparent in travel to many areas, with different regions experiencing either increased travel or decreased travel over winter, depending on climate. Further, some areas appear to show marked changes in mobility patterns since the onset of the pandemic. In addition, accounting for travelers in population denominators leads to about a 1% difference in implied mortality rates, with substantial variation across demographic groups and regions.

[1] Statistical Sciences, University of Toronto, Toronto, Ontario, Canada. lindsay.katz@mail.utoronto.ca.
[2] Statistical Sciences, University of Toronto, Toronto, Ontario, Canada.
[3] Statistical Sciences and Sociology, University of Toronto, Toronto, Ontario, Canada.

CONCLUSIONS
Short-term population mobility can vary substantially over the course of a year, which has implications for resource planning and the population at risk of health outcomes by geography.

CONTRIBUTION
This work highlights the potential for data collected through social media websites to provide insight into short-term mobility patterns.

1. Introduction

Migration is an important component of population change. But compared to fertility and mortality, migration is fundamentally more difficult to define and measure. Migrants can be considered across different time scales, and migration can be either permanent or temporary. Demographers are often interested in migrants who move into or out of a geographic area in a particular year, as the net number of migrants is an important component in determining how much the annual population in an area has increased or decreased. While annual migration rates are important to measure and have traditionally been the focus in demographic estimation, considering more granular temporal scales can shed light on both the measurement of migration and the dynamics of population compositional change (Bell and Ward 2000; Bernard and Perales 2021; Fiorio et al. 2017; Panczak, Charles-Edwards, and Corcoran 2020).

With a focus on shorter time scales, population mobility need not even constitute "migration" but rather might reflect more transient movements for work or leisure. For example, across the course of a year, people may travel for summer vacations, seasonal employment, or major holidays. These patterns in short-term mobility within and across populations are important to understand. Seasonal fluctuations in the number and types of people in a geographic area can drastically affect demand for different types of goods and services, and the magnitude of movements between geographic areas can have a substantial impact on the spread of infections, as evidenced by the spread of COVID-19 (Auliya and Wulandari 2021; Zargari et al. 2022). From a purely demographic perspective, substantial seasonal fluctuations in the size and composition of populations in an area could have a substantial effect on counting at-risk populations for various demographic outcomes. For example, deaths due to influenza peak during the winter months (Iuliano et al. 2018), and if there is substantial out-migration from cold areas in winter, particularly of older people, then the corresponding infection mortality rates due to influenza in these places may be underestimated.

While understanding temporary mobility is essential, data on such movements are often not widely available from traditional sources. Censuses, large-scale surveys, and other administrative data sources, such as tax file records, tend to focus on permanent changes of address and more broadly often lack the infrastructure to capture temporary visitors in a geographic area. Some previous work has utilized differences in address of residence and place of enumeration to glean information on temporary populations (Bell and Ward 1998). While giving an in-depth snapshot of temporary populations at one point in time, these approaches do not have the collection frequency required to explain seasonal fluctuations, for example.

More recently, there has been an increase in the use of nontraditional data sources — such as cellphone or digital trace data — to study short-term mobility. The most common of these data sources, and the fastest growing in the United States, is cellphone data. This source is incredibly rich from the point of view of spatial and temporal granularity (Panczak, Charles-Edwards, and Corcoran 2020). A major issue here is one of access to data: cellphone records are usually privately owned and so can be accessed only through purchase of the data or some other data-sharing agreement for research purposes. Even if cellphone data area accessible to researchers, there is often a delay in retrieving the data, which means current patterns are difficult to examine. In addition, key demographic variables may not be available in cellphone data, so patterns across age and sex, for example, cannot be disentangled.

In this paper, we investigate using data extracted from the social media website Facebook to measure patterns in short-term mobility. Data on the number of people traveling in each US state; Washington, DC; and the Canadian provinces/territories were collected from Facebook's advertising platform every day, beginning in July 2020, over a period of more than a year. Importantly, extracted Facebook data can be disaggregated by sex and broad age group, which allows us to track seasonal variation over the course of a year in short-term travel across subpopulations of interest. While Facebook data have been previously used in migration research to measure migrant stocks across multiple different countries (Alexander, Polimis, and Zagheni 2020; Rampazzo et al. 2021; Zagheni, Weber, and Gummadi 2017), less work has focused on using these data to gain insights into more temporary and short-range movements. Analyses presented in this paper demonstrate clear seasonal and other temporal patterns in travelers' data, which vary substantially by geographic area, age, and sex. We compare these observed trends in travel to patterns present in inbound flight data for each US state as another means of capturing short-term movement. Further, we demonstrate an application of the Facebook data in the adjustment of population exposures to recalculate COVID-19 mortality rates in the United States, and we show that the impact on mortality rates is in the range of 0.6%–1%.

The remainder of this paper is structured as follows. First we give a brief background of existing literature, followed by a detailed description of each dataset used in this work. We then present the methods used to analyze the Facebook data descriptively and to recompute and adjust 2021 COVID-19 mortality rates. In the results section, we summarize main observations based on the Facebook data collected, compare trends to those found in national flight data, and present results of using seasonal patterns observed in mobility rates in the Facebook data to recalculate mortality rates due to COVID-19 in 2021. We conclude with a discussion of the implications, potential uses, and limitations of these data for demographic research.

2. Background

A growing body of literature utilizes nontraditional data sources to study patterns of migration. Much of this work emphasizes the lack of both timeliness and granularity characteristic of traditional survey data, motivating an interest in the use of new types of data to study human mobility. For research purposes, social media data have been sourced from cellphone records and a variety of platforms, including Twitter, Facebook, LinkedIn, and Airbnb. For instance, Zagheni et al. (2014) use geolocated Twitter data to investigate both international and domestic trends of out-migration in OECD countries. Fiorio et al. (2017) used geolocated tweets of more than 60,000 Twitter users to calculate measures of internal migration in the United States, assuming different time scales, which allowed the authors to consider the relationship between short- and long-term migration. An increasing body of work has utilized data from LinkedIn to study the international movements of high-skilled workers (State et al. 2014; Vieira et al. 2022). Relatedly, while not strictly social media data, bibliometric data are increasingly being used to track movements of academics over time (Miranda-González et al. 2020; Subbotin and Aref 2021; Zhao et al. 2023).

In terms of using the Facebook social media platform in migration estimation, the majority of work has focused on using data sourced from Facebook's advertising platform, which currently can be extracted from the website free of charge. The advertising platform gives estimates of population groups on Facebook with various demographic, social, cultural, or economic characteristics. The population perspective makes it particularly amenable to demographic research and allows the benefit of avoiding any confidentiality or accessibility concerns with using individual-level data. Work by Zagheni, Weber, and Gummadi (2017) was the first to demonstrate the utility of these data in picking up patterns of international migration in the United States. This work illustrated that data scraped from Facebook's advertising platform correlated well with other gold standard measures of migration (Zagheni, Weber, and Gummadi 2017).

We chose to use this platform because it is well established, currently free, and relatively stable. Alexander, Polimis, and Zagheni (2020) extended this work to consider a combination of Facebook data and American Community Survey data to “nowcast” migrant stocks in the United States. The authors propose a statistical framework that uses both data sources while accounting for the bias inherent to Facebook’s advertising platform data. Similar efforts to combine the strengths of the timely but biased Facebook estimates with more traditional survey or census data have been applied to other parts of the world, most notably the United Kingdom (Rampazzo et al. 2021) and other parts of Europe (Spyratos et al. 2018). While Facebook advertising platform data have been widely used in migration estimation, the focus has been on measuring changes in longer-term migration. Exceptions include Alexander, Polimis, and Zagheni (2019), who used these types of data to estimate short-term fluctuations in Puerto Rican migrants to the continental United States in the months following Hurricane Maria in 2017. In addition, Leasure et al. (2023) employed Facebook advertising data to nowcast daily population displacement in the weeks following the Russian invasion of Ukraine.

Much of the existing literature on short-term mobility and seasonal patterns focuses on either measuring tourism for economic purposes or measuring seasonal mobility patterns as they relate to the spread of infectious diseases. For example, Xu et al. (2021) use cellphone data to investigate international tourism patterns in three major cities in South Korea. Their analysis uncovers heterogeneity in day-to-day traveler behavior across these cities, highlighting the value of cellphone records for studying micro-level patterns of tourism. Similarly interested in travel, Cebeillac and Vaguet (2021) assess the value of Airbnb data as a resource to capture tourist flows in Iceland by comparing these data to official Icelandic statistics. They find that seasonal variation in tourism flows based on Icelandic official statistics are well captured in the Airbnb data and assert that while the data are biased, they have potential as a proxy for annual tourism estimates. In Canada, Wilton and Wirjanto (1998) aim to identify and quantify patterns of seasonality in Canadian tourism using national tourism indicators data from Statistics Canada. Focus is placed on the seasonality of commodities and industries relating to tourism and the associated economic implications. In terms of studying the spread of infectious diseases at a macro level, a large body of literature uses air flight data to understand seasonal patterns. For example, Brownstein, Wolfe, and Mandl (2006) study the impact of air travel and its seasonal variation on the spread of influenza in the United States using disease surveillance data. Other studies show similar relationships in other regions of the world (Findlater and Bogoch 2018; Grais, Hugh Ellis, and Glass 2003).

