

Empirical Research with Large Datasets

Workshop

BPLIM



BANCO DE
PORTUGAL
EUROSISTEM



Large surveys and other continuous data streams in statistics production
Frauke Kreuter – LMU Munich, University of Maryland

BoP 17.12.22

PORTO | 19 - 20 DEC. 2022



Kathleen Stewart
UMD - Center for
Geospatial Information
Science



Samantha Chiu
UMD - Joint Program in
Survey Methodology



Xiaoyi Deng
UMD - Joint Program in
Survey Methodology



Yao Li
UMD - Center for
Geospatial Information
Science



Zheng Liu
UMD - Center for
Geospatial Information
Science

**Jesse Klein
Wichada La Motte-Kerr
et al.**
UMD - Social Data
Science Centre

**Anna-Carolina Haensch
Jacob Beck
et al.**
LMU Munich



Stanley Presser
UMD - Joint Program in
Survey Methodology
and Sociology



Elizabeth Stuart
JHU - Bloomberg School
of Public Health



Elena Goicoechea
JHU - Bloomberg School
of Public Health



Ting-Hsuan Chang
JHU - Bloomberg School
of Public Health



Alex Reinhart
Delphi Group
Carnegie Mellon
University

**Esther Kim
Sarah LaRocca
Kathleen Morris
Adrienne Bradford
Curtiss Cobb
and many more**
Facebook / Meta



Florian Keusch
University of
Mannheim, Germany



Sebastian Bähr
Institute for Employment
Research, Germany



Mark Trappmann
Institute for Employment
Research; University of
Bamberg, Germany



Georg-Christoph Haas
Institute for Employment
Research, University of
Mannheim, Germany



Sonja Malich
Institute for Employment
Research,
LMU Munich, Germany

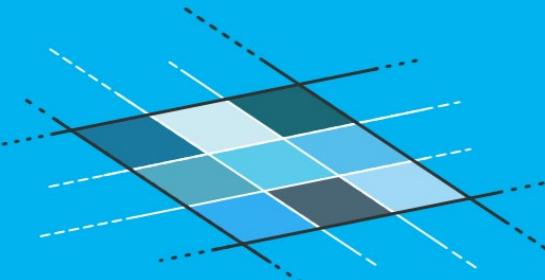
**Ralf Klueber
Michael Wennesheimer
Christoph Boegershausen
Christopher Kraemer
at umlaut / accenture**

FEDERAL STATISTICS, MULTIPLE DATA SOURCES, AND PRIVACY PROTECTION

Next Steps

MODERNIZING U.S. DATA INFRASTRUCTURE:

**Design Considerations for
Implementing a National Secure
Data Service to Improve Statistics
and Evidence Building**



DATA FOUNDATION



Structure and
Data Collection

Instruments and
Weights

Validation and
Access



Structure and
Data Collection

Participation
and Privacy

Validation



Prepare
Pipelines

Create
Communication

Train the Team



Global COVID-19 Symptom Survey. A Facebook partnership

Structure and Data Collection

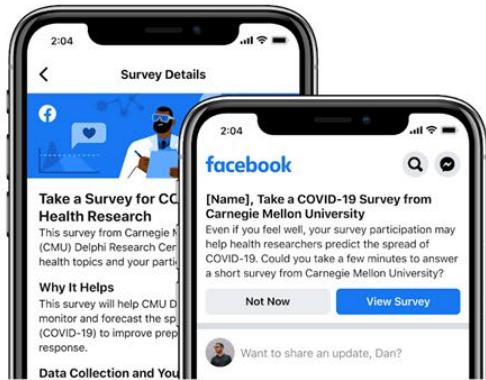
Responding to the Need for Syndromic Surveillance

Syndromic surveillance enables policymakers and public health systems to make decisions before diagnosis data are available, especially in low resource areas with limited testing capabilities.

Facebook can reach large segments of the target population daily with the technical infrastructure to provide bias correction. And, the speed and scale of the symptom surveys allow them to act as early warning systems.

Project Overview

1 Who's Taking the Survey



Facebook invites a new, random sample of users to participate each day.

2 How the Survey Works



Users are sent to the survey hosted by UMD or CMU using Qualtrics.

Facebook does not receive responses, but does calculate weights to correct for non-response bias and sampling frame coverage bias using internal Facebook data for 115 countries or territories.

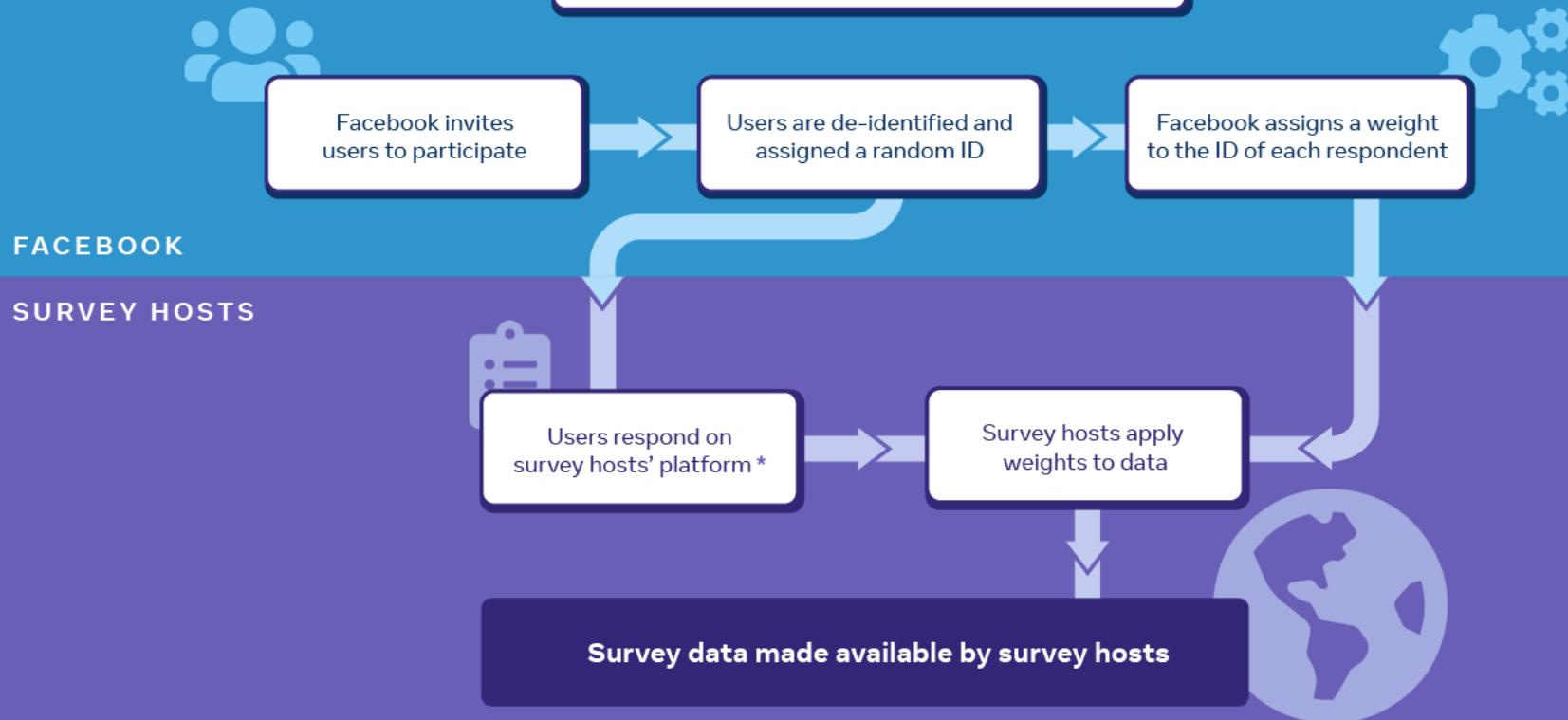
3 Using the Survey Data



Using the aggregated data, Facebook created a map visualization to help policymakers and public health systems make decisions.

The non-aggregate data are available to eligible academic and nonprofit researchers by request.

CTIS Data Collection Process



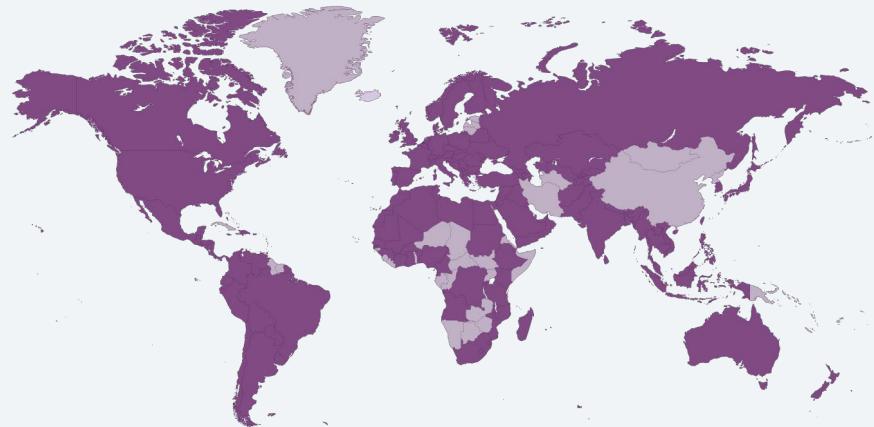
100 M +

Completed the UMD global survey launched in 200+ countries or territories, including 114 for which we provided weights.

