Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections

Chenfeng Xiong[a,b,1], Songhua Hu[a], Mofeng Yang[a], Weiyu Luo[a], and Lei Zhang[a,1]

[a]Maryland Transportation Institute, Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742; and [b]Shock Trauma and Anesthesiology Research Center, School of Medicine, University of Maryland, Baltimore, MD 21201

Edited by Arild Underdal, University of Oslo, Oslo, Norway, and approved September 12, 2020 (received for review May 29, 2020)

Accurately estimating human mobility and gauging its relationship with virus transmission is critical for the control of COVID-19 spreading. Using mobile device location data of over 100 million monthly active samples, we compute origin–destination travel demand and aggregate mobility inflow at each US county from March 1 to June 9, 2020. Then, we quantify the change of mobility inflow across the nation and statistically model the time-varying relationship between inflow and the infections. We find that external travel to other counties decreased by 35% soon after the nation entered the emergency situation, but recovered rapidly during the partial reopening phase. Moreover, our simultaneous equations analysis highlights the dynamics in a positive relationship between mobility inflow and the number of infections during the COVID-19 onset. This relationship is found to be increasingly stronger in partially reopened regions. Our study provides a quick reference and timely data availability for researchers and decision makers to understand the national mobility trends before and during the pandemic. The modeling results can be used to predict mobility and transmissions risks and integrated with epidemics models to further assess the public health outcomes.

COVID-19 | mobile device location data | mobility | partial reopening

COVID-19 is one of the worst global health crises seen in decades. One approach to delay its spread is to reduce human travel by imposing nonpharmaceutical interventions such as self-isolation that were deemed effective in other countries such as China and Italy (1–3). However, in the United States, where there were over 75,000 confirmed deaths as of the first week of May 2020, people were still engaging in a considerable amount of travel, although social distancing and stay-at-home orders were issued in most regions (e.g., refs. 4 and 5). Human mobility further increased with the release of guidelines by the White House for reopening America in the third week of April. As of May 1, 18 states had loosened restrictions on their "stay-at-home" orders (6).

Further, increased mobility may positively contribute to the spread of COVID-19. For instance, Jia et al. (7) and Badr et al. (8) found a high correlation between the number of infections and mobility in each prefecture/county in China and the United States, using mobile phone location data. In the United States, the detailed daily interregional travel mobility patterns at the county level are not yet available to inform the public. While the existing studies have revealed the positive correlation between mobility and transmissions, the dynamics of the relationship between these two are not yet quantified.

To overcome these knowledge gaps and provide rapid evidence for policy making, we use mobile device location data and validated algorithms to derive daily trips and aggregate them into county-level mobility inflow data. This aggregated information is made available to the public via our COVID-19 impact analysis platform (https://data.covid.umd.edu) and has already supported a number of agencies' daily decision-making. Then we highlight the pattern changes in mobility inflow in US counties and find a time-dependent relationship between inflow and infections in destination counties via a simultaneous equations modeling process.

Results

Using anonymized mobile device location data from over 100 million monthly active samples, we observe dramatic changes in mobility patterns across the United States. Compared to the reference metrics in January 2020 (excluding New Year's Day, January 1, and Martin Luther King Jr. Day, January 20), mobility is estimated to drop by as much as 35% nationally and then gradually rebound, especially after partial reopening. (These results and data are reported in https://github.com/SonghuaHu-UMD/Mobility_COVID19_PNAS).

