



OPEN

## Novel mobility index tracks COVID-19 transmission following stay-at-home orders

Peter Hyunwuk Her<sup>1,2,8</sup>, Sahar Saeed<sup>3,4,8</sup>, Khai Hoan Tram<sup>5</sup> & Sahir R Bhatnagar<sup>6,7</sup>✉

Considering the emergence of SARS-CoV-2 variants and low vaccine access and uptake, minimizing human interactions remains an effective strategy to mitigate the spread of SARS-CoV-2. Using a functional principal component analysis, we created a multidimensional mobility index (MI) using six metrics compiled by SafeGraph from all counties in Illinois, Ohio, Michigan and Indiana between January 1 to December 8, 2020. Changes in mobility were defined as a time-updated 7-day rolling average. Associations between our MI and COVID-19 cases were estimated using a quasi-Poisson hierarchical generalized additive model adjusted for population density and the COVID-19 Community Vulnerability Index. Individual mobility metrics varied significantly by counties and by calendar time. More than 50% of the variability in the data was explained by the first principal component by each state, indicating good dimension reduction. While an individual metric of mobility was not associated with surges of COVID-19, our MI was independently associated with COVID-19 cases in all four states given varying time-lags. Following the expiration of stay-at-home orders, a single metric of mobility was not sensitive enough to capture the complexity of human interactions. Monitoring mobility can be an important public health tool, however, it should be modelled as a multidimensional construct.

While highly effective vaccines are readily available in the United States, uptake remains low<sup>1,2</sup> and interventions aimed at minimizing human contact continue to be necessary to mitigate the spread of SARS-CoV-2<sup>3–8</sup>. The potential of monitoring population-level mobility patterns using geolocated mobile phone data as a public health tool has been demonstrated<sup>9–12</sup>. In March 2020, worldwide adherence to lockdowns was measured using various mobility metrics<sup>13–18</sup>. A modelling study from China showed 20–60% reductions in mobility notably controlled the spread of SARS-CoV-2<sup>19</sup>. A study from Canada showed that reductions in mobility strongly predicted future control of SARS-CoV-2 growth rates<sup>6</sup>. However, in the absence of social distancing interventions, the association between changes in population-level mobility and COVID-19 remains unclear<sup>11,12</sup>. This is particularly important now as highly transmissible variants of concerns (Delta, Omicron) have become dominant strains<sup>20</sup> and transmission is likely to occur before clinical cases are diagnosed<sup>21</sup>. Additionally, relying on case detection alone to predict surges in transmission continues to underestimate the pandemic as molecular diagnostic tests have become scarce worldwide<sup>22</sup>. Thus, the need to rapidly identify populations and locations at heightened risk of exposure remains necessary.

Population-level mobility, as it pertains to human interaction, is multidimensional. This is particularly true when assessing distinct geographical areas that vary by population density and socioeconomic factors across the United States<sup>23,24</sup>. While the measurement of mobility is complex, most studies to date have used single metrics such as the percentage of people remaining at home or changes in the distance travelled to summarize human interactions and evaluate trends and associations with COVID-19<sup>9</sup>. These single metrics may oversimplify mobility associated with human interaction. As social distancing policies loosened from strict “lock down” to business-as-normal, the utility of continuously monitoring mobility may require a robust definition that is able to capture the complexity of population-level movement<sup>25</sup>. To this end, the aim of this study was to use advanced statistical methods to create a novel index that summarizes mobility as a latent construct using a combination of six mobility metrics. We evaluated how our mobility index varied across 365 counties in 4 states as a function of

<sup>1</sup>Department of Pharmacology and Therapeutics, McGill University, Montreal, Canada. <sup>2</sup>Department of Medical Biophysics, University of Toronto, Toronto, Canada. <sup>3</sup>Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, USA. <sup>4</sup>Department of Public Health Sciences, Queen's University, Ontario, Canada. <sup>5</sup>Division of Infectious Diseases, Department of Medicine, University of Washington, Seattle, USA. <sup>6</sup>Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada. <sup>7</sup>Department of Diagnostic Radiology, McGill University, Montreal, Canada. <sup>8</sup>These authors contributed equally: Peter Hyunwuk Her and Sahar Saeed. ✉email: sahir.bhatnagar@mcgill.ca

time. Finally, we assessed the validity of our mobility index by evaluating how mobility correlated with COVID-19 cases from four states in the Midwest compared to a single metric from the time stay-at-home orders expired.

## Methods

**Data sources.** *Mobility metrics.* We used aggregated mobility data publicly available through SafeGraph from January 1 to December 8, 2020, via the Social Distancing Metric database. SafeGraph uses a panel of GPS pings from anonymous mobile devices from a representative sample of the US Census population to derive metrics of mobility. These data includes a range of spatial behaviors from >45 million mobile devices ( $\approx 10\%$  of devices in the United States). To enhance privacy, SafeGraph excludes census block group information if fewer than two devices visited an establishment in a month.

*A priori*, we choose six mobility metrics commonly used in the literature as a proxy of human contact and that could be attributable to mobility behavior changes as associated with COVID-19 infections. Each metric is defined for a given day ( $t$ ) for a given county ( $j$ ). The metrics ( $s$ ) included:

- The fraction of devices leaving home in a day
- The fraction of devices away from home for 3–6 hours (Part-time work behavior)
- The fraction of devices away from home longer than 6 hours (Full-time work behavior)
- The median time spent away from home
- The median distance traveled from home
- The average number of short stops (>3 stops for less than 20 min) (Delivery behaviors).

*COVID-19 cases.* Confirmed COVID-19 cases data were retrieved from the New York Times open-source project<sup>26</sup>. This publicly available dataset aggregates county-level daily counts of diagnosed cases, from Health services' official reports. All methods were carried out in accordance with the relevant guidelines.