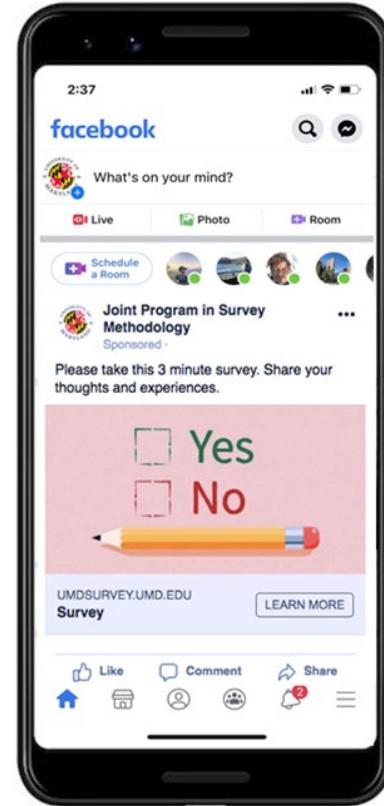
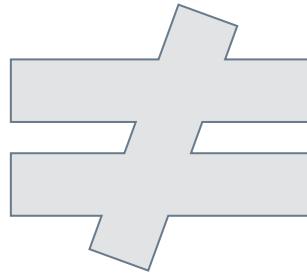
~50000

Completed the CMU survey launched in the United States **per day**

In UK: about 1800 respondents daily between **April 2020 and April 2022**



Project Overview



Survey through advertisement on FB

- Controlled Sampling
- Stable Sample Selection
- Known Selection Probabilities
- Known Population
- Transparent Methodology

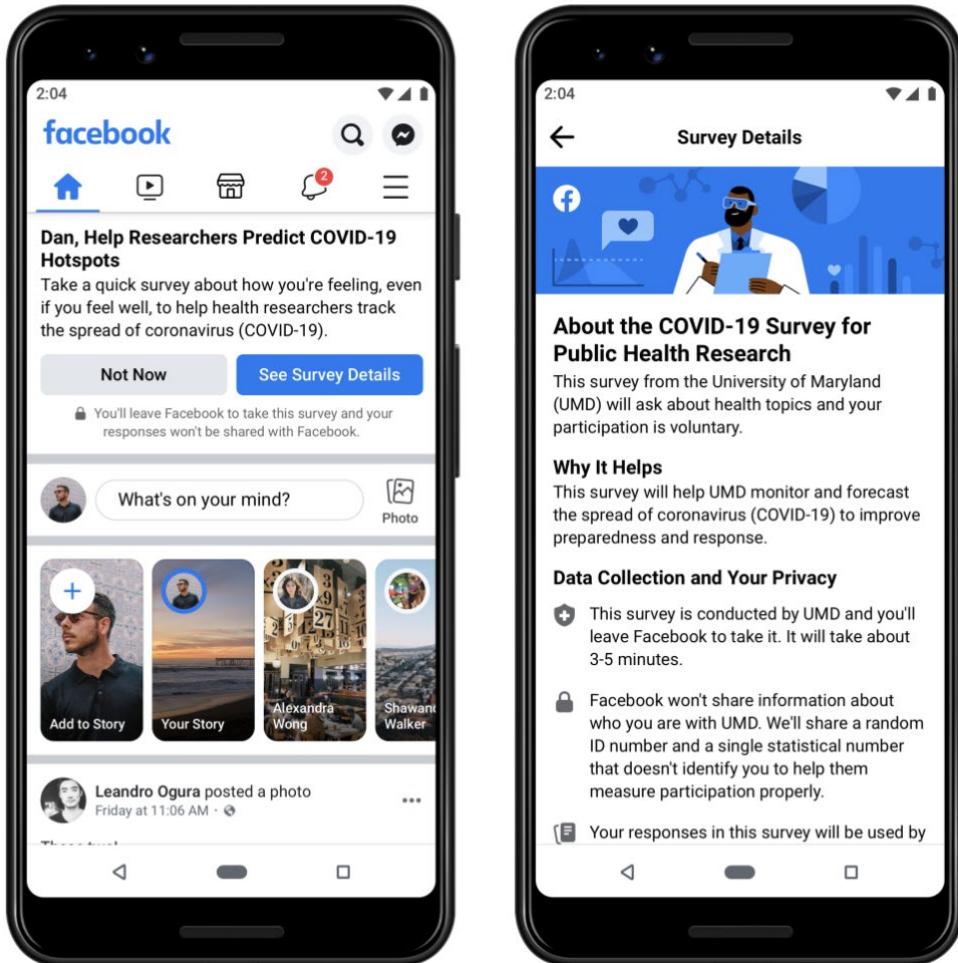
Survey Instrument and Weights

UMD Global Survey Instrument

Available in 50+ languages

Survey Instrument has 5 Sections
(13 versions over 2 years):

- Consent
- Health symptoms
- Contacts with others
- Mental health
- Economic security
- Demographic characteristics



April 6, 2020

April 15, 2020

May 21, 2020

Sept. 8, 2020

Nov. 24, 2020

Dec. 19, 2020

Jan 12., 2021

Feb. 8, 2021

March 2, 2021

May 21, 2021

W1	W2	W3	W4	W5	W6	W7	W8	W10	W11
	Community CLI	Translated Instrument Launches	Testing and Preventive Behaviors Additional Demographics	Schooling and School Policy	Vaccination Status and Intent	Revisions to the Vaccination Items	Reasons for Vaccine Hesitancy	Revisions to the Vaccination Items	Harmonization with the UMD Global Instrument and the Preventive Health Survey Barriers to Vaccine Access

Figure 12. This figure shows the estimated proportion of the adult (18+) population that is open to receiving a COVID-19 vaccine based on Facebook survey responses

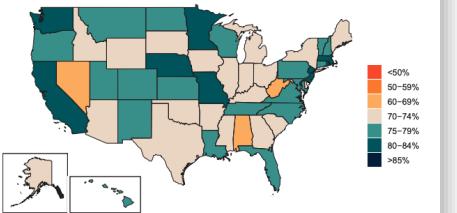


Figure 13. The number of people who receive any vaccine and those that are immune accounting for efficacy, loss to follow up for 2 dose vaccines, and a 28 day delay between first dose and immunity for 2 dose vaccines.

April 6, 2020

April 15, 2020

May 21, 2020

Sept. 8, 2020

Nov. 24, 2020

Dec. 19, 2020

Jan 12., 2021

Feb. 8, 2021

March 2, 2021

May 21, 2021

W1	W2	W3	W4	W5	W6	W7	W8	W10	W11
	Community CLI	Translated Instrument Launches	Testing and Preventive Behaviors Additional Demographics	Schooling and School Policy	Vaccination Status and Intent	Revisions to the Vaccination Items	Reasons for Vaccine Hesitancy	Revisions to the Vaccination Items	Harmonization with the UMD Global Instrument and the Preventive Health Survey Barriers to Vaccine Access

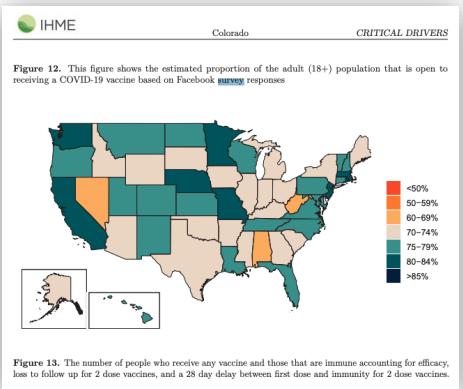
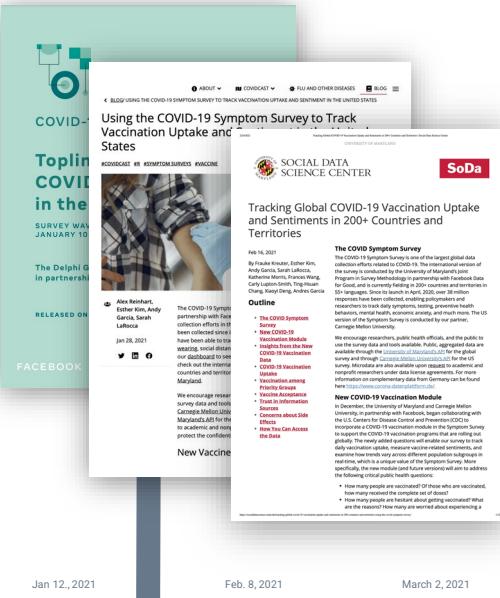


Figure 13. The number of people who receive any vaccine and those that are immune accounting for efficacy, loss to follow up for 2 dose vaccines, and a 28 day delay between first dose and immunity for 2 dose vaccines.



W1	W2	W3	W4	W5	W6	W7	W8	W10	W11
Community CLI	Translated Instrument Launches	Testing and Preventive Behaviors Additional Demographics	Schooling and School Policy	Vaccination Status and Intent	Revisions to the Vaccination Items	Reasons for Vaccine Hesitancy	Revisions to the Vaccination Items	Harmonization with the UMD Global Instrument and the Preventive Health Survey	Barriers to Vaccine Access

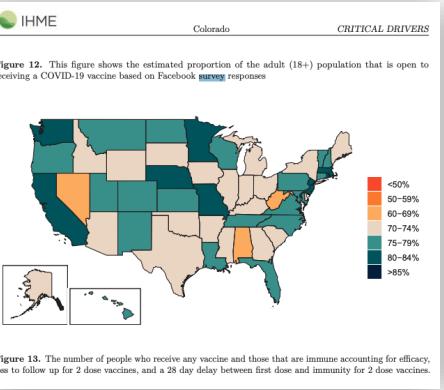


Figure 13. The number of people who receive any vaccine and those that are immune accounting for efficacy, loss to follow up for 2 dose vaccines, and a 28 day delay between first dose and immunity for 2 dose vaccines.