In addition to exploring trends in the Facebook data, we illustrate a potential application of this dataset to adjusting COVID-19 mortality rates in the United States to account for fluctuations in the number of travelers in a particular region over the course of a year. Previous research has focused on assessing the impact of both micro- and macro-level population mobility on trends in COVID-19 mobility from the point of view of increasing the spread and speed of the transmission of an infectious disease (Alessandretti 2022). For example, Slater et al. (2022) extended classical spatial statistical models to incorporate mobility data derived from cellphones to account for population mobility in regions of Spain. Other researchers have incorporated mobility data from cellphones into infectious disease compartment models (Fields et al. 2021).

However, the size of a population in a region at a particular point in time can also influence the calculation of COVID-19 mortality rates by changing the implied number of people at risk to exposure to the disease. Adjusting for the seasonal dynamics of populations over the course of a year is important, particularly in the case of infectious diseases, but is rarely done, most likely due to data availability issues. Previous work has shown the substantial impact of population mobility on the estimation of disease incidence and mortality. For example, cellphone data were used to show that malaria incidence varied by as much as 30% in Namibia (zu Erbach-Schoenberg et al. 2016).

When calculating mortality rates at a temporal granularity of less than a year, standard practice is to interpolate annual population counts to get an estimate of inter-year population size. For example, this interpolation method is used to produce weekly mortality rates in the short-term mortality fluctuations (STMF) database maintained by the Human Mortality Database team (Jdanov et al. 2021). However, this interpolation method, which assumes a constant level of population increase or decrease, ignores the seasonal fluctuations in population levels we have demonstrated using Facebook data. So it is possible that the COVID-19 mortality rates previously reported may be under- or overestimated, depending on the specific dynamics of the region and demographic subgroup of interest.

To summarize, valuable work has highlighted the potential of nontraditional data to complement or proxy traditional survey data in the study of human mobility. However, the majority of this work focuses on either long-term migration or short-term, short-distance movement at the individual level. While existing literature considers short-term, seasonal mobility from an economic and epidemiological standpoint, there is limited work considering seasonal changes in population sizes through a demographic lens. Our work aims to bridge that gap utilizing data from Facebook's advertising platform at the state/province level in North America.

3. Data

3.1 Facebook data

The data collected for this work are sourced from Facebook's advertising platform. This platform lets advertisers on Facebook target specific audiences by selecting characteristics such as age, sex, and current location, all of which can be specified with the appropriate parameters in an API query statement. Once all desired characteristics are selected, the platform gives an estimate of the potential reach (daily and monthly active users) of the advertisement to this subgroup. We use these estimates as a data point – an observation (albeit an imperfect one) of the current number of travelers in a particular geographic region and in a particular demographic subgroup of interest – and use this to build a database of traveler information. We were interested in extracting information about people who are flagged as traveling in a certain location.

To query the Facebook marketing API for traveler and population counts, we filled the "location_types" query field with "travel_in" and "home" arguments, respectively. The "travel_in" specification was defined to capture individuals who were most recently active in the location (the destination) of interest and were more than 100 miles away from the city displayed as current on their Facebook profile (their home location) (Meta 2023). In contrast, we used the "home" location type to extract data on users whose current city displayed on their Facebook profile is within the location of interest. This was validated by Facebook based on IP addresses and data from users' friends' profile locations (Meta 2023).

In this paper we consider data collected every day over the period July 1, 2020, to July 1, 2021. We collected estimates of daily and monthly users traveling in each province in Canada and state in the United States. In addition to traveler counts, estimates of the total daily and monthly population (total number of Facebook users) for each state and province were collected. The population estimates were collected to serve as denominators to compute a traveler rate for each subgroup of interest.

We collected four sets of observations: traveler counts with and without age breakdown and population counts with and without age breakdown. In the age-disaggregated data, three broad age groups are available: 13–29, 30–50, and 50–65. All data were disaggregated by sex and every US state (and Washington, DC) and Canadian province/territory. For the age-specific data, this corresponded to 2 gender groups $\times 3$ age groups $\times (51$ states $+ 13$ provinces$) = 384$ observations of travelers and population collected every day. For the non-age-specific data, this corresponded to 2 gender groups $\times (51$ states $+ 13$ provinces$) = 128$ observations of travelers and population collected every day. We collected data that were not disaggregated by age to increase the chances of having meaningful observations for smaller populations and geographic regions, which are generally rounded up in the Facebook advertising estimates.

We then addressed any obvious errors in the raw data. First, population counts of zero were treated as missing and filled with the previous available day's data for that region and demographic group. Next we noticed a number of observations where the daily traveler count was zero. Intuitively, we expected that this was the case for regions where the daily population was relatively small. However, for regions with a large daily population count and a typical daily travel count in the thousands, we suspected that these were erroneous data points. To address this concern, we removed rows with a daily traveler count of zero unless they belonged to one of the Canadian territories (Yukon, Northwest Territories, and Nunavut), which have the lowest populations by far of the regions collected. Finally, observations where the traveler count exceeded the total population were dropped.

3.2 National flights data

To allow for a comparison between trends observed in our travel data and inbound flight patterns, we obtained data on the actual arrival time of daily inbound flights to the largest international airport in each US state for four major US airlines (where available): American Airlines, Delta Airlines, Southwest Airlines, and United Airlines. In states with multiple large international airports such as New York and Florida, we collected this data for more than one airport to get a more accurate observation of flight volume for that state. A full list of these airports is provided in the appendix. These data were collected only for the United States because daily inbound flight data by Canadian province/territory is not readily available.

We calculated the number of inbound flights for each state per day. The flight volume data were extracted to cover the same time span covered in the Facebook data. While the flight volume data were meant to serve as a rough proxy for capturing travel trends, it is important to recognize that they do not necessarily provide a complete and accurate picture of travel patterns. Notably, not all airports and airlines are accounted for in the data, and some airports serve primarily as connecting airports. It is also worth noting that during the COVID-19 pandemic, many flights were being run empty. For consistency and completeness, we considered only states with flight data for at least three airlines, which was the case for 41 states.

3.3 COVID-19 deaths data

To calculate adjusted monthly COVID-19 mortality rates for 2020–2021, we obtained data on COVID-19 death counts and population counts for the regions and demographic groups of interest. Specifically, we obtained counts of daily deaths due to COVID-19 by US state, sex, and age group over the period July 1, 2020, to July 1, 2021, from National Center for Health Statistics data published on the Centers for Disease Control and Prevention (CDC) data catalog (National Center for Health Statistics 2023). We obtained annual population counts by US state, sex, and age group for 2020 and 2021 from the US Census Bureau's state population by characteristics 2020–2022 data (United States Census Bureau 2023b).

The provisional COVID-19 deaths dataset we used contains death counts for a number of causes of death, including COVID-19, pneumonia, and influenza, based on the *International Classification of Diseases*, tenth edition (ICD-10) (National Center for Health Statistics 2023). For the purpose of our analysis, we used only the COVID-19 death counts from this dataset.

We aggregated deaths and population counts into monthly counts by US state, by sex, and for the three age groups, 18–29, 30–49, and 50–64, to be as consistent as possible with the Facebook data.

Note that for validation purposes, we ensured that we could reproduce the official quarterly COVID-19 mortality rates by state published online by the CDC (Ahmad and Cisewski 2023). These rates were age-adjusted using data from the US Census Bureau's 2000 national population projections for the year 2021 as a standard population (United States Census Bureau 2023a).

3.4 Assumptions and biases

A number of assumptions and biases inherent to the Facebook data can act as mechanisms influencing change observed in the data over time and across demographic groups. These assumptions should be taken into consideration when interpreting analytical findings.

First, it is important to consider that the period of time for which we collected these data was the height of the COVID-19 pandemic. This was a period when numerous lockdowns and flight restrictions were put into effect, with large variability in the degree to which these restrictions on mobility were taking place over time and across regions. For instance, by December 2, 2020, businesses were mostly open and masks were not required in some states, including Florida, Georgia, and Tennessee, but in California there was a curfew in place in certain counties, with masks being mandatory and some businesses closed (New York Times 2020). At the same time, other states, such as

Illinois, Michigan, and New Mexico, had closed most businesses and made masks mandatory. New Mexico also had a stay-at-home advisory in place at this time (New York Times 2020). Many of these restrictions had changed significantly by the end of our time period. For example, by June 21, 2021, California and Illinois were completely reopened and masks were required only indoors for unvaccinated people (New York Times 2021). Approaches to COVID-19 restrictions also varied across provinces in Canada. In April 2021, Dickson (2021) reported on the greater degree of lockdown stringency seen in Quebec compared to Ontario at this time. During the pandemic, many US states and Canadian provinces also implemented travel and flight restrictions, which differed across regions. In their article, Marshall and Syed (2020) highlighted the significant differences in travel limitations across states. By August 2020, all northeastern and most mid-Atlantic states had implemented statewide restrictions on travel, while more than half of the states, including Florida and Texas, had not implemented any travel restrictions (Marshall and Syed 2020).