*Covariates.* Demographic variables including population size and population density for each county were collected from the American Community Survey, data available through the the US Census Bureau. To capture the variability of health, social, and economic factors across counties we used the COVID-19 Community Vulnerability Index (CCVI)<sup>27,28</sup>. The CCVI is similar to the Social Vulnerability Index, which was developed by the CDC to support disaster management<sup>29</sup> except with additional elements specific to the COVID-19 pandemic. The index combines 40 indicators of vulnerability from seven themes (socioeconomic status, minority status, household composition, epidemiology, healthcare system, high risk environments, population density) to create the county and state-level index. The CCVI is recognized by the Centers for Disease Control and Prevention as a valuable tool in COVID-19 research and pandemic response planning<sup>30,31</sup>. Dates for when stay-at-home orders were enforced and lifted were obtained from the state-level social distancing policies database<sup>32</sup>. There were several entries for each state due to policy revisions or updates. We used the first entry for each state in this database.

**Analysis.** *Population.* Given the magnitude of the available data, we reduced the number of states in our analysis to four populous and neighboring states in the Midwest. These states had similar numbers of counties and socioeconomic indicators. The counties from each state represented varying population densities and rural vs. urban areas. Based on the 2020 Presidential elections, Illinois (IL) and Michigan (M) were considered Democratic states while Ohio (OH) and Indiana (IN) were predominately Republican states. We used every county from each of the selected states to avoid any preferential selection.

*Mobility index.* We defined mobility as a change of each mobility metric relative to the average of the week before (time-updated rolling average). For each county  $j = 1, \dots, 365$ , we index each of the 6 mobility metrics  $s = 1, \dots, 6$  by calendar day  $t = 1, \dots, m_j$ , where  $m_j$  is the total number of observed days since re-opening in county  $j$ . We define the following quantities:

- $X_{j,t,s}$ : the scalar value of mobility metric  $s$  measured on day  $t$  in county  $j$ .
- $\mathbf{X}_{j,t-8,\dots,t-1,s} = (X_{j,t-8,s}, \dots, X_{j,t-1,s}) \in \mathbb{R}^8$ : the value of mobility metric  $s$  measured on days  $t-8, \dots, t-1$ , i.e., the 7 days prior to day  $t$  in county  $j$ . This is a vector quantity.
- $\bar{\mathbf{X}}_{j,t-8,\dots,t-1,s}$ : the average of the  $\mathbf{X}_{j,t-8,\dots,t-1,s}$ .

The change from baseline mobility metric  $s$  for day  $t$  in county  $j$  is given by

$$\Delta X_{j,t,s} = \frac{X_{j,t,s} - \bar{\mathbf{X}}_{j,t-8,\dots,t-1,s}}{\bar{\mathbf{X}}_{j,t-8,\dots,t-1,s}} \quad (1)$$

The use of a rolling average is unique to this analysis. Most studies have used a static relative baseline period such as mobility trends between January until February 2020<sup>9,33,34</sup>. This common approach does not account for seasonal mobility variability or changes as a result of the pandemic<sup>35–37</sup>. In contrast, our baseline (rolling average) takes into consideration temporal trends that were likely changing with evolving public health policies. To check for outliers and appropriateness of the use of the average, we calculated the coefficient of variation (CV) for each individual metric in each county (Supplemental Figure S13). We found that there were no strong outliers, as all the CV were less than 2.5, suggesting that the average was a valid metric to use. We also considered using the median but did not find a significant difference in the results relative to the average.

Since our hypothesis was each metric could be attributed to a common underlying notion of mobility, we used an unsupervised machine learning method known as functional principal component analysis (fPCA) to create our latent mobility index<sup>38</sup>. Briefly, PCA is a technique for reducing the dimensionality of multiple variables while minimizing information loss. This is done by creating new uncorrelated variables (principal components) that successively maximize variance. A “functional” PCA accounts for the longitudinal nature of the data. We applied fPCA on  $\Delta X_{j,t,s}$  separately for each county and extracted the first principal component, i.e., the linear combination of individual mobility metrics that explained the most variance. We denote this first principal component by  $fPCA_{j,t}$ , a score summarizing mobility in each county ( $j$ ) on a given day ( $t$ ). To enable comparability between counties and states,  $fPCA_{j,t}$  were scaled as Z-scores, which defined our mobility index (MI) given by:

$$MI_{j,t} = \frac{fPCA_{j,t} - \overline{fPCA_{j,t}}}{\sqrt{\text{Var}(fPCA_{j,t})}}, \quad (2)$$

where  $\overline{fPCA_{j,t}} = \frac{1}{m_j} \sum_{k=1}^{m_j} fPCA_{j,k}$  and  $\text{Var}(fPCA_{j,t})$  are the average and variance of the fPCA scores in county  $j$  over the observed time period, respectively.

The interpretation of  $MI$  is as follows:  $MI = 0$ , on average there was no change in mobility relative the previous week;  $MI = 1$  on average there was an increase in mobility by one standard deviation relative to the last week and:  $MI = -1$  on average there was a decrease in mobility by one standard deviation relative to the last week. An animation was created to visualize the relative daily changes of  $MI$  by counties.

**Association with COVID-19.** For each county  $j$ , let  $y_{j,t}$  be the number of confirmed COVID-19 cases on day  $t = 1, \dots, m_j$ , and  $\mathbf{q}_{j,t} = [MI_{j,t-0}, \dots, MI_{j,t-21}]$  denote the vector of lagged occurrences of our mobility index (defined in Eq. (2)) with 0 days and 21 days as minimum and maximum lags, respectively. In words, the first element of  $\mathbf{q}_{j,t}$  represents the value of our mobility index on day  $t$ , the second element represents the value of our mobility index one day prior to  $t$ , and so on. From the time stay-at-home orders expired until December 8 2020, the relationship between daily counts of COVID-19 cases ( $y_{j,t}$ ) and mobility ( $\mathbf{q}_{j,t}$ ), accounting for up to 21 days of lag, was estimated with a quasi-Poisson hierarchical generalized additive model (HGAM)<sup>39,40</sup> of the form:

$$\log(E(y_{j,t})) = \beta_0 + s(\mathbf{q}_{j,t}) + s(\text{time}_t) + s_j(\text{time}_t) + CCVI_j + \text{density}_j, \quad (3)$$