We further analyze a critical metric for decision support (9), the daily origin–destination (OD) matrices at the county level, to assess interregional mobility of the population. Fig. 1 illustrates the percentage change in the OD mobility inflow, that is, the amount of travel flowing into a specific county from other places. Three phases are compared in Fig. 1: 1) prepandemic and no social distancing (March 2 to 8), 2) stay-at-home orders and mobility reduction (March 30 to April 5), and 3) partial reopening and mobility increase (April 27 to May 3). Compared to January, slightly heavier intercounty travel is observed during the prepandemic period, reflecting seasonal mobility patterns. After the nation entered emergency on March 13, most states issued "stay-at-home" orders, and the peak of social distancing behavior took place in the first half of April (mobility inflow decreased by about 35%). By late April, several states announced partial reopening to varying degrees, while fatigue in following social distancing was shown. Nationally, mobility inflow has recovered to only 20% lower than the typical inflow observed in January. This increasing trend in travel across county borders is found to be consistent with the other mobility measures that still remained at a fraction of their typical patterns during the partial reopening phase (e.g., the average daily person-miles traveled was still 22% lower than the January benchmark).

The increase of inflow in the "reopened" regions during the partial reopening phase is notably higher (Fig. 1*C*). On average, these counties' mobility inflows have recovered to 86% of the January numbers, while the inflows for the counties remaining locked down were only 73% recovered. Significantly more external visits bring in more exposure and epidemiological risks. A surge in the

Author contributions: C.X. and L.Z. designed research; C.X., S.H., M.Y., and W.L. performed research; S.H., M.Y., and W.L. analyzed data; and C.X., S.H., and L.Z. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: cxiong@umd.edu or lei@umd.edu.

First published October 15, 2020.

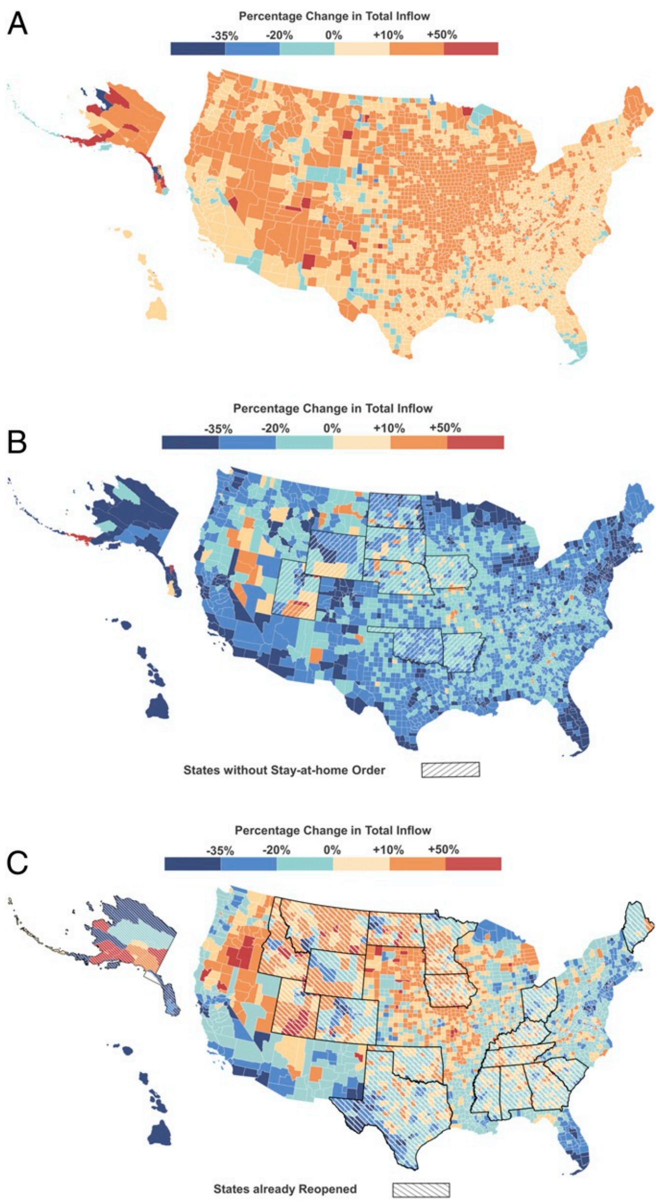

Fig. 1. Percentage change in total inflow of counties in contiguous United States in different time periods, compared to the benchmark mobility in January 2020. (*A*) Pre-pandemic, no social distancing (March 2 to March 8). (*B*) Stay-at-home orders and mobility reduction (March 30 to April 5). (*C*) Partial reopening and mobility increase (April 27 to May 3).

number of daily confirmed cases has been observed in reopened regions such as Kentucky, North Carolina, and Minnesota.