COVID-19 Symptom Survey

Using the COVID-19 Symptom Survey to track vaccination uptake and states

RELEASED ON JANUARY 10, 2021

The Delphi Group in partnership with IHME

FACEBOOK

Acknowledgments

Feb 14, 2021 By Frédéric Mentré, Esther Kim, Andy Garcia, Sarah Margolis, Katherine Morris, Rita Chang, Karen Deng & others

Outline

- The COVID-19 Symptom Survey is a partnership with Fieldwork Health and IHME. Data have been collected since Jan 10, 2021. We are currently updating, social distant and self-reporting data. Please check out the entire countries and territories Margolis et al.
- We encourage reader survey data and tools. Contact us at ihme@uw.edu or [MargolisLab.org](https://www.margolislab.org/) for more academic and non-profit collaboration opportunities.
- How You Can Access the Data

New Vaccine

ANNOUNCEMENT / PRESS RELEASE

Surgo Ventures and Resolve to Save Lives release new blueprint for states, counties, and cities to improve COVID-19 vaccine equity

APHA and Duke-Margolis Center for Health Policy endorse blueprint as "critical" and "practical" tool for increasing coverage at the local level

WASHINGTON, June 17, 2020—Today, Surgo Ventures and Resolve to Save Lives, an initiative of Vital Strategies, released a four-step, data-driven blueprint for public health officials and local leaders to help them increase vaccination coverage in their communities, and do it equitably. It has been endorsed by the American Public Health Association (APHA) and the Duke-Margolis Center for Health Policy

View Press Release



Incorporate a COVID-19 vaccination module in the Symptom Survey to track vaccination coverage and trends across the U.S. and globally. The newly added questions will enable our survey to track vaccination coverage and trends over time. This will allow us to examine how trends vary across different population subgroups in the U.S. and globally. Specifically, the new module (and future versions) will aim to address the following questions:

- How many people are vaccinated? Of those who are vaccinated, how many are fully vaccinated?
- How many people are hesitant about getting vaccinated? And are they? How are we worried about experiencing a side effect?

April 6, 2020 April 15, 2020 May 21, 2020 Sept. 8, 2020 Nov. 24, 2020 Dec. 19, 2020 Jan. 12, 2021 Feb. 8, 2021 March 2, 2021 May 21, 2021

W1	W2	W3	W4	W5	W6	W7	W8	W10	W11
	Community CLI	Translated Instrument Launches	Testing and Preventive Behaviors Additional Demographics	Schooling and School Policy	Vaccination Status and Intent	Revisions to the Vaccination Items	Reasons for Vaccine Hesitancy	Revisions to the Vaccination Items	Harmonization with the UMD Global Instrument and the Preventive Health Survey Barriers to Vaccine Access



Colorado

Figure 12. This figure shows the estimated proportion of people receiving a COVID-19 vaccine based on Facebook survey responses.

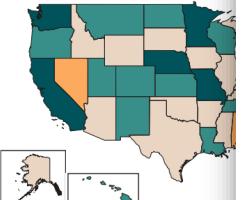


Figure 13. The number of people who receive any vaccine are at risk to loss follow up for 2 dose vaccines, and a 28 day delay between doses.

REMARKS

COVID-19 Household COVID-19 risk and in-person schooling

Author(s): Michael L. Krieger, Daniel J. Gitterman, and Daniel S. Goldstein
Source: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8070000/>

In-person schooling has proved contentious and difficult to study throughout the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Data from a national white survey of U.S. households in January 2021 provide a unique opportunity to examine the relationship between household transmission risk, particularly daily symptom scores, teacher masking, and dosage of in-school vaccination. We find that the risk of transmission is higher among households with children attending school in person, but when one or more mitigation measures are reported, a significant reduction in transmission risk is observed. For example, when all three mitigation measures (i.e., mask, distancing, and vaccination) are in place, transmission risk is reduced by nearly 50% compared with households that do not practice any mitigation measures. Even if transmission is eliminated, however, transmission risk remains high, particularly during the winter months, when mask use and drop-off.

WHY IT'S SO IMPORTANT THAT TWELVE-YEAR-OLDS CAN NOW GET A COVID-19 VACCINE

We are in a pandemic from which, as much as we might wish it, children have never been exempt.

By Amy Davidson Sorkin
www.nytimes.com/2021/06/17/health/covid-vaccine-kids.html



Leaving the age for the vaccine can open a range of questions. But it's an important moment for public health and for children themselves. —PHOTOGRAPH BY ROBERT GOLDBECK FOR THE NEW YORK TIMES

Fig. 14. Spatial distribution of survey responses. (A) Number of survey responses per county. (B) Percentage of households with children aged 12–17 years old who received their first dose of COVID-19 vaccine. (C) Percentage of households with children aged 12–17 years old who report having received their second dose of COVID-19 vaccine. (D) Average number of children aged 12–17 years old per household. (E) Percentage of households with children aged 12–17 years old reporting in-person schooling.

Linton et al., *Nature* 590, 400–403 (2021); doi:10.1038/s41586-021-03386-w

April 6, 2020

April 15, 2020

May 21, 2020

Sept. 8, 2020

Nov. 24, 2020

Dec. 19, 2020

Jan. 12, 2021

Feb. 8, 2021

March 2, 2021

May 21, 2021

W1

Community CLI

Translated Instrument Launches

Testing and Preventive Behaviors
Additional Demographics

W4

Schooling and School Policy

Vaccination Status and Intent

W6

Revisions to the Vaccination Items

W7

Reasons for Vaccine Hesitancy

Revisions to the Vaccination Items

Harmonization with the UMD Global Instrument and the Preventive Health Survey
Barriers to Vaccine Access

ANNOUNCEMENT / PRESS RELEASE
Surgo Ventures and Resolve to Save Lives release new blueprint for states, counties, and cities to improve COVID-19 vaccine equity

APHA and Duke-Margolis Center for Health Policy endorse blueprint as "critical" and "practical" tool for increasing coverage at the local level

WASHINGTON, June 17, 2020—Todays Surgo Ventures and Resolve to Save Lives, an initiative of Vital Strategies, released a four-step, data-driven blueprint for public health officials and local leaders to help them increase vaccination coverage in their communities, and do it equitably. It has been endorsed by the American Public Health Association (APHA) and the Duke-Margolis Center for Health Policy



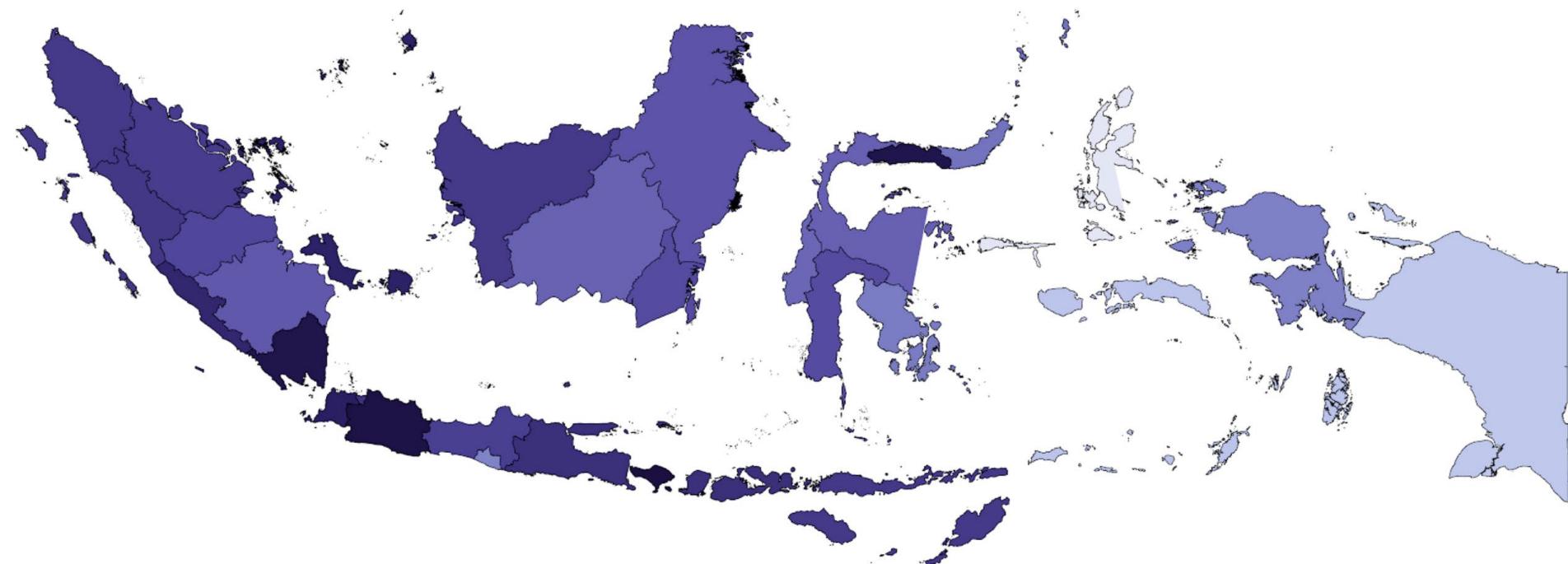
- The COVID-19 Symptom Survey partnership with Field Operations and the CDC has been collected since January 10. We encourage you to start using the survey to track vaccination rates across your country and territory.
- We encourage regular survey data and tools. Continue to use the survey to track vaccination rates across your country and territory.
- Check out the new COVID-19 Vaccination Equity Dashboard.
- How You Can Access the Data

Incorporate a COVID-19 vaccination module in the Symptom Survey to track vaccination rates across your country and territory. The newly added questions will enable our survey to track vaccination rates across your country and territory and examine how trends vary across different population subgroups in your community. The survey will also allow you to specifically track vaccination rates across different racial and ethnic groups.

- How many people are vaccinated? Of those who are vaccinated, how many are Black, Hispanic, Asian, and other?
- How many people are hesitant about getting vaccinated? And are they the same race and ethnicity as those who are vaccinated?

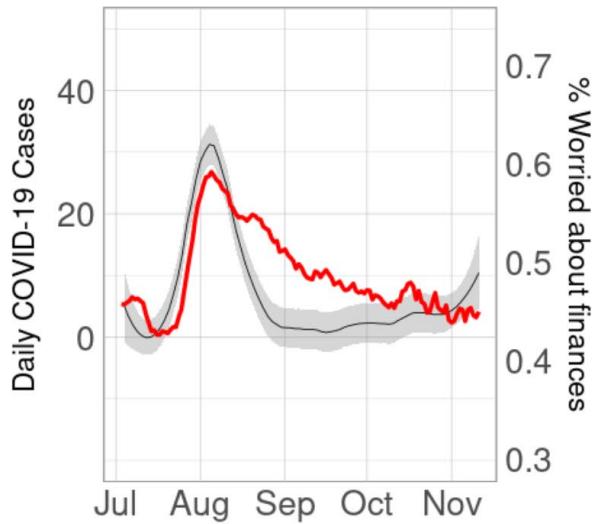
Indonesia Household Finances, November 2020

Survey Question: How worried are you about your household's finances in the next month?