The lockdowns and flight restrictions put in place throughout the pandemic may have had a significant influence on the behaviors and trends of travelers over that time period. For instance, travelers may have chosen to visit destinations with more lenient COVID-19 protocols or planned road trips to destinations close to home in response to flight limitations. These varying restrictions and their impacts on mobility are both important and necessary to consider when contextualizing any patterns of travel observed in these data.

Further, different demographic groups may use Facebook differently, which may impact the coverage and accuracy of its traveler data. For example, it is possible that observed gender differences are related to differences in the behavior of men on Facebook compared to that of women, such as the frequency of posting or location updates. For instance, if women have a greater tendency to keep the current city up to date on their profiles compared to men and are more active on Facebook when traveling in general, this would result in more accurate data for female users compared to male users. The same goes for travel differences across age groups. In sum, behavioral variation on Facebook across demographic groups could lead to differences in observed travel rates, which are not actually true differences that reflect the population at large.

Throughout this paper, we focus on using estimates of traveler rates rather than estimates of raw traveler counts. This lets us avoid some issues with the Facebook data, in particular the fact that the population of users on Facebook is a smaller subset of the total population of interest. However, using the rate of travelers still involves several important assumptions about the characteristics of Facebook users. In particular, we assume that the Facebook traveler rate in a particular region and demographic subgroup is a reasonable estimate of the true traveler rate in the population of interest. This assumption would be violated if people who use Facebook have a higher or lower propensity to travel than people who don’t use Facebook. For example, it might be reasonable to believe that a higher rate of Facebook users travel for leisure compared to those who don’t use Facebook because people who travel regularly may want to share their experiences on social media. Nevertheless, we believe that the order of magnitude of the traveler rates obtained from Facebook is realistic and that adjustments to COVID-19 mortality rates will give important insights into the possible effects of short-term mobility on assessing the true impact of the disease. Further, these data are valuable for understanding relative increases in mortality throughout the pandemic.

In addition, note that we were unable to obtain COVID-19 death counts starting at age 13, so the youngest age group in the death and population data starts at age 18, whereas the traveler rate used to make an adjustment for this group starts at age 13. This assumes the rate of travelers aged 13–17 is similar to that of those aged 18–29, which is probably not true; it’s likely that the traveler rate for 18- to 29-year-olds is on average higher than that for 13- to 17-year-olds. This is a limitation of the data collection strategy; in the future we will collect data that disaggregate 13- to 17-year-olds from those 18 and older.

In lieu of gold standard data sources to explain short-term mobility patterns, we believe these data are a valuable resource despite the acknowledged potential biases. We discuss interpretation of the results in light of these limitations further in Section 6.

4. Methods

4.1 Descriptive methods for Facebook data

Using the merged, cleaned Facebook datasets, we computed daily rates of travel; total traveler count was divided by total population. We then visualized these data with a number of time series plots. For each region, we plotted the daily traveler rate over time for male and female travelers separately. For the plots using the data with age breakdowns, observations were categorized by age group to allow for clear visual differentiation of the data.

The time series plots uncovered interesting patterns in the raw data across subpopulations. These preliminary findings motivated us to compute summary metrics to empirically capture those visual patterns of interest. Before doing so, we extracted smooth estimates of the time series data in an effort to reduce noisiness and focus on systematic trends. We used the locally estimated scatterplot smoothing (Loess) method to derive these estimates. This method is non-parametric, meaning it does not rely on any structural assumptions about the data. We first split our data by all unique demographic groups for each region and fit a Loess model for each subpopulation with daily rate as the outcome variable and date as the explanatory variable. We then used these models to predict the smoothed estimates. Before computing summary metrics for our analysis, we calculated the root mean square error (RMSE) between the observed data and our smoothed estimates for each demographic group in each region. A smaller RMSE suggests a closer fit of the model to the data. Using our RMSE computation, we filtered out data for any subpopulations where the RMSE exceeded the 95th quantile (~ 0.01). We did this for both the age-disaggregated data and the non-age-disaggregated data. This resulted in the removal of data from four Canadian provinces/territories: Northwest Territories, Nunavut, Yukon, and Prince Edward Island. Visual checks of data from these groups confirmed that the omitted data were too noisy for our analyses. We manually filtered out one additional demographic group from Yukon – females aged 30–50 – because every traveler count collected for that group is zero, resulting in travel rates of zero and a subsequent RMSE of zero. All data and code to reproduce the analyses can be found at https://github.com/lindsaykatz/fb-short-term-mobility. We conducted all analyses in R, version 4.2.2 (R Core Team, 2022).

We then used the smoothed travel rate estimates to compute metrics, which summarize our initial findings. First we computed the year-on-year change from July 1, 2020, to July 1, 2021, for each region and demographic group. The second summary metric we computed is intended to capture seasonal variation in travel rates, motivated by the observation that some groups reach peaks or troughs in the winter months. To capture these peaks and troughs, we computed the growth rate from July 1, 2020, to each individual date ranging from October 1, 2020 to March 31, 2021. We then considered the maximum absolute growth rate for each region, based on the date with the most extreme positive or negative rate of change from July 1, 2020. The final summary metric we computed is a male-to-female ratio of the daily travel rate over time. We used the data without age groups to compute this, and for each region and date, we divided the smoothed daily travel rate estimate for male travelers by its female counterpart.

4.2 Methods for COVID-19 mortality rate computations

Our goal in this application is to compare unadjusted monthly COVID-19 mortality rates to those adjusted for fluctuations in traveler rates based on estimates from Facebook.

The unadjusted, baseline monthly mortality rate due to COVID-19 in a particular age group a, state r, sex s, and month t is defined as

$$M_{a,r,s,t} = \frac{D_{a,r,s,t}}{P_{a,r,s}^{2020} + \frac{t}{12}\left(P_{a,r,s}^{2021} - P_{a,r,s}^{2020}\right)}$$

where D is the total number of deaths for that month, P^{2020} is the annual population estimate at July 1, 2020, and P^{2021} is the annual population estimate at July 1, 2021. Note that the month number t starts at July (of 2020) as $t = 0$, with July (of 2021) as $t = 12$.

The adjusted mortality rate is defined as

$$M_{a,r,s,t} = \frac{D_{a,r,s,t}}{\left(P^{2020}_{a,r,s} + \frac{t}{12}\left(P^{2021}_{a,r,s} - P^{2020}_{a,r,s}\right)\right) \cdot \left(1 + N_{a,r,s,t}\right)}$$

where N is the net monthly traveler rate – that is, the in- minus out-traveler rate, obtained from the corresponding Facebook estimates. Note that the Facebook data we collected allow us to obtain the rate of in-travelers in a particular state but not the rate of travelers from a particular state. To estimate the rate of travelers out of a state r in a particular month t, $O_{a,r,s,t}$, we assume that the proportion of travelers to each other state x that comes from state r is proportional to the population in state r:

$$O_{a,r,s,t} = \sum_{x \neq r} T_{a,x,s,t} \frac{P^{f}_{a,r,s,t}}{\sum_{x \neq r} P^{f}_{a,x,s,t}}$$

where $T_{a,x,s,t}$ is the number of travelers and $P^{f}_{a,r,s,t}$ is the population estimate in age group a, state r, sex s, and month t collected from Facebook. The net traveler rate is then

$$N_{a,r,s,t} = I_{a,r,s,t} - O_{a,r,s,t}$$

where $I_{a,r,s,t}$ is the rate of in-travelers in age group a, state r, sex s, and month t collected from Facebook. That is:

$$I_{a,r,s,t} = \frac{T_{a,r,s,t}}{P^{f}_{a,r,s,t}}$$

5. Results

5.1 Descriptive summary

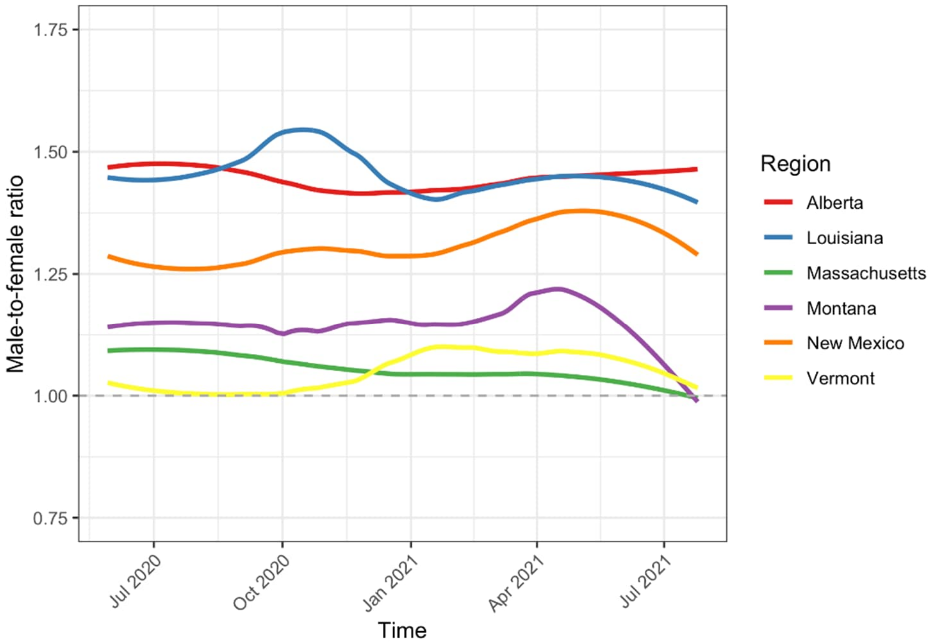
This subsection descriptively summarizes the main patterns observed in the traveler dataset, including broad age and sex patterns, and provides a discussion of four different types of geographic regions based on common seasonal patterns. The discussion draws on some regions as examples. For a full set of plots for all geographic areas and demographic groups, see our companion interactive application here: https://lindsaykatz.shinyapps.io/travel_rate_plots/.