where  $\beta_0$  is the intercept,  $s(\cdot)$  are the smooth non-parametric functions of the predictor variables,  $CCVI_j$  is the COVID-19 Community Vulnerability Index and  $\text{density}_j$  is the population density (people per square kilometer) at the county-level. The term  $s(\mathbf{q}_{j,t})$  in Eq. (3) captures the potentially non-linear and delayed effect of mobility on COVID-19 cases through a cross-basis function<sup>41</sup>. We used penalized cubic regression splines<sup>39</sup> for both dimensions, with interior knots placed at Z-scores of  $-3, -2, -1, 0, 1, 2, 3$  for  $MI_{j,t}$ , and 7 and 14 days for the lag. The penalty is given by  $\lambda \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta}$ , where  $\boldsymbol{\beta}$  are the regression parameters,  $\mathbf{S}$  is a matrix of known coefficients<sup>42</sup>, and  $\lambda$  is the tuning parameter that controls the degree of smoothing and chosen via generalized cross-validation. Given the heterogeneity of COVID-19 epidemiology across counties, models included both a state level calendar time effect  $s(\text{time}_t)$  using thin plate regression splines<sup>43</sup> and a county level calendar time effect  $s_j(\text{time}_t)$  using a factor-smoother interaction basis<sup>40</sup>. Population size was used as offset in each model. A separate model was run for each of the selected states.

There were two main advantages for using a HGAM to evaluate the association between mobility and COVID-19 cases: (1) it can quantify the non-linear functional relationships over time where the shape of each function varies across counties, and (2) it has the capacity of modelling varying lags<sup>44</sup>. Based on a recent systematic review of 42 studies, the mean incubation period of SARS-CoV-2 was 8 days (95% CI 10.3, 16 days)<sup>45</sup>. This lag between time of infection and becoming symptomatic/testing positive can vary at both the patient-level (the time it takes to develop symptoms and get tested) and also at the county-level (the time it takes to perform and report the test results). Given this variation, we were able to control for varying lagged exposures (up to 21 days) at the county-level. To evaluate the utility of our mobility index, we compared a dose-response relationship between mobility and COVID-19 cases and goodness-of-fit-statistics of our latent MI compared to a single measure of mobility (the fractions of devices leaving the home). All analyses were performed using R version 4.0.2<sup>46</sup> along with the `mgcv`<sup>39</sup> and `dlnm`<sup>47</sup> packages. Code and data for reproducing all the results, figures and animation in this paper is available at <https://github.com/sahirbhatnagar/covid19-mobility>.

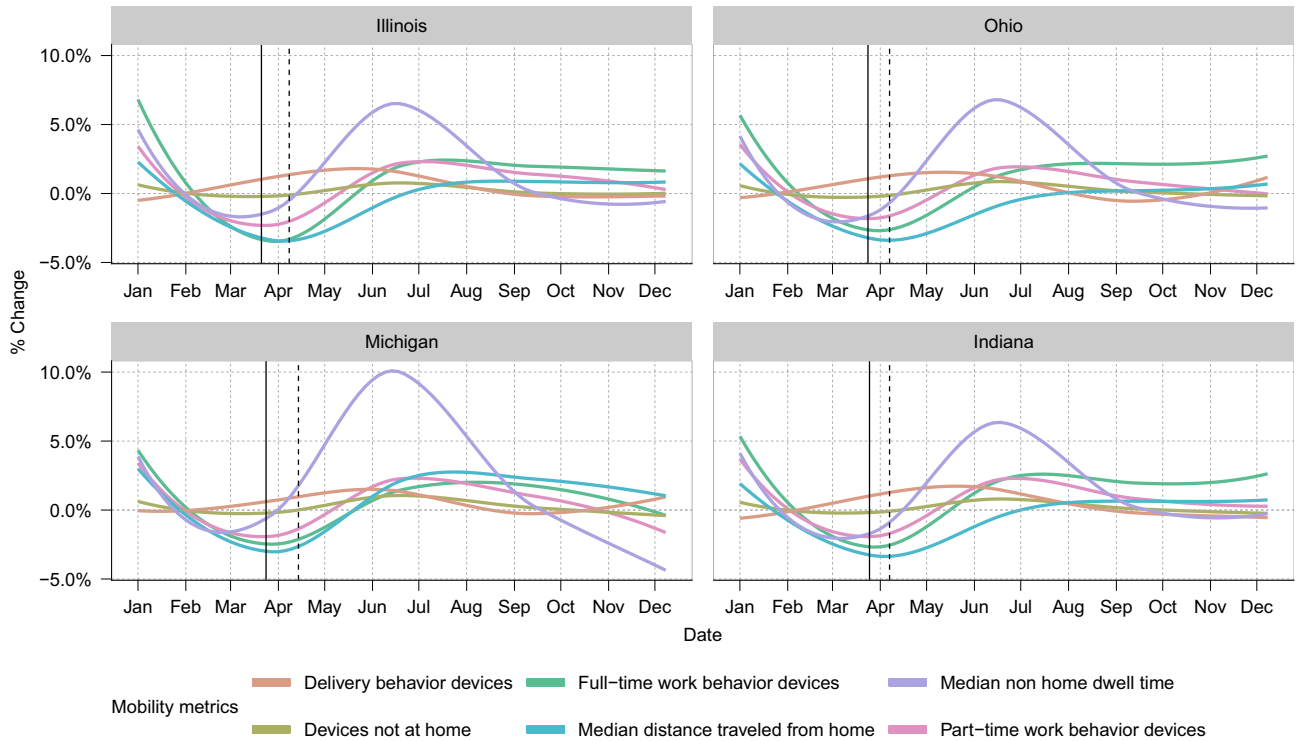
## Results

**Mobility patterns.** Daily mobility changes of three hundred sixty-five counties from the four most populous states in the Midwest: Illinois, Ohio, Michigan and Indiana were analyzed between January 1 2020 until December 8 2020. State-level sociodemographic and economic characteristics were similar across four states and are summarized in Table 1.

Figure 1 illustrates the average daily changes of the six mobility metrics between January and December 2020 of each state (average of all counties) relative to the week before. Overall, each metric had a unique trajectory but trends were similar across four states. Based on the average change, the number of devices not at home and delivery behavior (more than 3 stops lasting for less than 20 mins) remained stable throughout time. There was more variation in the metrics related to full-time work and the time spent away from home. Of the four states, mobility changes were more pronounced in Ohio. Across all states, relative to the previous seven days, mobility increased daily between March and May. Mobility metrics varied considerably by counties (S1, S2, S3, S4), illustrating how aggregating changes at the state-level may mask granular changes at the county-level.