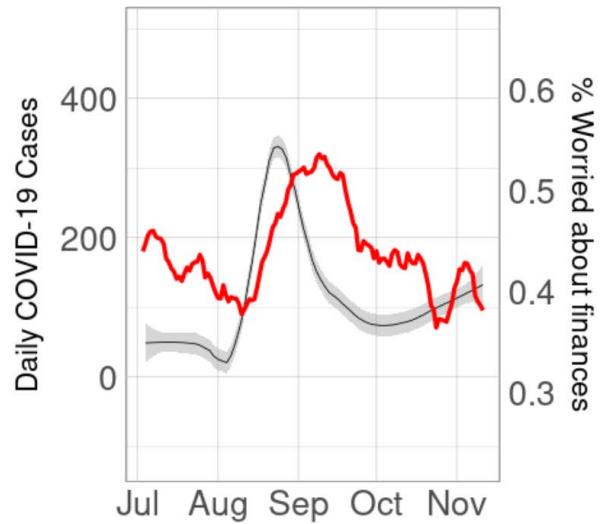


Financial anxiety and COVID-19 cases

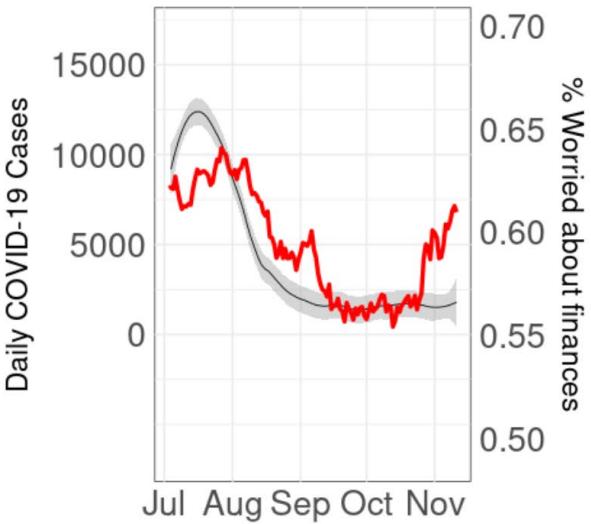
Vietnam



South Korea



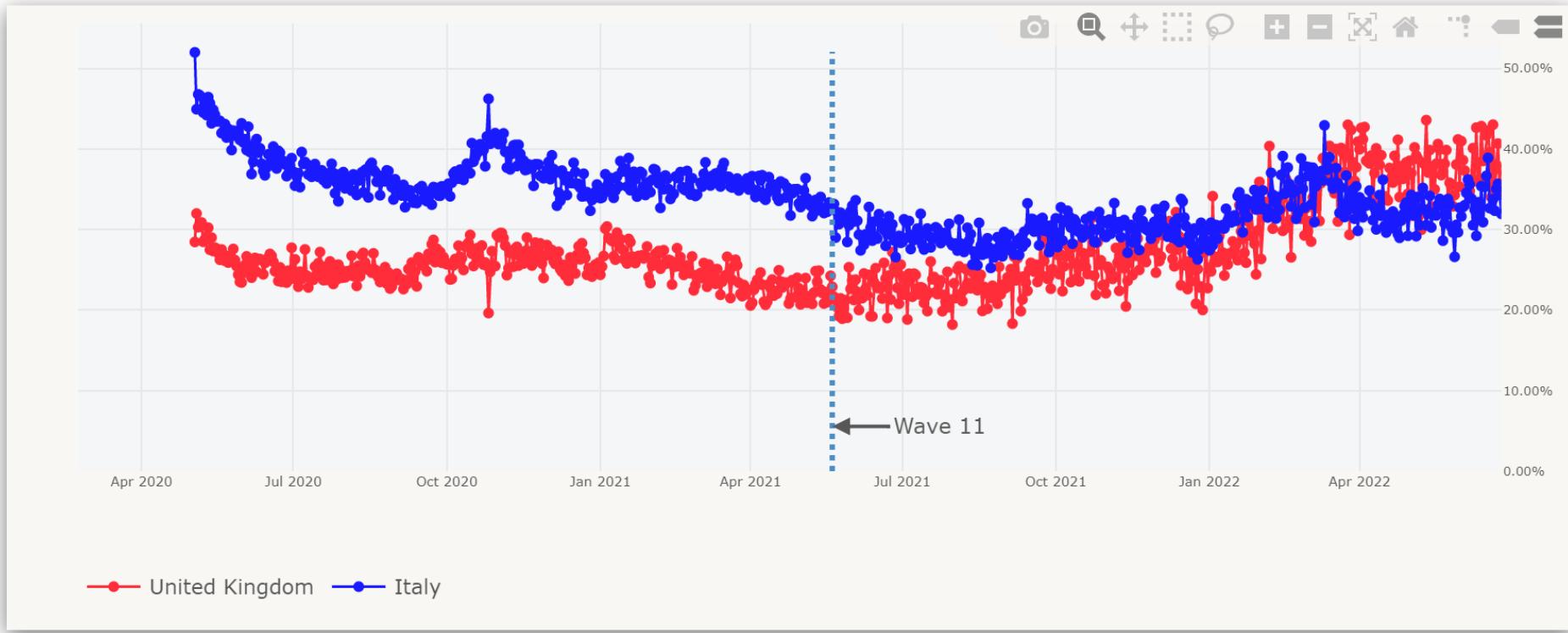
South Africa



Red line represents weighted proportion of individuals who responded 'very worried' or 'somewhat worried' to the question 'How worried are you about your household's finances in the next month?' Black line represents smoothed daily COVID-19 cases available at COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

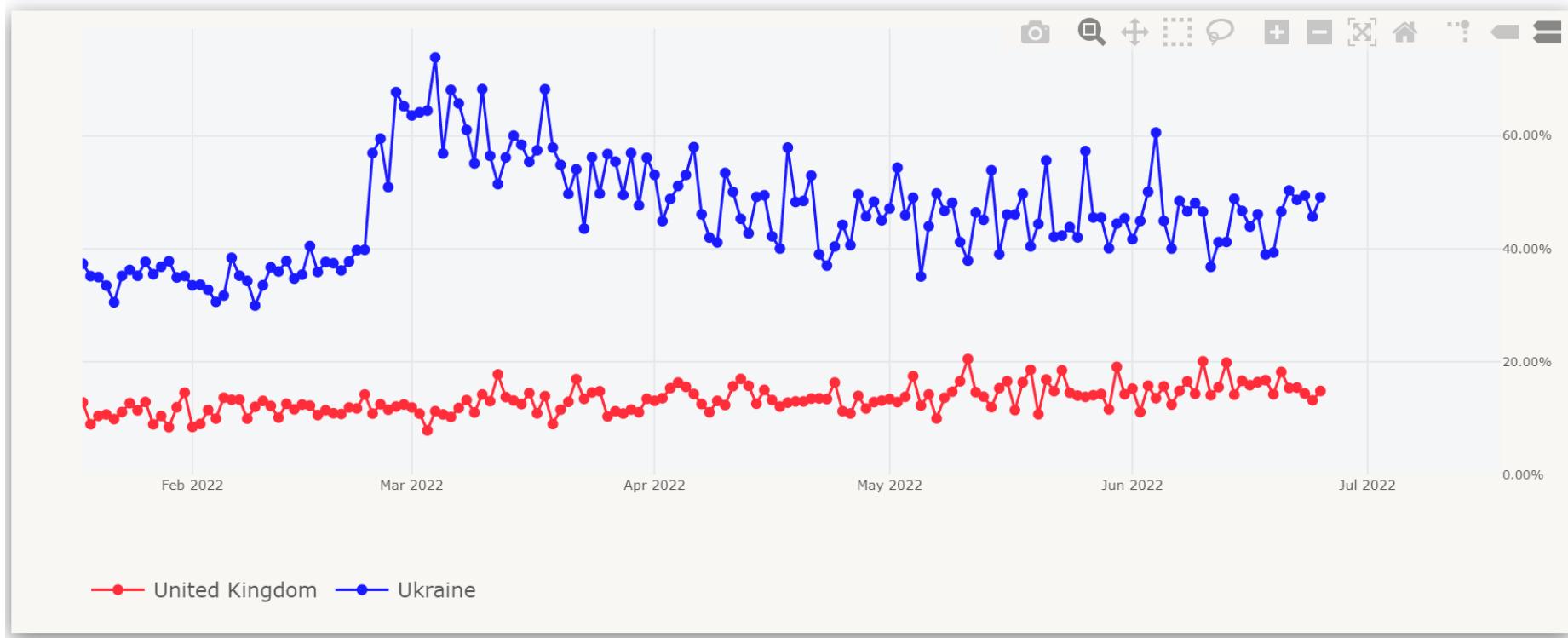
Household Financial Worry

Percentage of respondents that are somewhat or very worried about themselves and their household's finances in the next month



Food Security

Respondents who are very worried or somewhat worried about having enough to eat in the next week



RESEARCH

Open Access

Global trends and predictors of face mask usage during the COVID-19 pandemic



Elena Badillo-Goicoechea^{1†}, Ting-Hsuan Chang^{2†}, Esther Kim³, Sarah LaRocca³, Katherine Morris³, Xiaoyi Deng⁴, Samantha Chiu⁴, Adrienne Bradford⁴, Andres Garcia⁴, Christoph Kern⁵, Curtiss Cobb³, Frauke Kreuter^{4,6,7} and Elizabeth A. Stuart^{1,2*}

Abstract

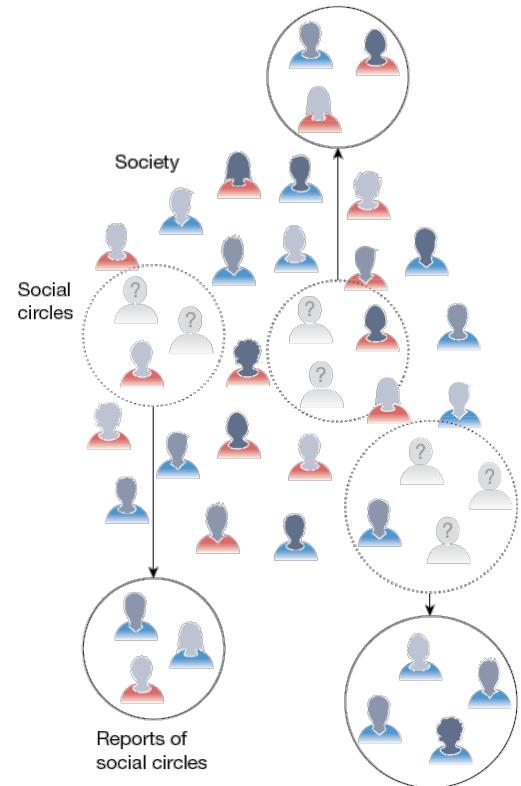
Background: Guidelines and recommendations from public health authorities related to face masks have been essential in containing the COVID-19 pandemic. We assessed the prevalence and correlates of mask usage during the pandemic.