5.1.1 Broad age and sex patterns

The data suggest that men in general travel more than women across the entire time period. This is true for all days across all geographic regions, except Montana and Massachusetts, where the ratio dips below one only in the last two and three days of the time series, respectively. The estimated sex ratios of traveler rates for several geographic regions are shown in Figure 1. The average male-to-female travel rate ratio across the entire time series is 1.29 for Canadian provinces/territories and 1.18 for US states and Washington, DC. The sex ratio of travelers is particularly high in regions such as Alberta and Louisiana, at around 1.5. That is, there are 50% more male travelers than female travelers. Regions with relatively low sex ratios include New England states such as Vermont and Massachusetts, which have traveler rate ratios closer to one.

In contrast to the observed gender difference in travel, we did not observe age groups that had consistently higher or lower travel rates. However, we did observe that seasonal variation in migration is more pronounced for the oldest age group (those aged 50–65). As mentioned in Section 4, many demographic groups reached a peak or trough in the travel rate during winter months. In 37 of the 61 regions of interest, the 50–65 age group experienced the largest variation in the travel rate from the July 1, 2020, baseline based on the seasonal variation summary measure. Further, in 24 of those 37 regions, the male 50–65 age group was attributed to the most extreme peak or trough.

Figure 1: Male-to-female ratio of the daily rate of travel over time for a range of states/provinces. Ratios are calculated based on smoothed traveler rates.

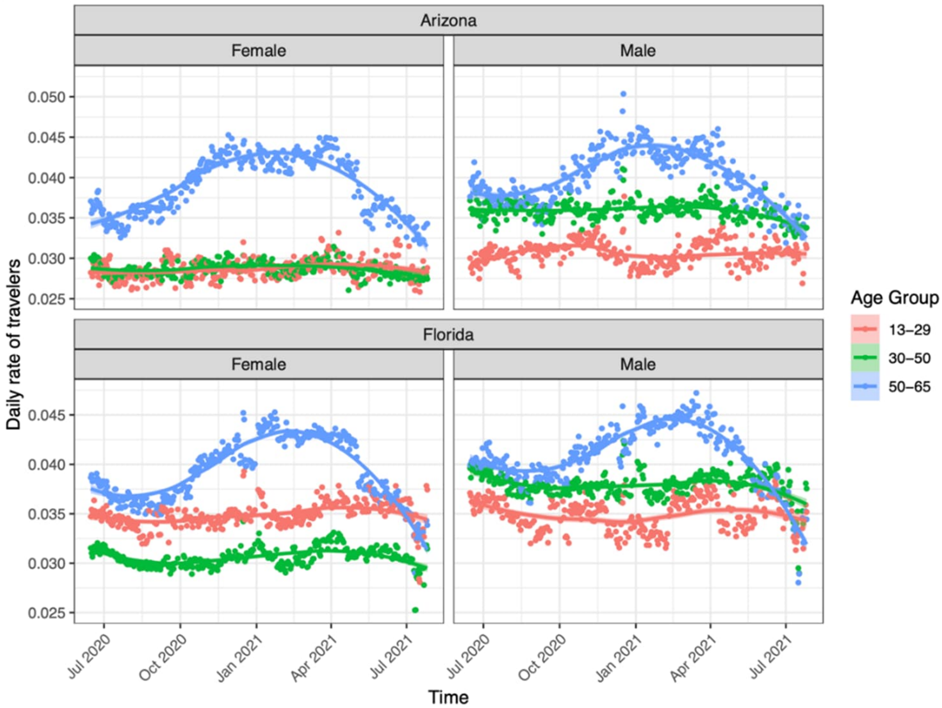


5.1.2 Characteristic seasonal patterns

Group 1: Seasonal fluctuations of in-migration during winter months in warm states

The first group is defined by large seasonal fluctuations of increased traveling during winter months in states that tend to be warm throughout the year. We presume that this influx of travelers is driven by people seeking warmer weather in winter. A prominent characteristic of seasonal winter in-migration is that after the peak is reached, there is a decline in the travel rate following the winter season. This pattern is seen in the 50–65 age group in Arizona and Florida for both male and female travelers (Figure 2). The traveler rate has, on average, a 15.8% relative increase in the number of travelers in these states at their peak compared to July 1, 2020. In contrast, for the 13–29 and 30–50 age groups for both genders, there is no pronounced pattern of seasonal migration in either state.

Figure 2: Daily traveler rate over time, by age group and sex, for Arizona and Florida



Group 2: Seasonal decreases of in-migration during winter months in cold states

The second group serves as an interesting contrast to the first, as there are large decreases in the number of travelers during winter months in states that tend to remain cold throughout the year, likely attributed to people wanting to avoid cold weather. This dip is followed by a consistent increase in the travel rate over the remainder of the time series. This trend is consistent in all the New England states and others, such as New Jersey, Michigan, and Idaho. Figure 3 illustrates this pattern for Maine and Idaho. In comparison to the first group, where the seasonal pattern was apparent for only the oldest age group, this phenomenon is apparent across all age groups, though it is still most prominent in the 50–65 age group for both male and female travelers. Another interesting observation is that the overall rates of traveling are similar across all age groups in this group. In contrast, for Group 1, not only were the seasonal fluctuations of a greater magnitude for the oldest age group but the overall traveler rates were also substantially higher than for other age groups.

Making use of the seasonal summary metric, Table 1 provides an overview of the ten largest observed drops in the travel rate from July 1, 2020, with the associated region, demographic group, and date on which that trough was observed. Geographically, this group is concentrated in the northern parts of the United States, with decreases in traveling rates as high as 41%.

Figure 3: Daily traveler rate over time, by age group and sex, for Idaho and Maine