State	Order	Lift	Population	Number of counties	Median household income	Cumulative cases per capita at opening <sup>a</sup>	Cumulative cases per capita until Dec 8 <sup>a</sup>	Party affiliation <sup>b</sup>
Illinois	March 21	April 8	12,741,080	102	\$65,030	118	6324	Democratic
Ohio	March 24	April 7	11,689,442	88	\$56,111	41	4363	Republican
Michigan	March 24	April 14	9,995,915	83	\$56,697	269	4427	Democratic
Indiana	March 25	April 7	6,691,878	92	\$55,746	83	5909	Republican

**Table 1.** The sociodemographic and economic characteristics of Illinois (IL), Ohio (OH), Michigan (M) and Indiana (IN). <sup>a</sup>Cumulative cases per 100,000 population <sup>b</sup>Based on the 2020 Presidential elections<sup>48</sup>.



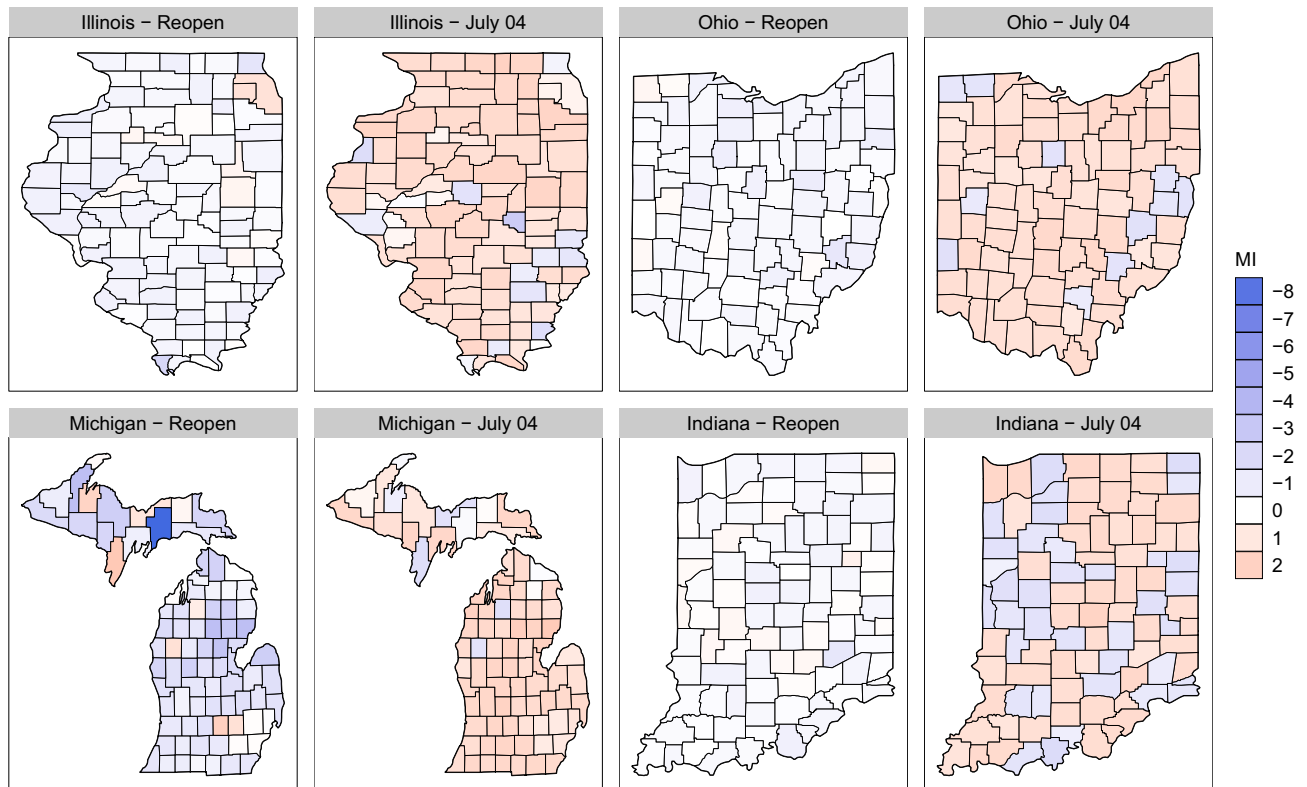
**Figure 1.** The average daily changes from baseline in the six mobility metrics for all counties of each state between January and December 2020. The baseline was calculated using a rolling average of the 7 previous days. The solid vertical lines represent the date the stay-at-home orders were put in place while the dotted vertical lines represent the dates the stay-at-home orders were lifted.

Illinois (n = 102)	Ohio (n = 88)	Michigan (n = 83)	Indiana (n = 92)
0.57 (0.50, 0.63)	0.71 (0.67, 0.74)	0.61 (0.53, 0.69)	0.65 (0.59, 0.70)

**Table 2.** Median (inter quartile range) of the proportion of variance explained by the first fPCA by state. *n* represents the number of counties in each state.

**First fPCA summarizes mobility patterns by counties.** We created a latent index of mobility by counties as given by Eq. (2) which is derived from the first fPCA. Table 2 provides the median and inter quartile range of the proportion of variance explained by the first fPCA across counties in a given state. We saw that more than 50% of the variance was explained by the first fPCA for a majority of all the counties analyzed, indicating good dimension reduction. In Supplemental Figures S5, S6, S7 and S8, we provide the absolute Pearson correlations between our MI and each individual metric by county for Illinois, Ohio, Michigan and Indiana, respectively. Furthermore, the correlations were particularly strong with full-/part-time work behavior as well as time spent away from home. Importantly, there was significant heterogeneity across counties, which would otherwise be missed when aggregating mobility metrics at the state level.

Figure 2 compares the changes of the MI from the day stay-at-home policies expired and July 4 (Independence Day). Blue shades indicate MI < 0 (decrease in mobility) and red shades indicate MI > 0 (increase in mobility). This graph provides some evidence that our MI is appropriately capturing mobility as we would expect there



**Figure 2.** MI values for each county of each state on the day the stay-at-home orders expired (reopen) and on July 4, 2020. Blue shades indicate a decrease in mobility ( $MI < 0$ ) and red shades indicate an increase in mobility ( $MI > 0$ ).

to be more movement on a traditionally busy U.S. holiday compared to earlier on in the pandemic when stay-at-home orders were lifted. In the Supplemental material, we also provide an animation illustrating the daily changes from reopening to December 8, 2020. The animation shows substantial difference in mobility patterns across counties that vary from day to day. The most dramatic change over time is increases in mobility from a weekday to a weekend.

