

Mobility data used to respond to COVID-19 can leave out older and non-white people

March 17 2021



Credit: Pixabay/CC0 Public Domain

Information on individuals' mobility—where they go as measured by their smartphones—has been used widely in devising and evaluating

ways to respond to COVID-19, including how to target public health resources. Yet little attention has been paid to how reliable these data are and what sorts of demographic bias they possess. A new study tested the reliability and bias of widely used mobility data, finding that older and non-White voters are less likely to be captured by these data. Allocating public health resources based on such information could cause disproportionate harms to high-risk elderly and minority groups.

The study, by researchers at Carnegie Mellon University (CMU) and Stanford University, appears in the *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, a publication of the Association for Computing Machinery.

"Older age is a major risk factor for COVID-19-related mortality, and African-American, Native-American, and Latinx communities bear a disproportionately high burden of COVID-19 cases and deaths," explains Amanda Coston, a doctoral student at CMU's Heinz College and Machine Learning Department, who led the study as a summer research fellow at Stanford University's Regulation, Evaluation, and Governance Lab. "If these demographic groups are not well represented in [data](#) that are used to inform policymaking, we risk enacting policies that fail to help those at greatest risk and further exacerbating serious disparities in the health care response to the pandemic."

During the COVID-19 pandemic, mobility data have been used to analyze the effectiveness of social distancing policies, illustrate how people's travel affects transmission of the virus, and probe how different sectors of the economy have been affected by social distancing. Yet despite the high-stakes settings in which this [information](#) has been used, independent assessments of the data's reliability are lacking.

In this study, the first independent audit of demographic [bias](#) of a smartphone-based mobility dataset used in the response to COVID-19,

researchers assessed the validity of SafeGraph data. This widely used mobility dataset contains information from approximately 47 million mobile devices in the United States. The data come from mobile applications, such as navigation, weather, and social media apps, where users have opted in to location tracking.

When COVID-19 began, SafeGraph released much of its data for free as part of the COVID-19 Data Consortium to enable researchers, nonprofits, and governments to gain insight and inform responses. As a result, SafeGraph's mobility data have been used widely in pandemic research, including by the Centers for Disease Control and Prevention, and to inform public health orders and guidelines issued by governors' offices, large cities, and counties. Researchers in this study sought to determine whether SafeGraph data accurately represent the broader population.

SafeGraph has reported publicly on the representativeness of its data. But the researchers suggest that because the company's analysis examined demographic bias only at Census-aggregated levels and did not address the question of demographic bias for inferences specific to places of interest (e.g. voting places), an independent audit was necessary.

A major challenge in conducting such an audit is the lack of demographic information—SafeGraph data do not contain demographics such as age and race. In this study, researchers showed how administrative data can provide the demographic information necessary for a bias audit, supplementing the information gathered by SafeGraph. They used North Carolina voter registration and turnout records, which typically include information on age, gender, and race, as well as voters' travel to a polling location on Election Day. Their data came from a private voter file vendor that combines publicly available voter records. In all, the study included 539,000 voters from North

Carolina who voted at 558 locations during the 2018 general election. The researchers deemed this sample highly representative of all voters in that state.

The study identified a sampling bias in the SafeGraph data that underrepresents two high-risk groups, which the authors called particularly concerning in the context of the COVID-19 pandemic. Specifically, older and minority voters were less likely to be captured by the mobility data. This could lead jurisdictions to under-allocate important health resources, such as pop-up testing sites and masks, to vulnerable populations.

"While SafeGraph information may help people make policy decisions, auxiliary information, including prior knowledge about local populations, should also be used to make policy decisions about allocating resources," suggests Alexandra Chouldechova, assistant professor of statistics and public policy at CMU, who coauthored the study.

The authors also call for more work to determine how [mobility data](#) can be more representative, including asking firms that provide this kind of data to be more transparent in including the sources of their data (e.g., identifying which smartphone applications were used to access the information).

Among the study's limitations, the authors note that in the United States, voters tend to be older and include more White people than the general population, so the study's results may underestimate the sampling bias in the general population. Additionally, since SafeGraph provides researchers with an aggregated version of the data for privacy reasons, researchers could not test for bias at the individual voter level. Instead, the authors tested for bias at physical places of interest, finding evidence that SafeGraph is more likely to capture traffic to places frequented by

younger, largely White visitors than to places frequented by older, largely non-White visitors.

More generally, the study shows how administrative data can be used to overcome the lack of demographic information, which is a common hurdle in conducting bias audits.

More information: Amanda Coston et al, Leveraging Administrative Data for Bias Audits, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021). [DOI: 10.1145/3442188.3445881](https://doi.org/10.1145/3442188.3445881)

Provided by Carnegie Mellon University

Citation: Mobility data used to respond to COVID-19 can leave out older and non-white people (2021, March 17) retrieved 22 May 2024 from <https://medicalxpress.com/news/2021-03-mobility-covid-older-non-white-people.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