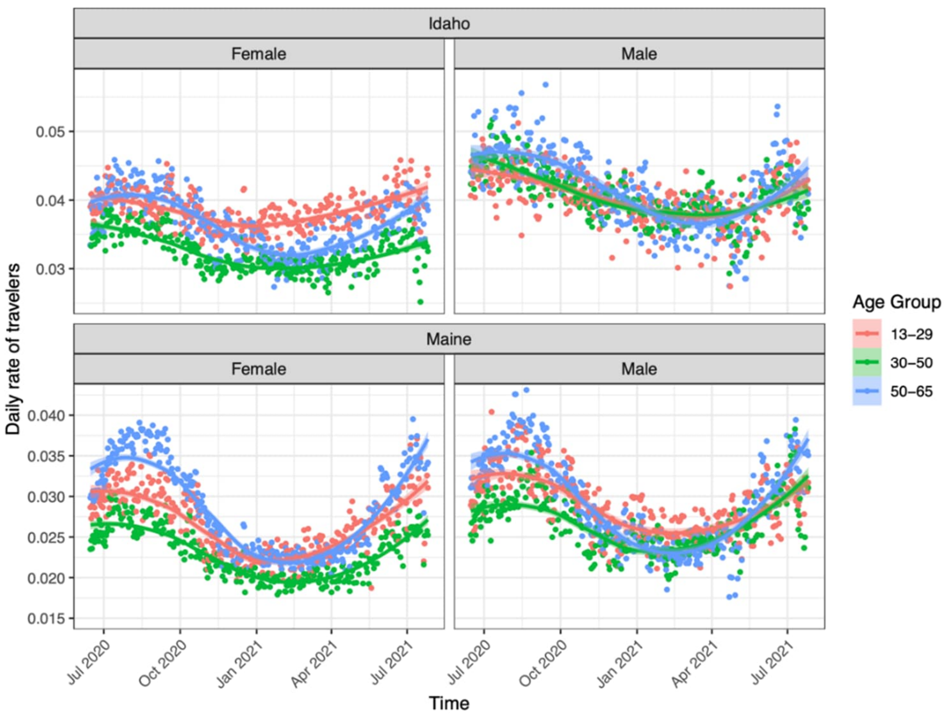


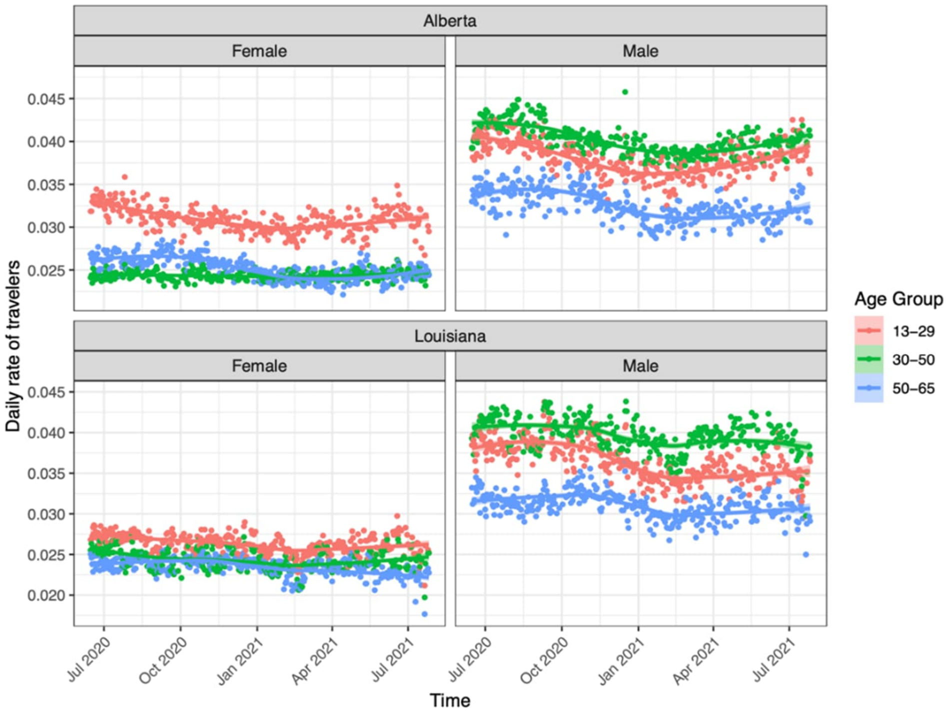
Table 1: Regions and demographic groups for the ten largest observed drops in the travel rate from July 1, 2020, with the date of observation

Date	Sex	Region	Age group	Rate of change from July 1, 2020
3/7/2021	M	Wyoming	50–65	0.4115
3/4/2021	F	Maine	50–65	0.3293
12/28/2020	M	Montana	50–65	0.3258
3/5/2021	M	New Hampshire	50–65	0.3236
3/7/2021	M	South Dakota	50–65	0.3207
2/10/2021	F	Delaware	13–29	0.3033
2/26/2021	M	Rhode Island	50–65	0.2990
2/15/2021	M	Iowa	50–65	0.2957
3/31/2021	M	Vermont	50–65	0.2936
10/14/2020	M	Alabama	13–29	0.2764

Group 3: Male-dominated travel

As noted above, we observed that men of all ages generally have a higher rate of travel than women of all ages, which may be attributed to gender differences in travel for work. To investigate the more extreme gender differences in travel, we used the male-to-female travel ratio metric and isolated those regions where the ratio was greater than or equal to the 75th sample percentile (~ 1.26) across the entire time series. In other words, the ratio of the male-to-female daily travel rate in the regions in this group is consistently in the top 25% of values. Eight regions – Alberta, Louisiana, Manitoba, Mississippi, North Dakota, Saskatchewan, Texas, and Wyoming – fall into this category. Of all the regions, Louisiana has the greatest average male-to-female travel ratio of ~ 1.46. Figure 4 provides two examples of regions characteristic of this group. Here it is clear that the daily rate of travel for men of all age groups is higher than that of women of all age groups, and there are no pronounced seasonal changes in the travel rate across the time series. We hypothesize that many regions in this group have male-dominated travel due to the strong presence of mining and agricultural sectors (Edwards 2013; Molloy, Smith, and Wozniak 2011).

Figure 4: Daily traveler rate over time, by age group and sex, for Alberta and Louisiana

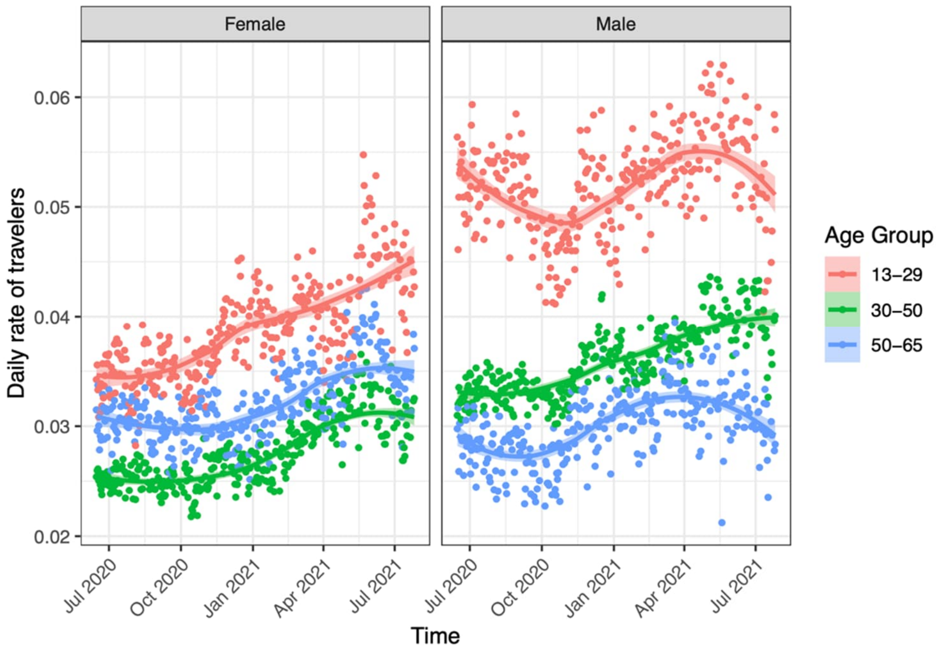


Group 4: Remaining regions (nonseasonal patterns)

The final group comprises regions that did not exhibit any patterns characteristic of the other three pathologies. These regions are Hawaii; Washington, DC; California; Northwest Territories; Yukon; Nunavut; and Prince Edward Island. The latter four have significant noise and sparsity in their data compared to the rest of the regions, so we cannot confidently classify them within our framework. This leaves three regions that exhibit unique nonseasonal changes across the time period. It should be noted that the nonseasonal trends observed in these regions might be associated with changing travel restrictions and behavior during the evolution of the COVID-19 pandemic, given that the period of our data collection covered a time of generally relaxing travel restrictions.

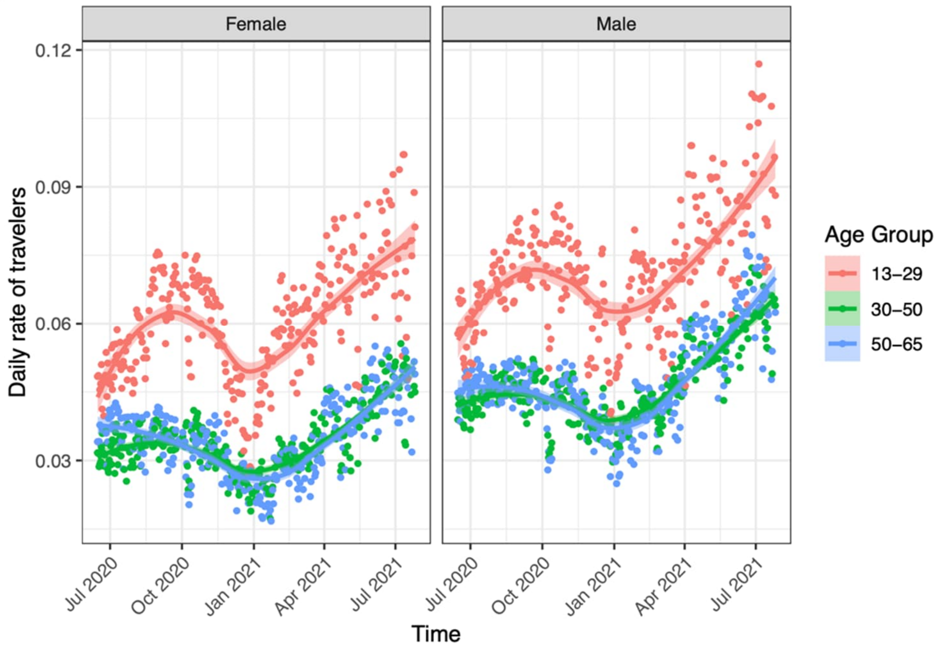
For the case of Hawaii, Figure 5 illustrates a slight but steady increase in the daily travel rate over time for women of all age groups and for men aged 30–50. This increase is most pronounced for women aged 13–29. For men aged 13–29 and 50–65, the rate appears to decrease, increase, and then decrease again. No other regions exhibited a pattern of consistent increase like what we see for most subpopulations in Figure 5. The observed increases could be related to a rebound in tourism after the initial COVID-19 restrictions.