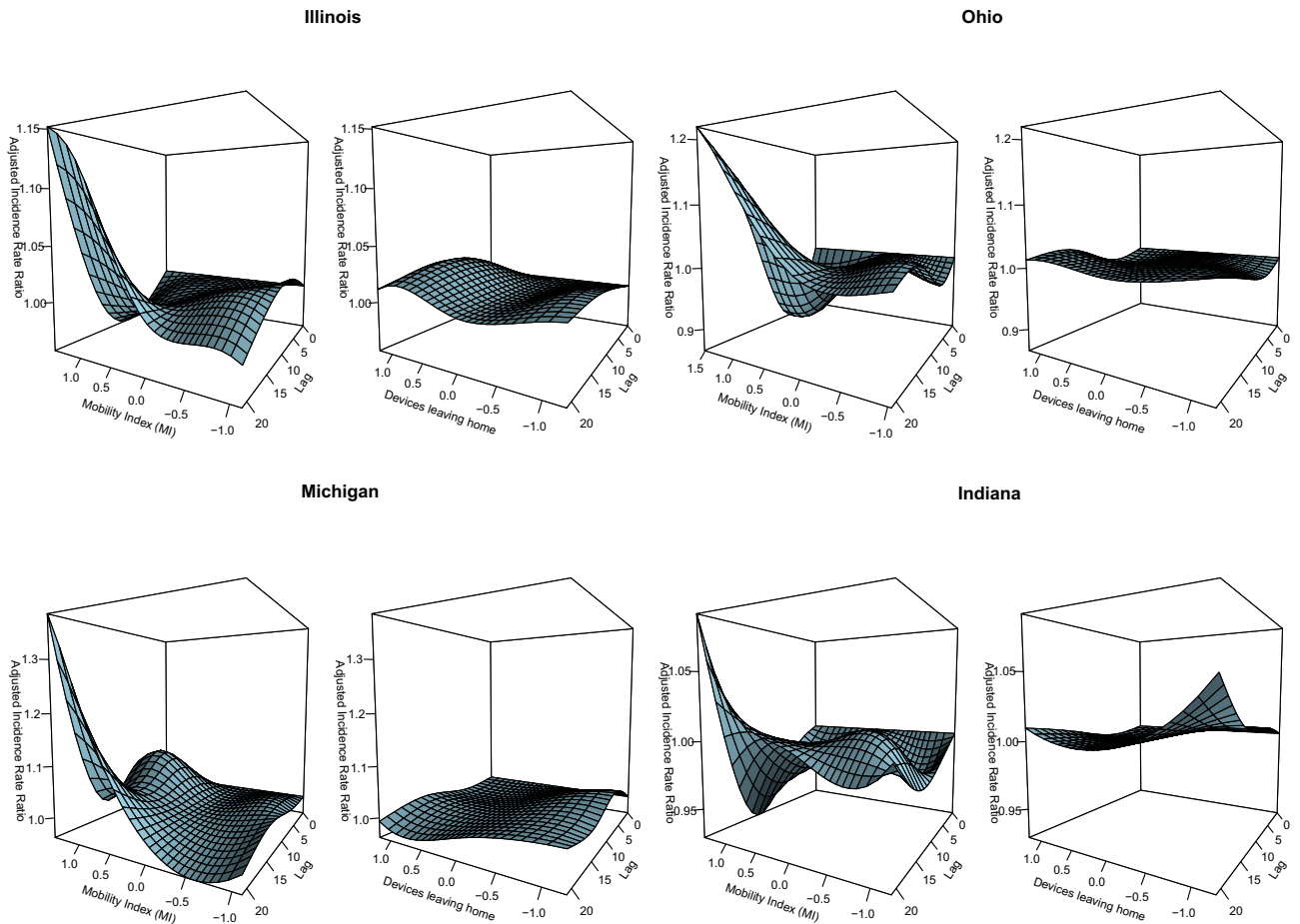
**Association with COVID-19.** To evaluate the validity of the MI, we compared its association with COVID-19 cases and a commonly used single metric of mobility (fraction of devices leaving home) (Fig. 3). Notably, the single metric was not associated with COVID-19 cases in any state at any lagged time point. However, there was a clear dose-response relationship between our MI and COVID-19 cases following a 10–21-day time lag in all four states. Across all four states the MI model resulted in significantly better goodness-of-fit statistics compared to the single metric (Table 3).

## Discussion

The COVID-19 pandemic is now fueled by highly transmissible variants of concern. Understanding the association between mobility and disease transmission can help tailor non-pharmaceutical interventions to mitigate outbreaks and potentially be used as an early indicator for surges in new infections. We leveraged freely available cell phone data with an unsupervised machine learning approach to create a multidimensional index of mobility. Results from our study suggest following the expiration of stay-at-home physical distancing policies, single metrics of mobility were not sensitive enough to capture the complexity of human mobility related to disease transmission. Our MI was correlated with COVID-19 cases from all counties in Illinois, Ohio, Michigan and Indiana. In comparison, the single metric of mobility (fraction of devices leaving home) was not associated with incident cases. Our results also demonstrate the importance of evaluating changes at a granular level as there was significant heterogeneity within states.

Tracking mobility has the potential of becoming a powerful tool to determine the impact of public health policies<sup>25</sup>. A growing subfield of COVID-19 research involves the analysis of mobility data and patterns. In the last 2 years, a variety of metrics and sources have been used to track mobility<sup>49,50</sup>. Initial studies evaluated how populations adhered to stay-at-home policies by tracking mobility trends<sup>41,51</sup>. Later it became evident that mobility may be useful as a public health surveillance tool as studies evaluated the correlation between mobility and COVID-19 diagnoses.<sup>3,52–59</sup> A study by Lasry et al. used Safegraph mobility data as a proxy for social distancing in the metropolitan areas of Seattle, San Francisco, New York City, and New Orleans and found an association between changes in mobility (% personal mobile devices leaving home) at the state-level and COVID-19 cases during the first COVID-19 wave<sup>9</sup>. In all four metropolitan areas, the number of mobile devices leaving home





**Figure 3.** Model results comparing the MI and its association with COVID-19 cases and a commonly used single metric of mobility (fraction of devices leaving home). For each state, the left panel summarizes the multidimensional MI; the right panel represents the percentage of devices leaving their home (x-axis); y-axis is the adjusted incidence rate ratio of COVID-19, at varying lagged response (0–21 days) (z-axis).