Methods: We examined a total of 13,723,810 responses to a daily cross-sectional online survey in 38 countries of people who completed from April 23, 2020 to October 31, 2020 and reported having been in public at least once during the last 7 days. The outcome was individual face mask usage in public settings, and the predictors were country fixed effects, country-level mask policy stringency, calendar time, individual sociodemographic factors, and

Bias and Weights

Social Sensing: Community COVID Like Symptoms

Do you personally know anyone in your local community who is sick with a fever and either a cough or difficulty breathing?



Adjusting for Sample Bias

Facebook calculates analytic weights aim to correct for random sampling, non-response, and coverage errors.

Survey weights are available for 115 countries but may be revised as Facebook and partners assess sample coverage.

The weight value does not identify the survey respondent.

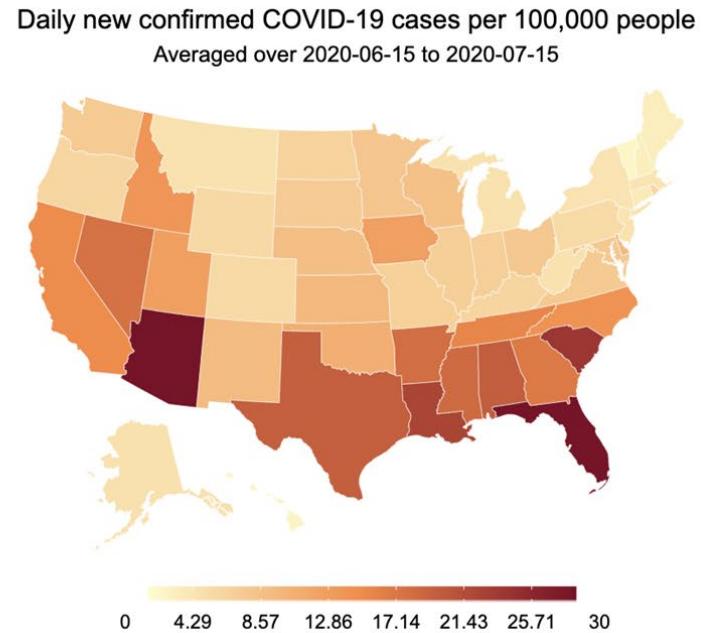
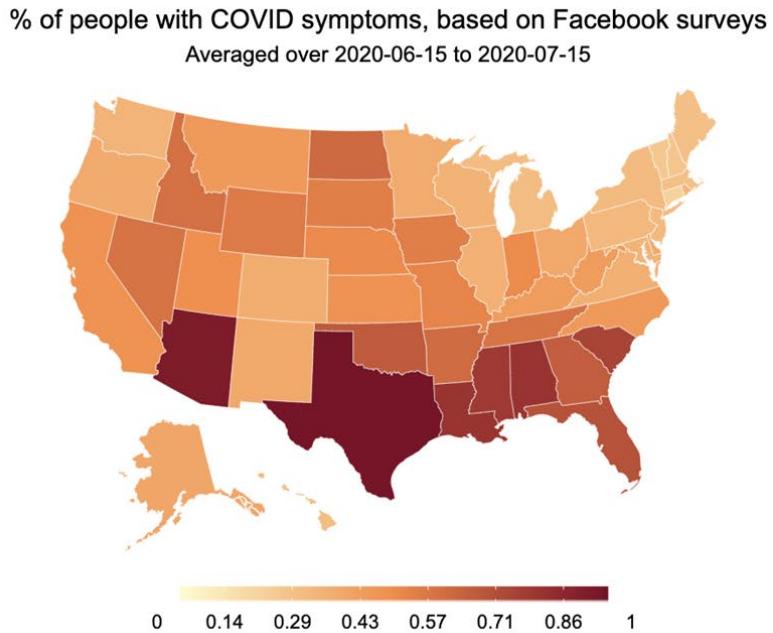
Selection probabilities fully known.
Random daily cross-sectional sample with cool-down periods.

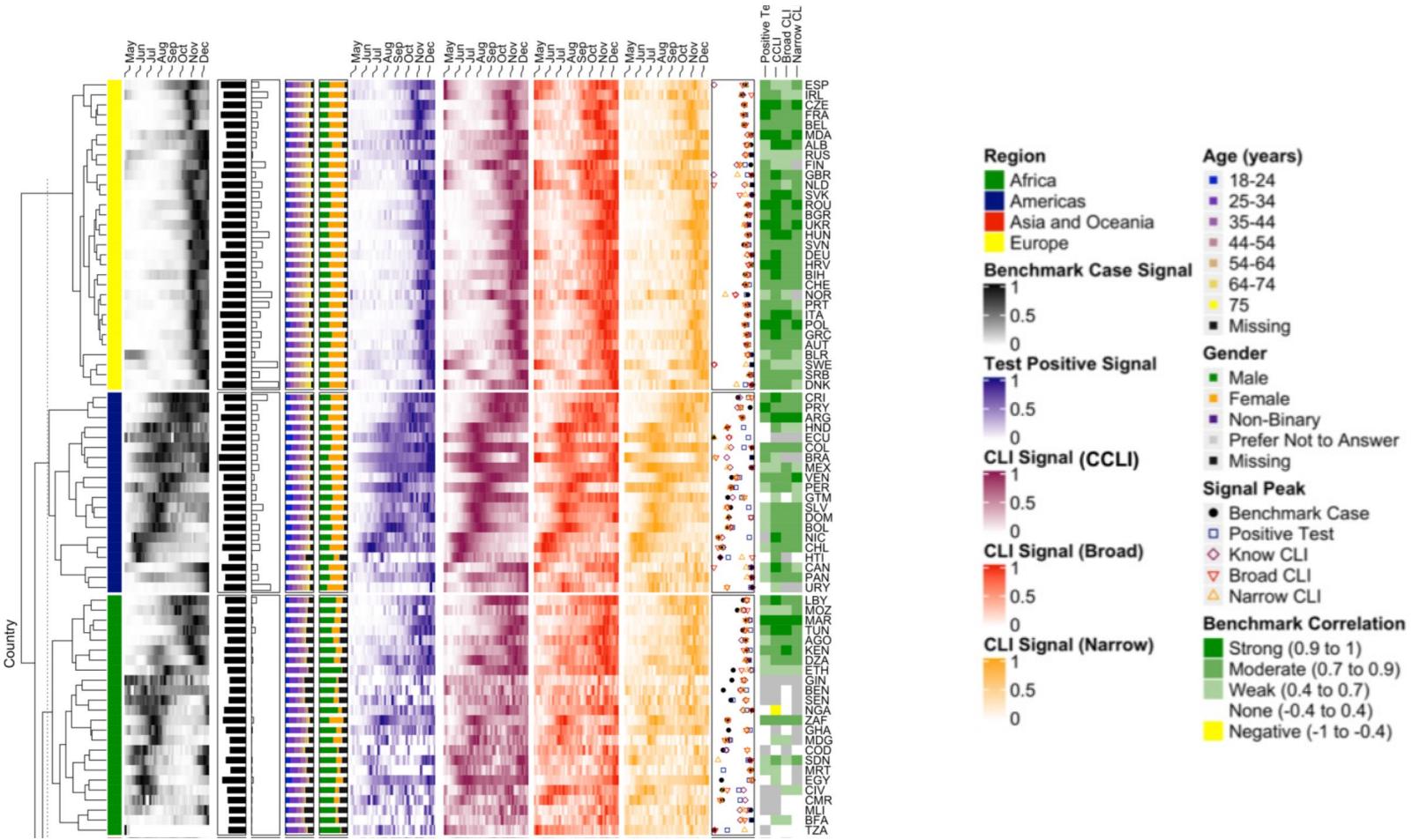
Nonresponse probabilities predicted using information available to Facebook from user data and general response patterns. Variables not shared for privacy reasons.

Poststratification by age, gender, territory using external benchmarks.