Figure 5: Daily traveler rate over time, by age group and sex, for Hawaii



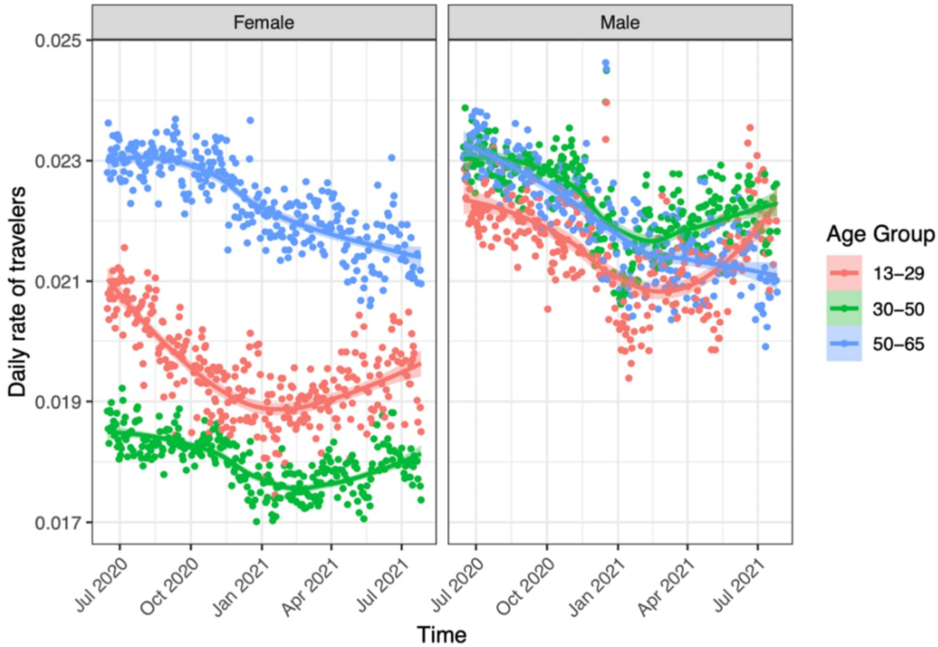
The data for Washington, DC, displayed a trend that is distinct from all other regions. As seen in Figure 6, all demographic groups exhibit the same general trend in the daily travel rate over time but at different magnitudes. The shape of the time series trend is as follows: There is a slight increase in travel from July 2020 to around October 2020, after which the rate decreases to reach a trough around January to February 2021. The rate then steadily increases. The travel rate for the 30–50 and 50–65 age groups is generally the same magnitude while that of the 13–29 age group is consistently larger. This is the case for both male and female travelers; the magnitude for each age group of male travelers is quite close to its female counterpart. It is also worth noting that the magnitude of travel rates observed for this region is substantially higher than what was observed in other regions.

Figure 6: Daily traveler rate over time, by age group and sex, for Washington, DC



Finally, California presents another unique case that is not characteristic of the three previous groups. Figure 7 shows that for the 50–65 age group for both male and female travelers, the daily travel rate steadily decreases over the time series, which we did not observe in any of the other regions. For men and women in the 13–29 and 30–50 age groups, there is a slight decrease in the travel rate from July 2020 to around January to March 2021 and then a slight increase until July 2021. There is also notable variation in the magnitude of the daily travel rate for female travelers across age groups, while for male travelers the travel rate appears to be closer in magnitude across all age groups.

Figure 7: Daily traveler rate over time, by age group and sex, for California

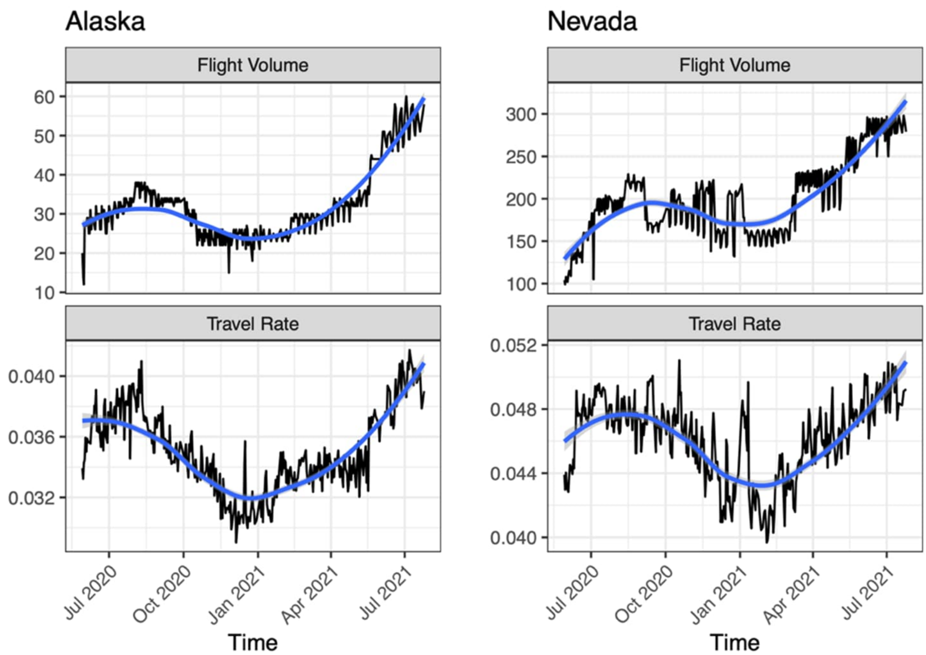


5.2 Comparison with flight volume data

The previous subsection illustrated the potential of traveler rates estimated from Facebook's advertising platform in picking up systematic trends in short-term mobility within a year, particularly seasonal patterns. However, we recognize that by the nature of their source, the data used in this analysis do not provide an accurate representation of the true population of travelers. As short-term mobility data are rarely collected, it is difficult to validate the patterns we see against a gold standard data source, particularly by demographic characteristics. However, we investigated flight volume as another means of assessing patterns in short-term mobility using data from the Bureau of Transportation Statistics (Bureau of Transportation Statistics 2023). Using these data, for each state we compared the number of inbound flights per day to the total daily travel rate, computed from the Facebook data.

Based on this comparison, we found that for a number of states, patterns in flight volume across the time series are comparable to the travel rate across the time series. In other words, patterns of seasonal variation in the travel rate observed in our initial analyses are also being picked up in the flight volume data. As shown in Figure 8, in both Alaska and Nevada, the seasonal patterns in flight volume across the time series are very similar to the seasonality of the travel rate seen across the time series. Specifically, in both states and for both travel metrics there is a trough around the winter season followed by an increase in magnitude over the remainder of the time series.

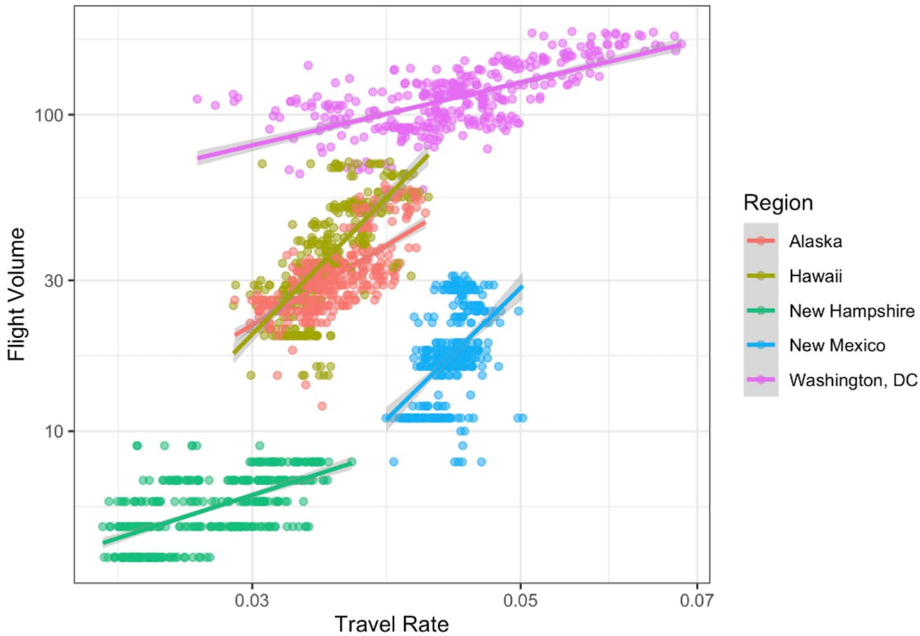
Figure 8: Time series plots of daily flight volume and daily travel rates for Alaska and Nevada



Note: Black lines represent collected counts and blue lines represent smoothed time series using the Loess method.

In general, there was a positive correlation between the flight volume data and the traveler rates (Figure 9 shows five selected areas), although in many states the relationship is not strong. Given that air travel is just one possible mode of transport, we did not expect to see perfect correspondence between these two data sources. An interesting line of future work could be to further investigate the similarities and differences in patterns across regions.