State	Residual deviance	Degrees of freedom	Reduction in deviance for the MI model (fPCA)
Illinois (single)	152,581		
Illinois (fPCA)	149,282	3.6	3,299
Ohio (single)	123,551		
Ohio (fPCA)	112,975	8.6	10,576
Michigan (single)	261,759		
Michigan (fPCA)	250,888	7.0	10,871
Indiana (single)	83,517		
Indiana (fPCA)	80,297	1.7	3,220

**Table 3.** Analysis of deviance table comparing the goodness of fit between the MI model (fPCA) and the fraction of devices leaving home (single) of the four states. Degrees of freedom shown is for the  $\chi^2$  test statistic.

declined from 80% (on February 26 2020) to 42% in New York City, 47% in San Francisco, 52% in Seattle, and 61% in New Orleans as stay-at-home policies were implemented. However, at this time NPI were more homogenous across counties and states. In contrast our study evaluated the association of mobility and COVID-19 cases following the expiration of stay-at-home orders reflecting mobility behavior that is more reflective of typical population-level movement. While we have used mobility data from Safegraph, several other mobility datasets have been studied with regards to COVID-19. Zachreson et al. used data from Facebook to determine the validity of aggregate human mobility data for COVID transmission patterns<sup>60</sup>. Aggregate mobility data is available through the Facebook Data For Good Program, which uses the mobile apps’ location service records of the users GPS locations. To be captured in this database, participants must be a member of Facebook and enable location services. To prevent users from being identified, Facebook removes users who do not meet a certain

threshold during the data aggregation period. This means that less densely populated areas are most likely to be underrepresented by Facebook mobility data. Google also provides another source of aggregate mobility using the time spent by users at several geolocations using Google Maps. While studies using Google data have found that increases in mobility lead to increased COVID-19 cases and death<sup>61</sup>, to date studies have not taken into account changes in mobility following the end of stay-at-home policies. Apple Mobility also provides mobility data using Apple Maps to create aggregated counts of direction requests. A study by James and Menzies used Apple Data and found mobility data and national financial indices exhibited similarities in their trajectories. Apple Mobility has several limitations as it includes any searches for directions as a measure of mobility and therefore may not be representative of actual community mobility.<sup>62</sup>

Mobility data offers several functions as a public health tool. While we focused directly on the number of COVID-19 cases, mobility can also be used to estimate and model of transmission rates<sup>52,60</sup>. Spatially explicit models of disease transmission using census data are often used to guide disease intervention decisions. However, it remains important to define mobility as a multidimensional construct. We demonstrated among hundreds of counties from four states, time-updated relative changes were associated with increases in COVID-19 cases. Furthermore, results from our study suggest our mobility should be considered an important confounder when evaluating the impact of other non-pharmaceutical interventions.

The strength of our study was the use of multiple advanced statistical methods to measure mobility and then validating its utility by evaluating its association with COVID-19 cases. The fPCA used to create the mobility index effectively captured the heterogeneity of the individual metrics over time and across counties within a given state. The unsupervised nature of this approach prevented the model from overfitting when evaluating the association with cases. Furthermore, we modelled a non-linear functional relationship between mobility and COVID-19 cases using a HGAM model while simultaneously fitting different lagged time periods. The expectation that the lag time should vary across states was confirmed by our results. The use of these methods has been under appreciated in the epidemiological and public health studies; we provide code and data to expand the use as we believe these methods could have wide applications in future research. We also highlight the need to track population-level mobility at a granular level, as we show significant heterogeneity across counties.

Our study also has limitations. The results of our study are based on data from all 365 counties from four states in the Midwest. While these counties represented varying population densities, socioeconomic conditions, and party affiliations (that may have resulted in different adherence and uptake of other NPI) our results may have limited generalizability to other larger metropolitan cities. Cell phone data was freely available and could help to predict trends during the pandemic but it is only a proxy for human contact. In this study we attempted to define a more robust definition of mobility, however it still remains a surrogate exposure. The association between mobility and COVID-19 cases may be underestimated, given our outcome is dependent on testing. Testing capacity has significantly changed throughout the pandemic in the United States. Seroprevalence studies estimate case detection is underrepresented by a factor of three times<sup>63</sup>. Although we do not believe this underrepresentation is differential between counties, outcomes such as COVID-19 related deaths and hospitalizations may be less biased. While the advantage is clear, the utility of these outcomes as a “real-time” public health tool is debatable as the latency period (time of infection to outcome) is long (greater than 21 days). As with all observational studies, associations should not be interpreted causally. Our model does not take into consideration confounding interventions that could also increase or mitigate transmission such as the proportion of the population adhering to physical distancing guidelines, wearing masks, interactions outside vs inside or air quality. To effectively measure social distancing patterns using individual-level data (either cell phones or wearable technology such as fitness trackers), would be more sensitive compared to aggregate data, but this raises ethical and privacy concerns<sup>64</sup>. Recent reports have hypothesized the COVID-19 pandemic may not be following a normal distribution but over dispersed or driven by “super spreader” transmission events which we did not account for in our model<sup>65</sup>. Finally, while PCA has the advantage of reducing overfitting, it has several assumptions and limitations. We must assume the features are related to each other in a linear fashion, and that the data can be appropriately summarised by the mean and variance<sup>66</sup>. Furthermore, PCA can be heavily influenced by outliers (three times the standard deviation from the sample mean), requires that the PCs are orthogonal to each other, and results in information loss due to selecting a relatively small number of PCs for downstream analysis. Specifically for our data, we show in Table 2 that selecting the first fPCA explains over 50% of the variance explained for a majority of all the counties analysed. The data did not have any significant outliers as seen in Supplemental Figure S13, which shows that the coefficients of variation (standard deviation divided by the mean) is less than 2.5 for all mobility metrics across all counties. We did not pursue nonlinear dimension reduction techniques such as kernel PCA, but think this would be an interesting direction for future research.