The dramatically increased mobility inflow may be hypothesized to transmit the virus within and across locations by "community transmission" and/or "importation." To capture the time-varying relationship between new cases and inflow trips, a simultaneous equations model (SEM) with dynamic panel and time-varying coefficients is employed. This model, as shown in *Materials and Methods*, uses a system of equations to depict the joint dependencies of the two variables of interest. The autocorrelations of new cases and inflows, lagged effect, and spatial and temporal heterogeneity are incorporated in the model. We divide the counties into two groups based on whether the state that each county is in has the effective reopen order before or after May 1 (referred to as reopened group and locked-down group hereafter). The model is then fitted separately for the entire country and for each county group, to compare the time-varying relationships. Most parameter estimates of the SEM model are found to be statistically significant at the 95% confidence level. The average R-squared statistics are 0.669 and 0.986 for the number of infections submodel and the mobility submodel for the nationwide SEM. We find similar goodness-of-fit measures for the models of the two county groups.

In accordance with the hypothesis, we find a strong and positive relationship between the mobility inflow and the number of infections in US counties, as depicted by the nationwide coefficient curve in Fig. 2*B*. The average coefficient is 0.243, indicating that, if a 10% increase is observed in the inflow today, given other things being equal, we expect to see a 2.34% increase in the number of infections a week later (i.e., model lag = 7). The dynamics in the relationship are also captured. The positive relationship sharply increased during the COVID-19 onset and then dampened when states gradually entered "locked-down" status. About 2 wk after the release of reopening guidelines, it started to bounce back.

Another interesting finding is the difference in the SEM coefficients for the two county groups. The coefficient of the locked-down group is consistently higher than that of the reopened group across the study period, indicating a higher risk embedded in the mobility inflow in locked-down counties. After all, traveling would be much less safe when there are more daily confirmed cases in the locked-down counties (Fig. 2*A*). The locked-down group's coefficient has a more dramatic drop after the first week of April, suggesting that travel carried less influence during that period, due to the implementation of "lock-down" and social distancing rules. It then bounces back quickly in May when people lose "social distancing" momentum. While the reopened group's coefficient exhibits much smoother dynamics, the gap between the coefficients of the two groups becomes increasingly smaller after 19 states announce reopening. It raises a concern about the epidemiological risks of partial reopening: that more external travel could lead to more coronavirus hospitalizations.

Conclusion

Accurately estimating human mobility and gauging its relationship with virus transmission during pandemic is critical for control of the spread of COVID-19 and any other highly contagious disease. A key contribution of the study lies in the daily updated OD travel demand analytics and mobility inflow for each of the 3,141 US counties, using mobile device location data. The analytics made available to the public reveal daily intercounty travels and has already provided timely support to a number of decision makers. Another contribution is that we robustly characterize the dynamics in a positive relationship between mobility inflow and the number of infections via a simultaneous equations modeling process with time-varying coefficients. And this positive relationship gets steadily amplified in reopened regions. Our findings warn us about premature loosening of restrictions and that a second spike in coronavirus could be a likely scenario in many early-opening regions even though people are still urged to stay at home unless necessary. The analysis and data provide a timely reference for researchers and decision makers about human mobility trends in the nation. The dynamics on how mobility influences COVID-19 infections are estimated, which can be used in predictions and integration with agent-based travel and epidemics models (see, e.g., ref. 10) to further assess the public health consequences of decisions such as reopening, school closure, etc. With proper incorporation of spatial correlation, the study can also be extended to critical urban and suburban areas with fine-grained mobility at the census-block level.