Figure 9: Scatterplots of travel rates and daily flight volumes for five selected regions

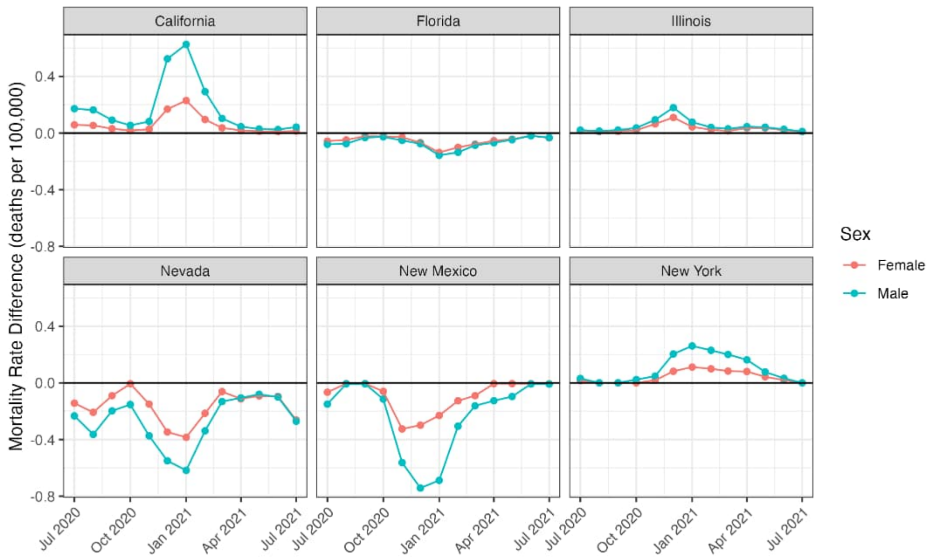


5.3 Application to adjusting COVID-19 mortality rates

Finally, we used seasonal patterns observed in mobility rates in the Facebook data to recalculate mortality rates due to COVID-19 in 2021. We found that adjusting the denominator for net traveler rates led to a 0.6% to 1% change in COVID-19 mortality rates. The largest increases in mortality rates were in California, New York, and Ohio, while the largest decreases were in Nevada, New Mexico, and Arizona. In general, mortality rates in the oldest age group (50- 64-year-olds) were the most impacted, which is a consequence of both relatively high traveler rates and high mortality rates.

Figure 10 illustrates the differences in the unadjusted and adjusted COVID-19 mortality rates for those in the 50–64 age group for six states. Differences generally peaked in January 2021, corresponding to the Alpha variant wave. Differences were higher for males, who generally have higher mortality and higher traveler rates, though this observed difference is driven only by the higher travel rate in this calculation. Florida, Nevada, and New Mexico show adjusted rates that are lower than unadjusted rates due to a net influx of travelers, particularly in the winter months. In general, the adjusted mortality rates showed strong geographic patterning, with southern states showing lower implied mortality rates and northeastern states (such as New York, New Jersey, and the New England states) showing higher implied mortality rates.

Figure 10: Difference in unadjusted and adjusted COVID-19 mortality rates (deaths per 100,000) for those in the 50–64 age group



6. Discussion

In this paper, we introduced a new data source for capturing short-term population movements across time and space. We extracted daily data on the number of people identified as traveling in a particular region from Facebook's advertising platform to construct age-specific rates of travelers across states and provinces in the United States and Canada.

We identified regional, sex, age, and seasonal and nonseasonal trends in travel, which may reflect processes in labor, recreation, and pandemic-related policy. In particular, we observed a higher overall rate of travel for males and wide variations in travel rates for different age groups across all regions. Further, we observed that many colder regions showed clear seasonal declines in travelers in winter, while Florida and Arizona exhibited an increase in travelers aged 50–65 in winter. A group of regions also showed much higher traveler rates for males than for females with very little seasonal pattern, which may suggest travel for economic reasons.

The comparison of the Facebook data with flight volume data from the Bureau of Transportation Statistics showed that the Facebook data are able to capture key seasonal patterns in short-term mobility over time. We also utilized traveler rate estimates obtained from Facebook's advertising platform to calculate age-specific COVID-19 mortality rates, adjusted for seasonal travel fluctuations, and demonstrated that accounting for these short-term movements can increase or decrease the implied mortality rates by up to about 1%, with the impact on estimated rates highest for the oldest age group.

Using traveler data from Facebook's advertising platform to measure short-term mobility has shown potential in terms of timeliness, accessibility, and granularity – attributes that are usually lacking from traditional survey data. However, the main limitation of using such data is the difficulty of validating whether traveler rates calculated from Facebook are representative of true population mobility. Previous work that uses Facebook advertising data for migration estimation focuses on international migrant stocks and flows, estimates of which can be somewhat validated and bias-adjusted based on comparison with data from censuses or nationally representative surveys (Alexander, Polimis, and Zagheni 2020; Zagheni, Weber, and Gummadi 2017). However, the equivalent gold standard data sources do not exist in this context. Future work will investigate the comparison and standardization of these data to other mobility data sources, such as air passenger data, tourism records, and cellphone data.

Another limitation of this particular dataset is the time period of collection. As discussed in Section 3.4, the period July 2020–July 2021 was in the early stages of the COVID-19 pandemic, when some states and provinces had travel restrictions and closures. Patterns observed in this dataset – for example, declines in travelers in California over the period – may have been impacted by these restrictions and may not be characteristic of short-term population mobility more generally. So we are continually collecting these data and plan to keep doing so in the near future to build up a longer time series of observations over multiple years. Finally, note that a limitation of these data is that we don't know where people travel from, only the relative increase or decrease of travelers in a particular region. Given that it's likely that many of the people traveling in the United States are residents, an increase in one area implies a decrease in another area. We cannot currently calculate net decreases because of how the data were collected, but in future work we plan to explore ways of obtaining more information about traveler origin.

Demographic rates are usually calculated with the inherent assumption that the population at risk is in essence static at a particular point in time. However, substantial short-term population mobility – whether for work or leisure – exists over the course of a year, and this short-term movement can affect the population composition and thus the types of people at risk due to prevailing health and mortality conditions in a particular area. This was shown at the state level, but future work on even more granular geographic scales (for instance, movement between rural and urban areas) is likely to enhance the impact even more. We see these data and this work as complementary to other efforts to investigate population mobility at different time scales and as a step forward to introducing more dynamic methods and treatment of population indicators.

References

Ahmad, F. and Cisewski, J. (2023). Quarterly provisional estimates for selected indicators of mortality, 2020-quarter 2, 2022. Hyattsville, MD: National Center for Health Statistics. National Vital Statistics System, Vital Statistics Rapid Release Program.

Alessandretti, L. (2022). What human mobility data tell us about COVID-19 spread. *Nature Reviews Physics* 4(1): 12–13. doi:10.1038/s42254-021-00407-1.

Alexander, M., Polimis, K., and Zagheni, E. (2019). The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data. *Population and Development Review* 45(3): 617–630. doi:10.1111/padr.12289.

Alexander, M., Polimis, K., and Zagheni, E. (2020). Combining social media and survey data to nowcast migrant stocks in the United States. *Population Research and Policy Review* 41:1–28. doi:10.1007/s11113-020-09599-3.

Auliya, S.F. and Wulandari, N. (2021). The impact of mobility patterns on the spread of the COVID-19 in Indonesia. *Journal of Information Systems Engineering and Business Intelligence* 7(1): 31–41. doi:10.20473/jisebi.7.1.31-41.

Bell, M. and Ward, G. (1998). Patterns of temporary mobility in Australia: Evidence from the 1991 census. *Australian Geographical Studies* 36(1): 58–81. doi:10.1111/1467-8470.00039.

Bell, M. and Ward, G. (2000). Comparing temporary mobility with permanent migration. *Tourism Geographies* 2(1): 87–107. doi:10.1080/146166800363466.

Bernard, A. and Perales, F. (2021). Is migration a learned behavior? Understanding the impact of past migration on future migration. *Population and Development Review* 47(2): 449–474. doi:10.1111/padr.12387.

Brownstein, J.S., Wolfe, C.J., and Mandl, K.D. (2006). Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine* 3(10): e401. doi:10.1371/journal.pmed.0030401.

Bureau of Transportation Statistics (2023). Airline on-time statistics, details statistics arrivals [dataset]. https://www.transtats.bts.gov/ontime/arrivals.aspx.

Cebeillac, A. and Vaguet, Y. (2021). What can we learn from Airbnb data on tourist flows? A case study on Iceland. *Espace populations sociétés. Space Populations Societies* 2020/3–2021/1. doi:10.4000/eps.10452.

Dickson, J. (2021). A tale of two COVID-19 hot spots: Ontario and Quebec take different approaches to confronting third wave [electronic resource]. *The Globe and Mail*. https://www.theglobeandmail.com/canada/article-a-tale-of-two-covid-19-hot-spots-ontario-and-quebec-take-different/.

Edwards, W. (2013). Temporary internal migration: Inferences from survey data. *Journal of Northern Studies* 7(2): 75–93.

Fields, R., Humphrey, L., Flynn-Primrose, D., Mohammadi, Z., Nahirniak, M., Thommes, E., and Cojocaru, M. (2021). Age-stratified transmission model of COVID-19 in Ontario with human mobility during pandemic's first wave. *Heliyon* 7(9): e07905. doi:10.1016/j.heliyon.2021.e07905.

Findlater, A. and Bogoch, I.I. (2018). Human mobility and the global spread of infectious diseases: A focus on air travel. *Trends in Parasitology* 34(9): 772–783. doi:10.1016/j.pt.2018.07.004.

Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., and Vinué, G. (2017). Using Twitter data to estimate the relationship between short-term mobility and long-term migration. *Proceedings of the 2017 ACM on web science conference.* New York, NY: ACM: 103–110. doi:10.1145/3091478.3091496.

Grais, R.F., Hugh Ellis, J., and Glass, G.E. (2003). Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology* 18: 1065–1072. doi:10.1023/A:1026140019146.