## Conclusion

Our study underscores the potential of using freely available cell phone data as public health tool. We show changes in mobility can be used a predictor of surges in COVID-19 cases. However, monitoring mobility in the absence of strict non-pharmaceutical interventions such as stay-at-home policies will require robust definitions.

Received: 15 September 2021; Accepted: 12 April 2022

Published online: 10 May 2022

## References

1. Tram, K. H. *et al.* Deliberation, dissent, and distrust: Understanding distinct drivers of Coronavirus disease 2019 vaccine hesitancy in the United States. *Clin. Infect. Dis.* (2021). In press.

2. Mayo Clinic. U.S. COVID-19 vaccine tracker: See your state's progress. Retrieved March 8, 2022 from <https://www.mayoclinic.org/coronavirus-covid-19/vaccine-tracker> (2021).
3. Zipursky, J. S. & Redelmeier, D. A. Mobility and mortality during the COVID-19 pandemic. *J. Gen. Intern. Med.* **35**, 3100–3101 (2020).
4. Finazzi, F. & Fassò, A. The impact of the COVID-19 pandemic on Italian mobility. *Significance (Oxford, England)* **17**, 17 (2020).
5. Jiang, J. & Luo, L. Influence of population mobility on the novel coronavirus disease (COVID-19) epidemic: Based on panel data from Hubei, China. *Glob. Health Res. Policy* **5**, 1–10 (2020).
6. Brown, K. A. *et al.* The mobility gap: Estimating mobility thresholds required to control SARS-CoV-2 in Canada. *CMAJ* **193**, E592–E600 (2021).
7. Bian, L. *et al.* Impact of the Delta variant on vaccine efficacy and response strategies. *Exp. Rev. Vacc.* **20**, 1201–1209 (2021).
8. León, T. M., Vargo, J., Pan, E. S., Jain, S. & Shete, P. B. Nonpharmaceutical interventions remain essential to reducing Coronavirus disease 2019 burden even in a well-vaccinated society: A modeling study. *Open Forum Infect. Dis.* **8**, ofab415 (2021).
9. Lasry, A. *et al.* Timing of community mitigation and changes in reported COVID-19 and community mobility—four US metropolitan areas, February 26–April 1, 2020. *Morb. Mortal. Wkly Rep.* **69**, 451–457 (2020).
10. Linka, K., Goriely, A. & Kuhl, E. Global and local mobility as a barometer for COVID-19 dynamics. *Biomech. Model. Mechanobiol.* **20**, 651–669 (2021).
11. Xiong, C. *et al.* Mobile device location data reveal human mobility response to state-level stay-at-home orders during the COVID-19 pandemic in the USA. *J. R. Soc. Interface* **17**, 20200344 (2020).
12. Khataee, H., Scheuring, I., Czirok, A. & Neufeld, Z. Effects of social distancing on the spreading of COVID-19 inferred from mobile phone data. *Sci. Rep.* **11**, 1–9 (2021).
13. Saha, J., Barman, B. & Chouhan, P. Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. *Child Youth Serv. Rev.* **116**, 105160 (2020).
14. Abu-Rayash, A. & Dincer, I. Analysis of mobility trends during the COVID-19 coronavirus pandemic: Exploring the impacts on global aviation and travel in selected cities. *Energy Re. Soc. Sci.* **68**, 101693 (2020).
15. Eckert, F. & Mikosch, H. Mobility and sales activity during the Corona crisis: Daily indicators for Switzerland. *Swiss J. Econ. Stat.* **156**, 9 (2020).
16. Noland, R. B. Mobility and the effective reproduction rate of COVID-19. *J. Transp. Health* **20**, 101016 (2021).
17. Kuo, C.-P. & Fu, J. S. Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions. *Sci. Total Environ.* **758**, 144151 (2021).
18. Nouvellet, P. *et al.* Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**, 1–9 (2021).
19. Zhou, Y. *et al.* Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: A modelling study using mobile phone data. *Lancet Digit. Health* **2**, e417–e424 (2020).
20. World Health Organization. Tracking SARS-CoV-2 variants. Retrieved March 11, 2022 from <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (2022).
21. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C. Pathophysiology, transmission, diagnosis, and treatment of Coronavirus disease 2019 (COVID-19): A review. *JAMA* **324**, 782–793 (2020).
22. Batista, C. *et al.* The silent and dangerous inequity around access to COVID-19 testing: A call to action. *EClinicalMedicine* **43**, 2 (2022).
23. Garg, S. *et al.* Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed Coronavirus disease 2019—COVID-NET, 14 States, March 1–30, 2020. *Morb. Mortal. Wkly Rep.* **69**, 458 (2020).
24. Van Dorn, A., Cooney, R. E. & Sabin, M. L. COVID-19 exacerbating inequalities in the US. *Lancet (London, England)* **395**, 1243 (2020).
25. Buckee, C. O. *et al.* Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
26. The New York Times. Coronavirus (COVID-19) Data in the United States. Retrieved December 10, 2021 from <https://github.com/nytimes/COVID-19-data> (2021).
27. Surgo Ventures. The U.S. covid community vulnerability index (CCVI). Retrieved August 17, 2021 from <https://precisionforcovid.org/ccvi> (2021).
28. Smittenaar, P. *et al.* A COVID-19 community vulnerability index to drive precision policy in the US. *medRxiv* doi:10.1101/2021.05.19.21257455 (2021). Preprint, <https://www.medrxiv.org/content/early/2021/05/20/2021.05.19.21257455.full.pdf>.
29. Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L. & Lewis, B. A social vulnerability index for disaster management. *J. Homeland Secur. Emerg. Manag.* **8**, 2 (2011).
30. Melvin, S. C., Wiggins, C., Burse, N., Thompson, E. & Monger, M. Peer reviewed: The role of public health in COVID-19 emergency response efforts from a rural health perspective. *Prevent. Chronic Dis.* **17**, 2 (2020).
31. Centers for Disease Control & Prevention. COVID-19 secondary data and statistics 2020. Retrieved March 8, 2022 from <https://www.cdc.gov/library/researchguides/2019novelcoronavirus/datastatistics.html> (2021).
32. Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N. & Wilkerson, J. Pandemic politics: Timing state-level social distancing responses to COVID-19. *J. Health Polit. Policy Law* **46**, 211–233 (2021).
33. Sulyok, M. & Walker, M. Community movement and COVID-19: A global study using Googles community mobility reports. *Epidemiol. Infect.* **148**, e284 (2020).
34. Gao, S., Rao, J., Kang, Y., Liang, Y. & Kruse, J. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSpatial Spec.* **12**, 16–26 (2020).
35. Post, L. A. *et al.* Surveillance metrics of SARS-CoV-2 transmission in central Asia: Longitudinal trend analysis. *J. Med. Internet Res.* **23**, e25799 (2021).
36. He, Y., Wang, X., He, H., Zhai, J. & Wang, B. Moving average based index for judging the peak of the COVID-19 epidemic. *Int. J. Environ. Res. Public Health* **17**, 5288 (2020).
37. Kim, E. H. & Bae, J.-M. Seasonality of tuberculosis in the Republic of Korea, 2006–2016. *Epidemiol. Health* **40**, e2018051 (2018).
38. Wang, J.-L., Chiou, J.-M. & Müller, H.-G. Functional data analysis. *Ann. Rev. Stat. Appl.* **3**, 257–295 (2016).
39. Wood, S. N. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science (CRC Press/Taylor & Francis Group, Boca Raton, 2017), second edition edn.
40. Pedersen, E. J., Miller, D. L., Simpson, G. L. & Ross, N. Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ* **7**, e6876 (2019).
41. Gasparrini, A., Armstrong, B. & Kenward, M. G. Distributed lag non-linear models. *Stat. Med.* **29**, 2224–2234 (2010).
42. Gu, C. & Gu, C. *Smoothing Spline ANOVA Models* Vol. 297 (Springer, 2013).
43. Wood, S. N. Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**, 95–114 (2003).
44. Gasparrini, A., Scheipl, F., Armstrong, B. & Kenward, M. G. A penalized framework for distributed lag non-linear models. *Biometrics* **73**, 938–948 (2017).
45. Dhoubib, W. *et al.* The incubation period during the pandemic of COVID-19: A systematic review and meta-analysis. *Syst. Rev.* **10**, 1–14 (2021).
46. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2020).