Validation and Access

Facebook symptom surveys appear to correlate with incidence (July)





Mask Wearing Insights / Validation

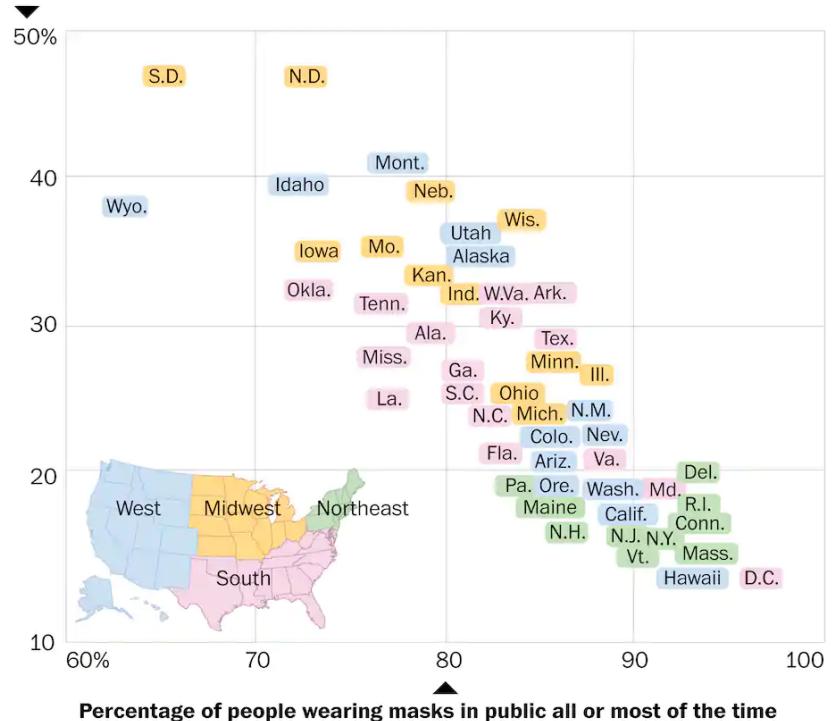
"If Facebook's users are different from the U.S. population generally in a way that the survey weighting process doesn't account for, then our estimates could be biased," cautioned Alex Reinhart, a Carnegie Mellon professor of statistics and data science [...] "But if that bias doesn't change much over time, then we can still use the survey to detect trends and changes."

<https://www.washingtonpost.com/business/2020/10/23/pandemic-data-chart-masks/>

Masking up

Fewer covid-19 symptoms reported in states with higher rates of mask use.

Percentage of people who know someone with covid-19 symptoms



Data as of Oct. 19

Source: Delphi CovidCast, Carnegie Mellon University

THE WASHINGTON POST

CTIS Data Dissemination

Three categories of data are disseminated via the Delphi and UMD APIs, including:

- 1. Individual-level, daily samples** for eligible users with fully executed Data Use Agreements (DUAs) with Meta and with CMU and/or UMD.
- 2. Daily aggregate estimates** of a subset of indicators at the country and subnational region levels disseminated via public APIs at both organizations.
- 3. Weekly and monthly aggregate estimates** broken out by respondent characteristics (e.g., age, gender, vaccination status) at the country and subnational region level publicly disseminated via CSV-formatted files.

Publicly Available, Aggregate Data

Global Survey Data:

<https://covidmap.umd.edu/api.html>

US Survey Data:

<https://cmu-delphi.github.io/delphi-epidata/api/covidcast.html>

Non-Aggregate Data for Research

Researchers from academic and non-profit institutions can request access.

Signed Data Use Agreements are required.

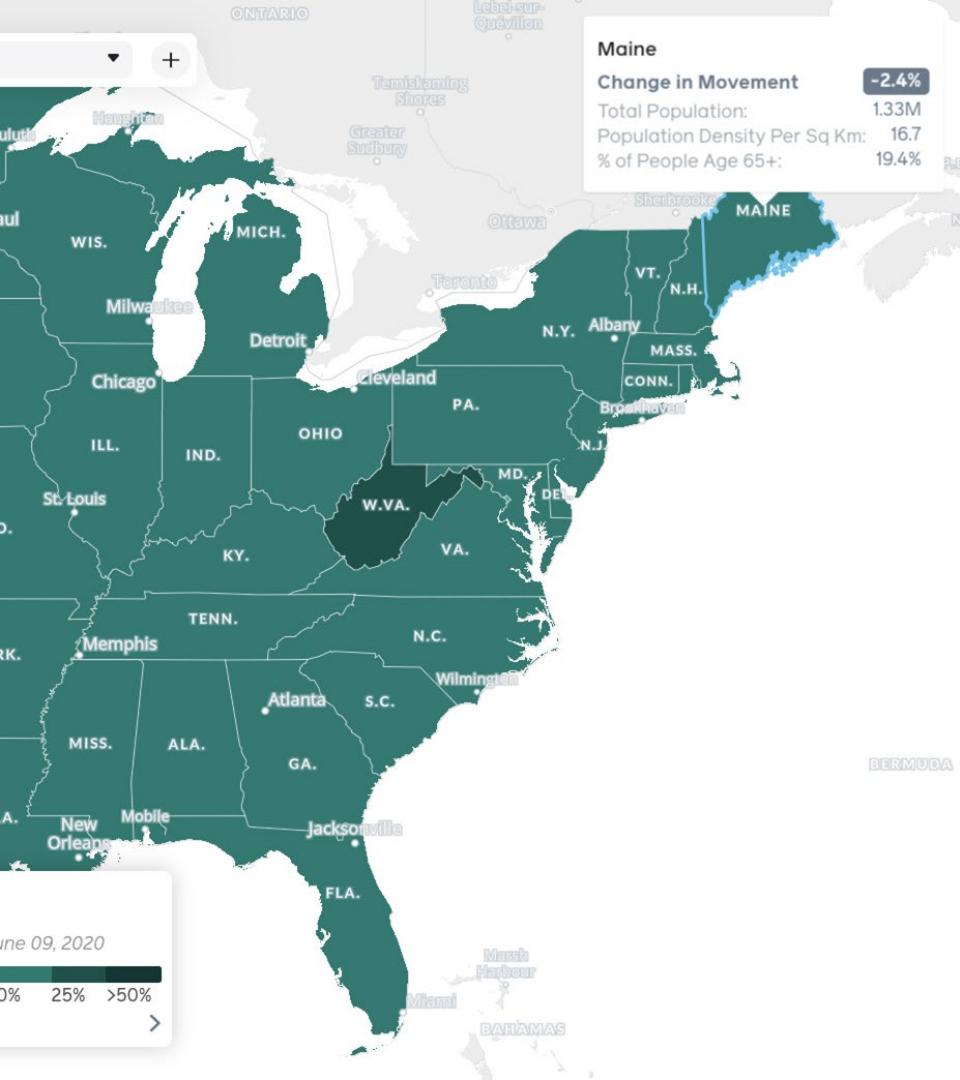
Central portal for project documentation and data access requests is on Facebook's Data for Good website: dataforgood.fb.com.

Additional Resources

Other Complimentary Data Sources Through Data for Good

Delphi, through its COVIDcast Epidata API, publishes case and death data, plus medical claims and aggregates from the US surveys: <https://cmu-delphi.github.io/delphi-epidata/api/covidcast.html>

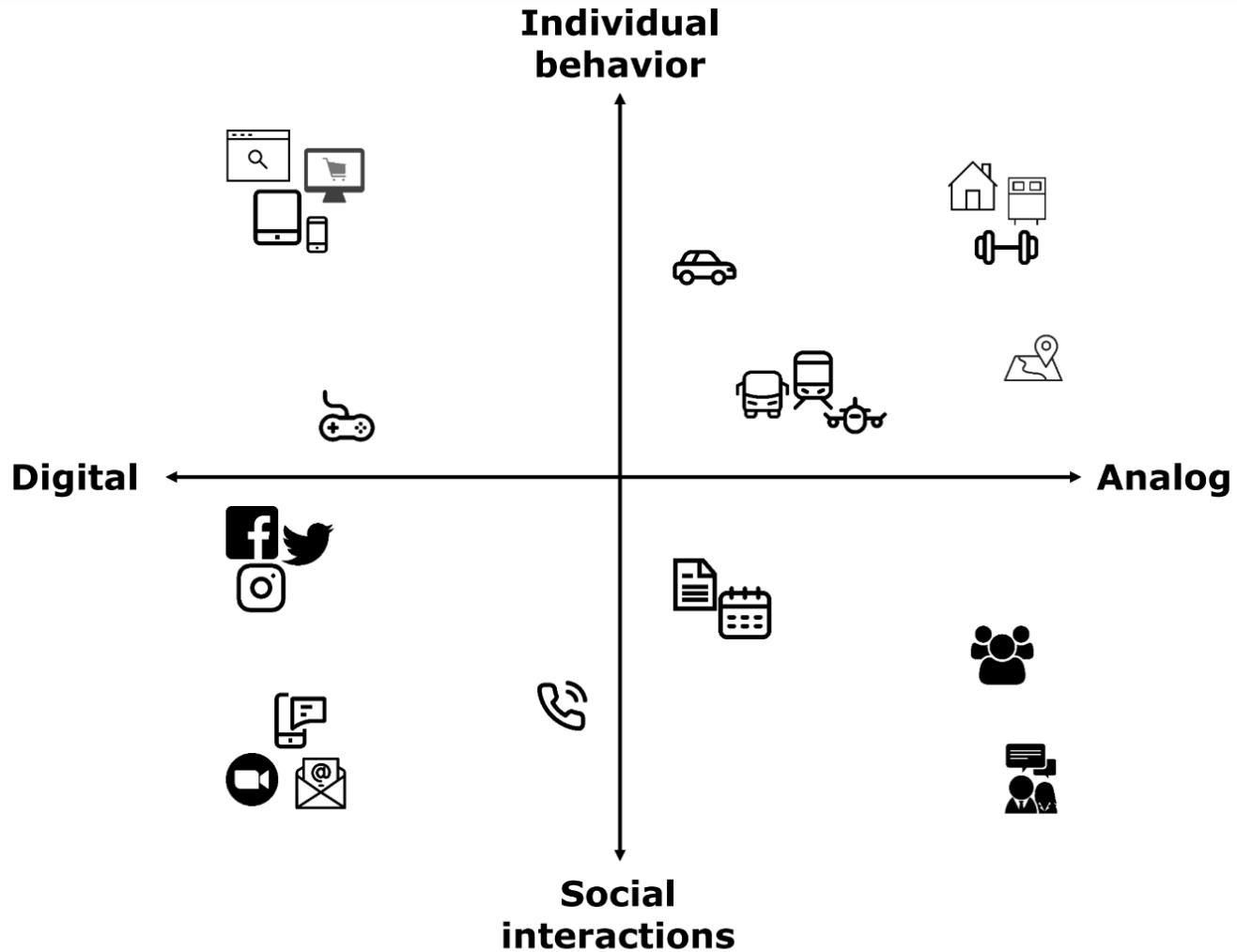
Facebook's Data for Good publishes additional datasets: dataforgood.fb.com.