Iuliano, A.D., Roguski, K.M., Chang, H.H., Muscatello, D.J., Palekar, R., Tempia, S., Cohen, C., Gran, J.M., Schanzer, D., Cowling, B.J., Wu, P., Kyncl, J., Ang, L.W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., Asten, L. van, Pereira da Silva, S., Aungkulanon, S., Buchholz, U., Widdowson, M.-A., and Bresee, J.S. (2018). Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *The Lancet* 391(10127): 1285–1300. doi:10.1016/S0140-6736(17)33293-2.

Jdanov, D.A., Galarza, A.A., Shkolnikov, V.M., Jasilionis, D., Németh, L., Leon, D.A., Boe, C., and Barbieri, M. (2021). The short-term mortality fluctuation data series, monitoring mortality shocks across time and space. *Scientific Data* 8(1): 235. doi:10.1038/s41597-021-01019-1.

Leasure, D.R., Kashyap, R., Rampazzo, F., Dooley, C.A., Elbers, B., Bondarenko, M., Verhagen, M., Frey, A., Yan, J., Akimova, E.T., Fatehkia, M., Trigwell, R., Tatem, A.J., Weber, I., and Mills, M.C. (2023). Nowcasting daily population displacement in Ukraine through social media advertising data. *Population and Development Review* 49(2): 231–254. doi:10.1111/padr.12558.

Marshall, M. and Syed, K. (2020). Map: Coronavirus travel restrictions by state. *The National Broadcasting Company*. https://www.nbcnews.com/news/us-news/map-coronavirus-travel-restrictions-inside-united-states-n1236157.

Meta (2023). Basic targeting. https://developers.facebook.com/docs/marketing-api/audiences/reference/basic-targeting.

Miranda-González, A., Aref, S., Theile, T., and Zagheni, E. (2020). Scholarly migration within Mexico: Analyzing internal migration among researchers using Scopus longitudinal bibliometric data. *EPJ Data Science* 9(34). doi:10.1140/epjds/s13688-020-00252-9.

Molloy, R., Smith, C.L., and Wozniak, A. (2011). Internal migration in the United States. *Journal of Economic Perspectives* 25(3): 173–196. doi:10.1257/jep.25.3.173.

National Center for Health Statistics (2023). Provisional COVID-19 deaths by sex and age [dataset]. https://data.cdc.gov/d/9bhg-hcku.

Panczak, R., Charles-Edwards, E., and Corcoran, J. (2020). Estimating temporary populations: A systematic review of the empirical literature. *Humanities and Social Sciences Communications* 6(87): 1–10. doi:10.1057/s41599-020-0455-y.

Rampazzo, F., Bijak, J., Vitali, A., Weber, I., and 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. doi:10.1215/00703370-9578562.

R Core Team (2022). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Slater, J.J., Brown, P.E., Rosenthal, J.S., and Mateu, J. (2022). Capturing spatial dependence of COVID-19 case counts with cellphone mobility data. *Spatial Statistics* 49: 100540. doi:10.1016/j.spasta.2021.100540.

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., and Rango, M. (2018). Migration data using social media: A European perspective. Luxembourg: Publications Office of the European Union. doi:10.2760/964282.

State, B., Rodriguez, M., Helbing, D., and Zagheni, E. (2014). Migration of professionals to the US: Evidence from LinkedIn data. *Social informatics: International conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014. Proceedings 6*. Cham: Springer: 531–543. doi:10.1007/978-3-319-13734-6_37.

Subbotin, A. and Aref, S. (2021). Brain drain and brain gain in Russia: Analyzing international migration of researchers by discipline using Scopus bibliometric data 1996–2020. *Scientometrics* 126(9): 7875–7900. doi:10.1007/s11192-021-04091-x.

The New York Times (2020). See coronavirus restrictions and mask mandates for all 50 states [electronic resource]. https://web.archive.org/web/20201202223053/https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html.

The New York Times (2021). See reopening plans and mask mandates for all 50 states [electronic resource]. https://web.archive.org/web/20210621230154/https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html.

United States Census Bureau (2023a). 2000 national population projections datasets: Annual projections of the resident population by age, sex, race, and Hispanic origin, 1999–2100: Middle series [dataset]. https://www.census.gov/data/datasets/2000/demo/popproj/2000-detailed-files-age-sex-race-his-origin.html.

United States Census Bureau (2023b). State population by characteristics: 2020–2022 [dataset]. https://www.census.gov/data/datasets/time-series/demo/popest/2020s-state-detail.html.

Vieira, C.C., Fatehkia, M., Garimella, K., Weber, I., and Zagheni, E. (2022). Using Facebook and LinkedIn data to study international mobility. SocArXiv Papers. doi:10.31235/osf.io/v9ay5.

Wilton, D. and Wirjanto, T. (1998). *An analysis of the seasonal variation in the national tourism indicators*. Ottawa: Canadian Tourism Commission.

Xu, Y., Xue, J., Park, S., and Yue, Y. (2021). Towards a multidimensional view of tourist mobility patterns in cities: A mobile phone data perspective. *Computers, Environment and Urban Systems* 86: 101593. doi:10.1016/j.compenvurbsys.2020.101593.

Zagheni, E., Garimella, V.R.K., Weber, I., and State, B. (2014). Inferring international and internal migration patterns from Twitter data. *Proceedings of the 23rd international conference on world wide web*. New York, NY: ACM: 439–444. doi:10.1145/2567948.2576930.

Zagheni, E., Weber, I., and Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review* 43(4): 721–734. doi:10.1111/padr.12102.

Zargari, F., Aminpour, N., Ahmadian, M.A., Samimi, A., and Saidi, S. (2022). Impact of mobility on COVID-19 spread: A time series analysis. *Transportation Research Interdisciplinary Perspectives* 13: 100567. doi:10.1016/j.trip.2022.100567.

Zhao, X., Akbaritabar, A., Kashyap, R., and Zagheni, E. (2023). A gender perspective on the global migration of scholars. *Proceedings of the National Academy of Sciences* 120(10): e2214664120. doi:10.1073/pnas.2214664120.

zu Erbach-Schoenberg, E., Alegana, V.A., Sorichetta, A., Linard, C., Lourenço, C., Ruktanonchai, N.W., Graupe, B., Bird, T.J., Pezzulo, C., Wesolowski, A., and Tatem, A.J. (2016). Dynamic denominators: The impact of seasonally varying population numbers on disease incidence estimates. *Population Health Metrics* 14(35): 1–10. doi:10.1186/s12963-016-0106-0.

Appendix

Below is a list of all airports included in the flights data analysis.

- Alabama – Birmingham–Shuttlesworth International Airport
- Alaska – Ted Stevens Anchorage International Airport
- Arizona – Phoenix Sky Harbor International Airport
- Arkansas – Bill and Hillary Clinton National Airport
- California – Los Angeles International Airport
- California – San Diego International Airport
- California – San Francisco International Airport
- Colorado – Denver International Airport
- Connecticut – Bradley International Airport
- Delaware – Wilmington Airport
- Florida – Fort Lauderdale–Hollywood International Airport
- Florida – Miami International Airport
- Florida – Orlando International Airport
- Florida – Tampa International Airport
- Georgia – Hartsfield–Jackson Atlanta International Airport
- Hawaii – Daniel K. Inouye International Airport
- Idaho – Boise Airport
- Illinois – Midway International Airport
- Illinois – O’Hare International Airport
- Indiana – Indianapolis International Airport
- Iowa – Des Moines International Airport
- Kansas – Wichita Dwight D. Eisenhower National Airport
- Kentucky – Cincinnati/Northern Kentucky International Airport
- Louisiana – Louis Armstrong New Orleans International Airport
- Maine – Portland International Jetport
- Maryland – Baltimore/Washington International Airport
- Massachusetts – Boston Logan International Airport
- Michigan – Detroit Metropolitan Airport
- Minnesota – Minneapolis–Saint Paul International Airport
- Mississippi – Jackson–Evers International Airport
- Missouri – St. Louis Lambert International Airport

- Montana – Bozeman Yellowstone International Airport
- Nebraska – Eppley Airfield
- Nevada – Harry Reid International Airport
- New Hampshire – Manchester–Boston Regional Airport
- New Jersey – Newark Liberty International Airport
- New Mexico – Albuquerque International Sunport
- New York – John F. Kennedy International Airport
- New York – LaGuardia Airport
- North Carolina – Charlotte Douglas International Airport
- North Dakota – Hector International Airport
- Ohio – Cleveland Hopkins International Airport
- Oklahoma – Will Rogers World Airport
- Oregon – Portland International Airport
- Pennsylvania – Philadelphia International Airport
- South Carolina – Charleston International Airport
- South Dakota – Sioux Falls Airport
- Tennessee – Nashville International Airport
- Texas – Austin–Bergstrom International Airport
- Texas – Dallas Fort Worth International Airport
- Texas – George Bush Intercontinental Airport
- Utah – Salt Lake City International Airport
- Vermont – Burlington International Airport
- Virginia/Washington, DC – Ronald Reagan Washington National Airport
- Virginia/Washington, DC – Washington Dulles International Airport
- Washington – Seattle–Tacoma International Airport
- West Virginia – Yeager Airport
- Wisconsin – Milwaukee Mitchell International Airport
- Wyoming – Casper–Natrona County International Airport