47. Gasparrini, A. Distributed lag linear and non-linear models in R: The package dlnm. *J. Stat. Softw.* **43**, 1–20 (2011).
48. The New York Times. Presidential Election Results: Biden Wins. Retrieved March 11, 2022 from <https://www.nytimes.com/interactive/2020/11/03/us/elections/results-president.html> (2021).
49. Carteni, A., Di Francesco, L. & Martino, M. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Sci. Total Environ.* **741**, 140489 (2020).
50. Kraemer, M. U. G. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
51. Wu, Y., Mooring, T. A. & Linz, M. Policy and weather influences on mobility during the early US COVID-19 pandemic. *Proc. Natl. Acad. Sci.* **118**, e2018185118 (2021).
52. Fox, S. J. *et al.* Real-time pandemic surveillance using hospital admissions and mobility data. *Proc. Natl. Acad. Sci.* **119**, e2111870119 (2022).
53. Han, X. *et al.* Quantifying COVID-19 importation risk in a dynamic network of domestic cities and international countries. *Proc. Natl. Acad. Sci.* **118**, e2100201118 (2021).
54. Persson, J., Parie, J. F. & Feuerriegel, S. Monitoring the COVID-19 epidemic with nationwide telecommunication data. *Proc. Natl. Acad. Sci.* **118**, e2100664118 (2021).
55. Hou, X. *et al.* Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proc. Natl. Acad. Sci.* **118**, e2020524118 (2021).
56. Singh, S., Shaikh, M., Hauck, K. & Miraldo, M. Impacts of introducing and lifting nonpharmaceutical interventions on COVID-19 daily growth rate and compliance in the United States. *Proc. Natl. Acad. Sci.* **118**, e2021359118 (2021).
57. Li, R. Mobility restrictions are more than transient reduction of travel activities. *Proc. Natl. Acad. Sci.* **118**, e2023895118 (2021).
58. Schlosser, F. *et al.* COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc. Natl. Acad. Sci.* **117**, 32883–32890 (2020).
59. Oztig, L. I. & Askin, O. E. Human mobility and Coronavirus disease 2019 (COVID-19): A negative binomial regression analysis. *Public Health* **185**, 364–367 (2020).
60. Zachreson, C. *et al.* Risk mapping for COVID-19 outbreaks in Australia using mobility data. *J. R. Soc. Interface* **18**, 20200657 (2021).
61. Yilmazkuday, H. Stay-at-home works to fight against COVID-19: International evidence from Google mobility data. *J. Hum. Behav. Soc. Environ.* **31**, 210–220 (2021).
62. James, N. & Menzies, M. Efficiency of communities and financial markets during the 2020 pandemic. *CHAOS Interdiscip. J. Non-linear Sci.* **31**, 083116 (2021).
63. Stone, M. *et al.* Use of US blood donors for national serosurveillance of severe acute respiratory syndrome Coronavirus 2 antibodies: Basis for an expanded national donor serosurveillance program. *Clin. Infect. Dis.* **74**, 871–881 (2022).
64. Anaya, L. S., Alsadoon, A., Costadopoulos, N. & Prasad, P. Ethical implications of user perceptions of wearable devices. *Sci. Eng. Ethics* **24**, 1–28 (2018).
65. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
66. Jolliffe, I. *Principal Component Analysis* (Springer Verlag, 2002).

## Author contributions

All the authors have seen and approved the final manuscript and have participated sufficiently in the work to take public responsibility for its content. S.S. and P.H.H. contributed equally to this project. S.S., K.H.T. and S.R.B. designed the study, interpreted the results, and drafted the manuscript. P.H.H. performed the literature review, extracted, and analyzed the data, produced the figures and revised and approved the manuscript under the supervision of S.R.B.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-10941-2>.

**Correspondence** and requests for materials should be addressed to S.R.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022