Materials and Methods

Data. We use a data panel of integrated and processed mobile device location data procured from multiple third-party data providers. The data

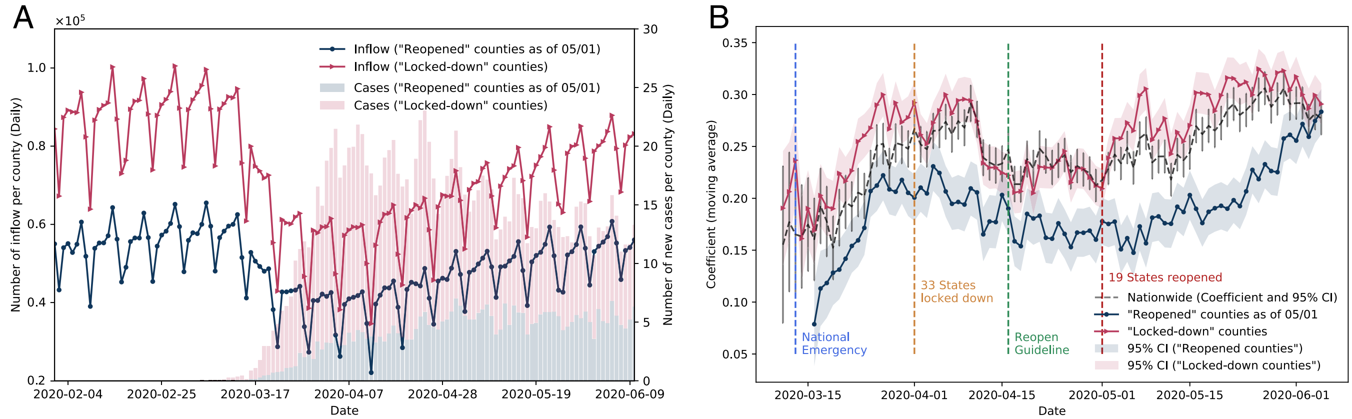

Fig. 2. The dynamics in the relationship between mobility inflow and county confirmed COVID-19 cases. (*A*) The evolution of mobility inflow and number of daily confirmed cases in reopened counties and locked-down counties. (*B*) Dynamic relationship between mobility inflow and daily confirmed cases in counties, indicated by the time-varying coefficient (α_{2T}) of the simultaneous equations model.

include daily movements of over 100 million anonymous, opted-in, and monthly active individuals, from January 1, 2020 to June 9, 2020. Then, a cloud-based computing platform is developed and deployed using a validated spatial–temporal algorithm (11), ingesting over 60 TB of data and utilizing over 75,000 CPU hours of computation to identify all trips. All computed metrics are made available to the general public via the COVID-19 analysis platform developed by the authors (https://data.covid.umd.edu/).

Measure. We study the US county-level daily mobility inflow by measuring the origin and destination information of all observed trips. If a trip's origin and destination are not in the same county, this trip will be counted in the mobility inflow for the destination county. Mobility measures used in other studies are also evaluated and shared (see our Github, mentioned in *Results*).

Modeling Method. To capture the time-varying relationship between the number of infections and mobility inflow, we have developed a SEM with dynamic panel and time-varying coefficients. For each day T, the dynamic panel is constructed on the time period of $[T, T+B)$, where B is the bandwidth of the dynamic panel.

$$\ln y_{it} = \alpha_{0T} + \alpha_{1T} \cdot \ln y_{i,t-1} + \alpha_{2T} \cdot \ln x_{i,t-l} + \gamma_T \cdot \mathbf{X}_{it} + \mu_i + r_t + \epsilon_{it} \qquad [1]$$

$$\ln x_{i,t-l} = \beta_{0T} + \beta_{1T} \cdot \ln x_{i,t-l-1} + \delta_T \cdot \mathbf{Z}_{i,t-l} + \mu_i' + r_{t-l}' + \epsilon_{i,t-l}', \qquad [2]$$