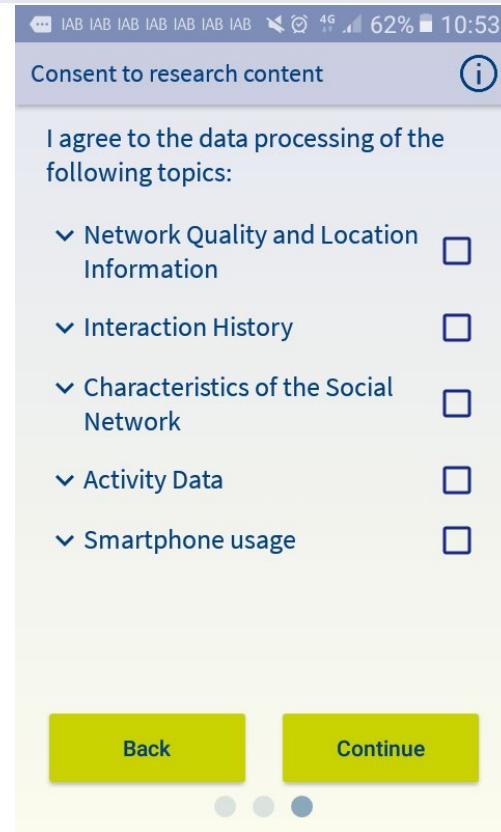
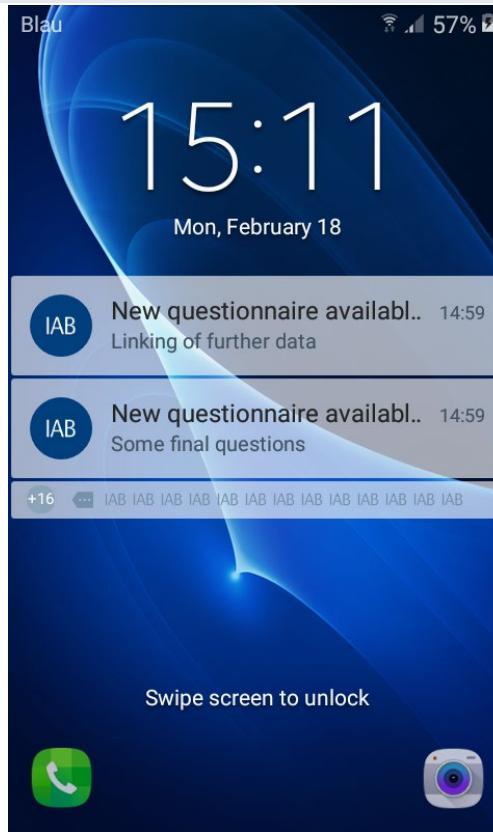
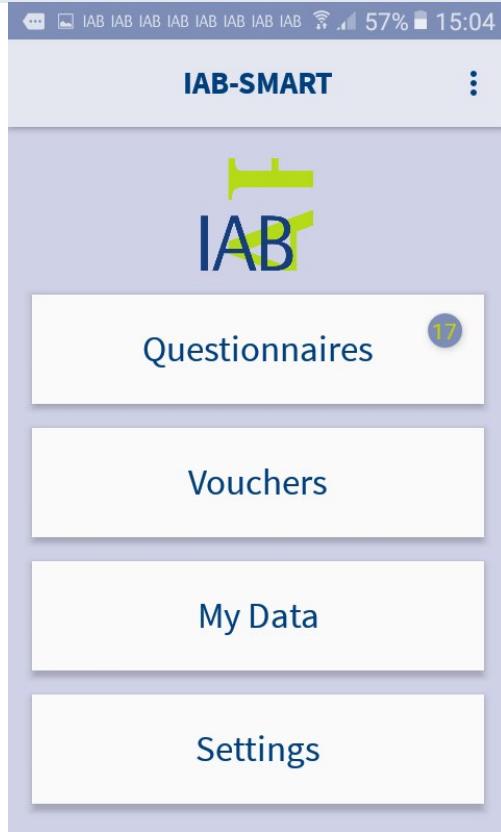


IAB-SMART Studie “Marienthal 2.0”





Structure and Data Collection



Consent to research content



Network Quality and Location Information

Interaction History

Characteristics of the Social Network

Activity Data

This feature allows us to gather information about selected means of transport (e.g. walk, bike, and vehicle) and periods of activity, which are measured by your smartphone sensors. Our aim is to investigate research questions relating to labour market mobility (e.g. commuting behaviour, radii of action).

Smartphone usage

Back

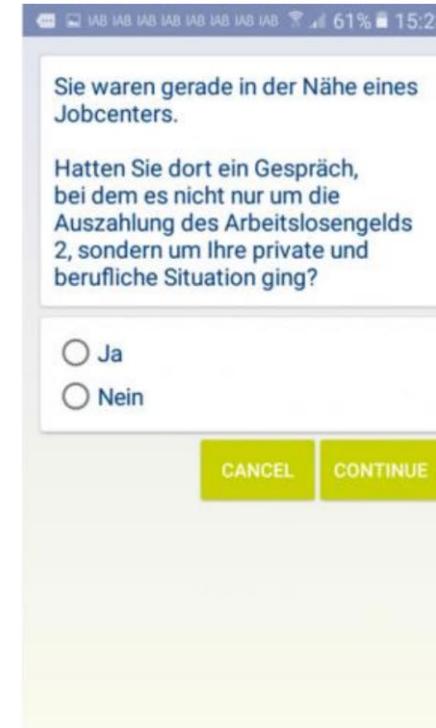
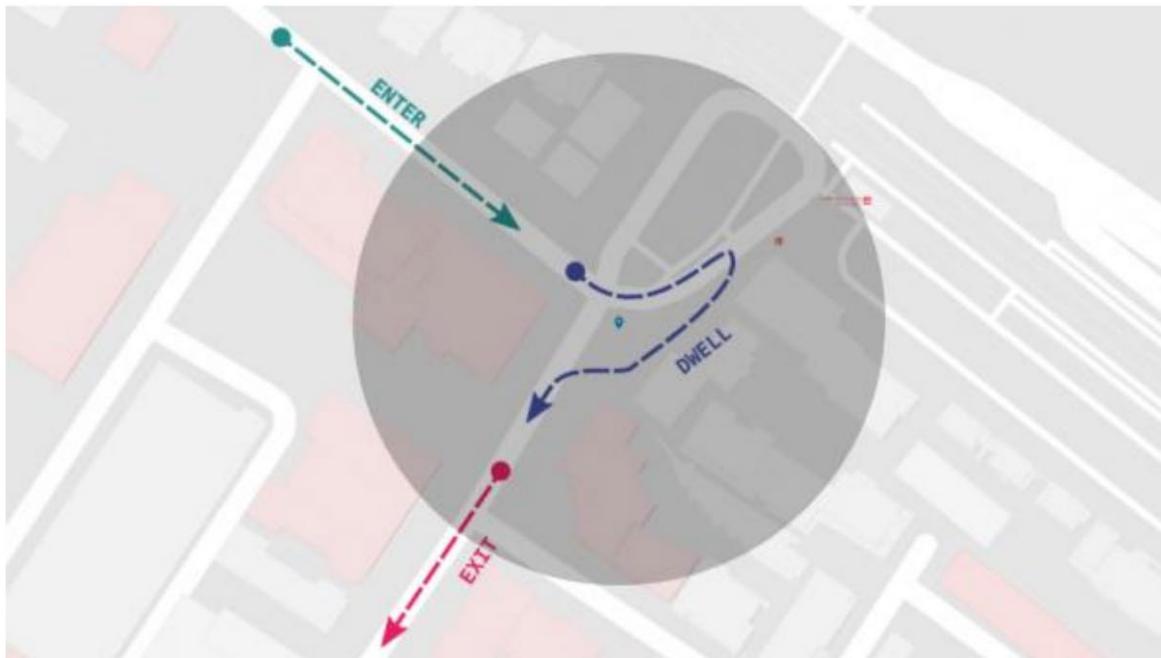
Continue



Data Types and Usage

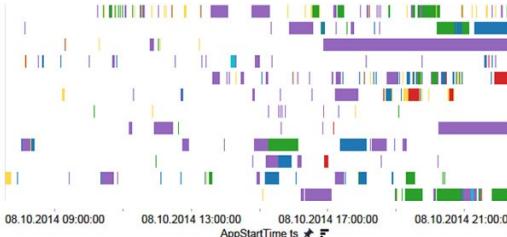
- Network quality and location (geofence triggered questions, radius)
- Interaction history (text messages and calls to measure participation and social inclusion)
- Network characteristics via the phone book (classification of names)
- Activity data (physical activity level, means of transportation etc.)
- Smart phone usage

Triggering: Geofencing

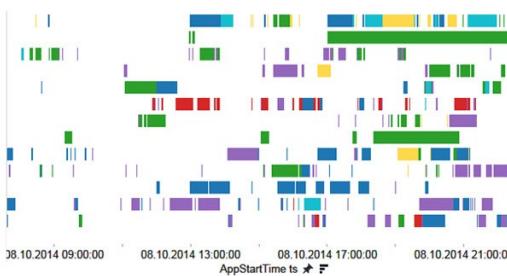


Haas, G.-C., Trappmann M., Keusch F., Bähr S. & Kreuter F. (2020). Using Geofences to Collect Survey Data: Lessons Learned From the IAB-SMART Study in Survey Methods: Insights from the Field, Special issue: 'Advancements in Online and Mobile Survey Methods'. <https://surveyinsights.org/?p=13405>

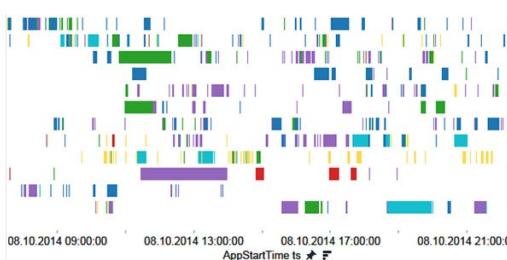
Example: App Usage



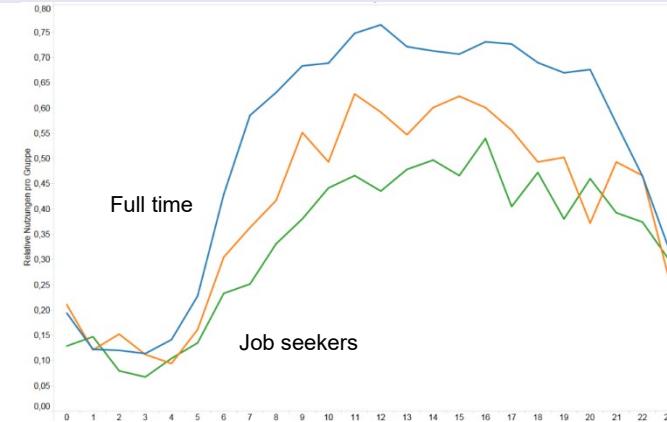
Full-time
employed
→ App use
past 5pm



Part-time
employed
→ App use at
noon



Job seekers
→ Continuous
app use



Email



Entertainment

Which functions get activated?