where y_{it} is the daily number of confirmed new cases in county i for day t, and $t \in [T, T+B)$; $x_{i,t-l}$ is the daily weighted mobility inflow received by county i from all of the other counties during day $t-l$; l denotes the time lag; α, β, γ, and δ are the time-varying coefficients for each day T; $\mathbf{X}$ and $\mathbf{Z}$ are the vectors of exogenous variables; μ, μ', r, and r' are the county-fixed effect and time-fixed effects to eliminate the spatial and temporal heterogeneity (12); and ϵ_{it} and ϵ_{it}' are the idiosyncratic error terms. Logarithmic transformation is employed here to address the nonnormality problem (13). The bandwidth B is set as 7 d to eliminate the disturbance of temporal fluctuation (14). First, the number of daily trips presents a 7-d weekly pattern. Second, $>$90% of the dynamic panels are found to be stationary, indicating the robustness of the model. Third, the mean incubation period of COVID-19 is proven to be 4.1 d to 7.0 d, with 95% confidence (15). More information about the model, including the selection and estimation of lag variable and exogenous variables, can be found in our Github. As a robustness check, we also evaluate models using different time windows and find consistent patterns.

Data Availability. The daily number of confirmed cases in each county was retrieved from Johns Hopkins University's COVID-19 Dashboard (16). County-level socio-demographic information was retrieved from U.S. Census Bureau's American Community Survey (17). The weather information was retrieved from National Centers for Environmental Information (18). Codes and spreadsheet data have been deposited in Github (https://github.com/SonghuaHu-UMD/Mobility_COVID19_PNAS). All computed metrics are made available to the general public via the COVID-19 analysis platform developed by the authors (https://data.covid.umd.edu/).

ACKNOWLEDGMENTS. We thank our partners: 1) various mobile device data providers and 2) Amazon Web Service and its Senior Solutions Architect, Jianjun Xu.

1. M. Chinazzi *et al.*, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
2. C. R. Wells *et al.*, Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7504–7509 (2020).
3. S. M. Moghadas *et al.*, Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9122–9126 (2020).
4. C. Xiong *et al.*, Data-driven modeling reveals the impact of stay-at-home orders on human mobility during the COVID-19 pandemic in the U.SarXiv:2005.00667. (2 May 2020).
5. A. Aktay *et al.*, Google COVID-19 community mobility reports: Anonymization process description (version 1.0)arXiv:2004.04145. (8 April 2020).
6. . New YorkTimes, See how all 50 states are reopening. https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html. Accessed 13 May 2020.
7. J. S. Jia *et al.*, Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).
8. H. S. Badr *et al.*, Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *Lancet Infect. Dis.*, 10.1016/S1473-3099(20)30553-3 (2020).
9. N. Oliver *et al.*, Mobile phone data and COVID-19: Missing an opportunity? arXiv:2003.12347 (27 March 2020).
10. A. Najmi, F. Safarighouzhdi, E. J. Miller, R. MacIntyre, T. H. Rashidi, Determination of COVID-19 parameters for an agent-based model: Easing or tightening control strategies. medRxiv:10.1101/2020.06.20.20135186 (23 June 2020).
11. L. Zhang, S. Ghader, "Data analytics and modeling methods for tracking and predicting origin-destination travel trends based on mobile device data" (Project Report, Federal Highway Administration Exploratory Advanced Research Program, 2020).
12. B. Baltagi, *Econometric Analysis of Panel Data* (John Wiley, 2008).
13. H. Fang, L. Wang, Y. Yang, "Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China" (Tech. Rep. 26906, National Bureau of Economic Research, 2020).
14. A. Cori, N. M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).
15. Q. Li *et al.*, Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
16. E. Dong, H. Du, L. Gardner, COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://github.com/CSSEGISandData/COVID-19. Accessed 12 June 2020.
17. U.S. Census Bureau, American Community Survey (ACS), 5-year estimates for county-level socio-demographics. https://data.census.gov/cedsci/. Accessed 12 June 2020.
18. National Centers for Environmental Information, Global Historical Climatology Network Place. ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/. Accessed 12 June 2020.