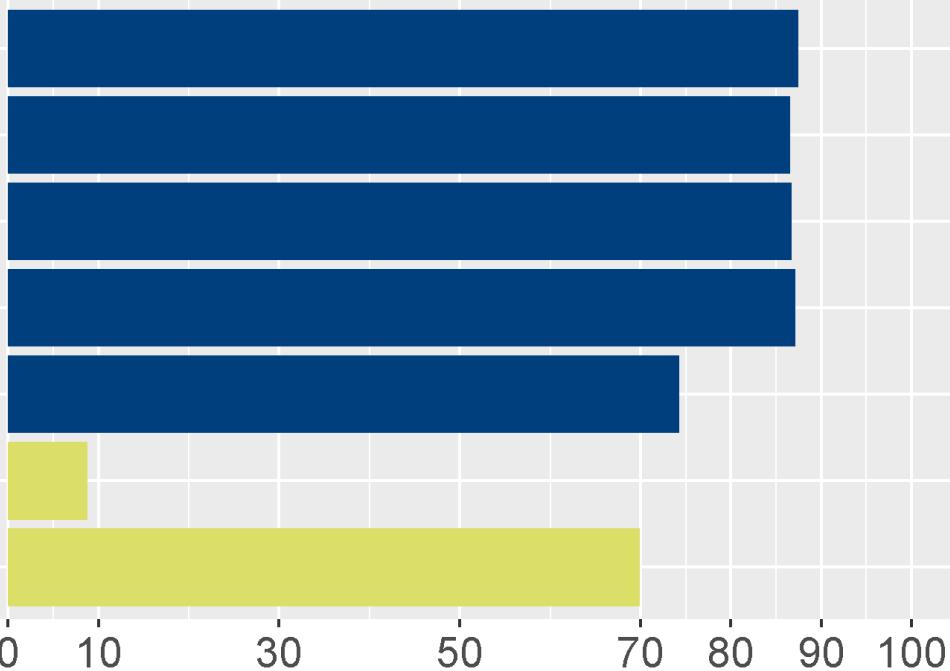
Einwilligung in Forschungsinhalte

Hiermit willige ich in die Datenverarbeitung folgender Inhalte ein:

- Netzqualität und Standortinformationen
- Interaktionsverlauf
- Merkmale des sozialen Netzwerks
- Aktivitätsdaten
- Smartphone-Nutzung

Zurück Weiter

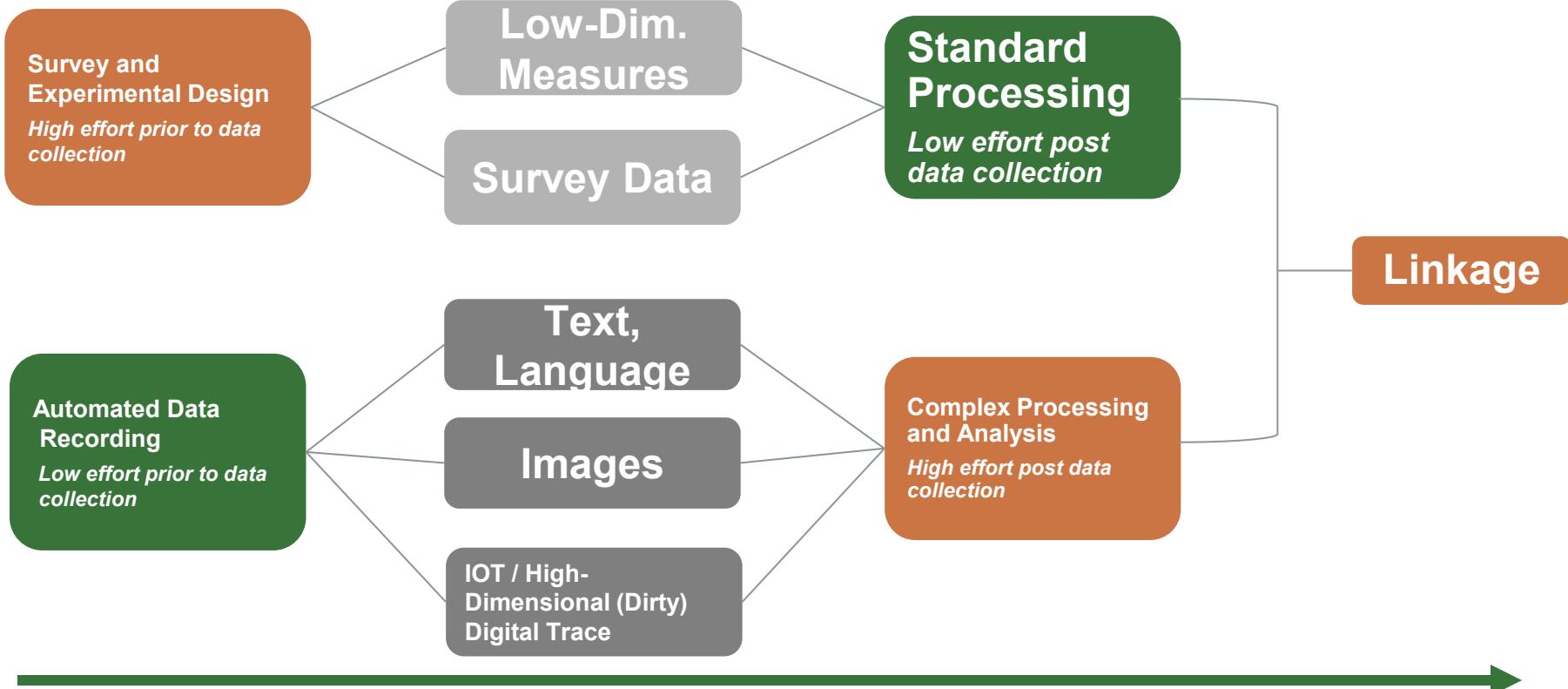
Network quality and location information



Activation of functions in percent

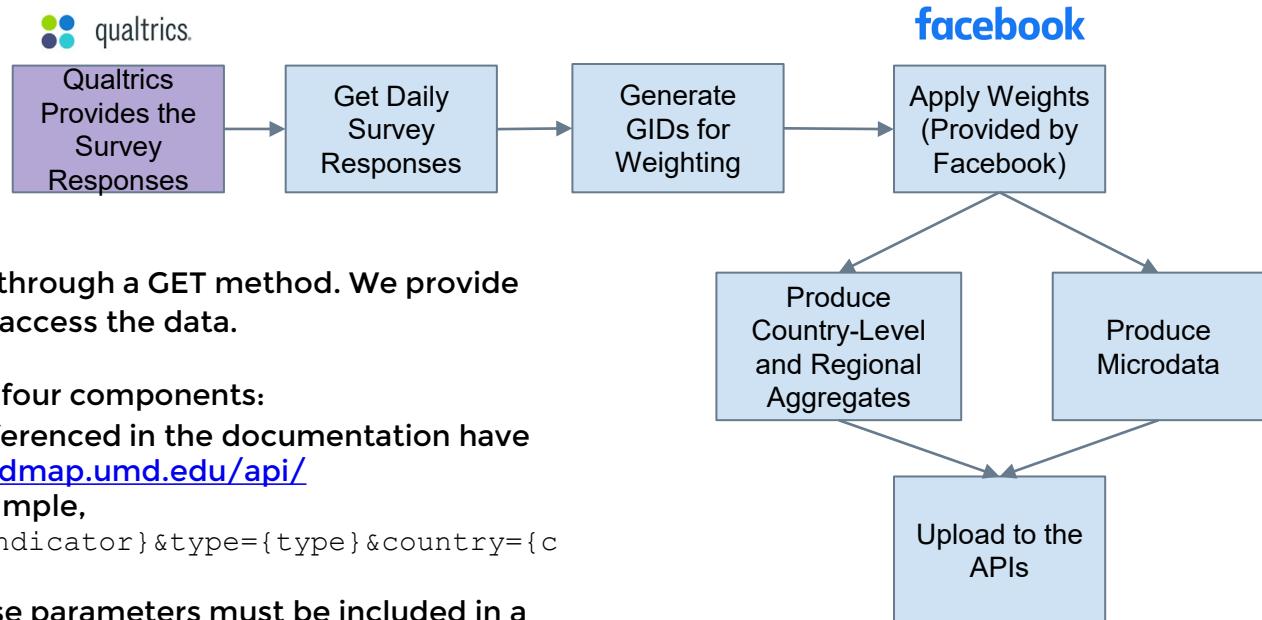


Pipelines, Outreach, Skills



Prepare Pipelines

Data Pipeline Workflow



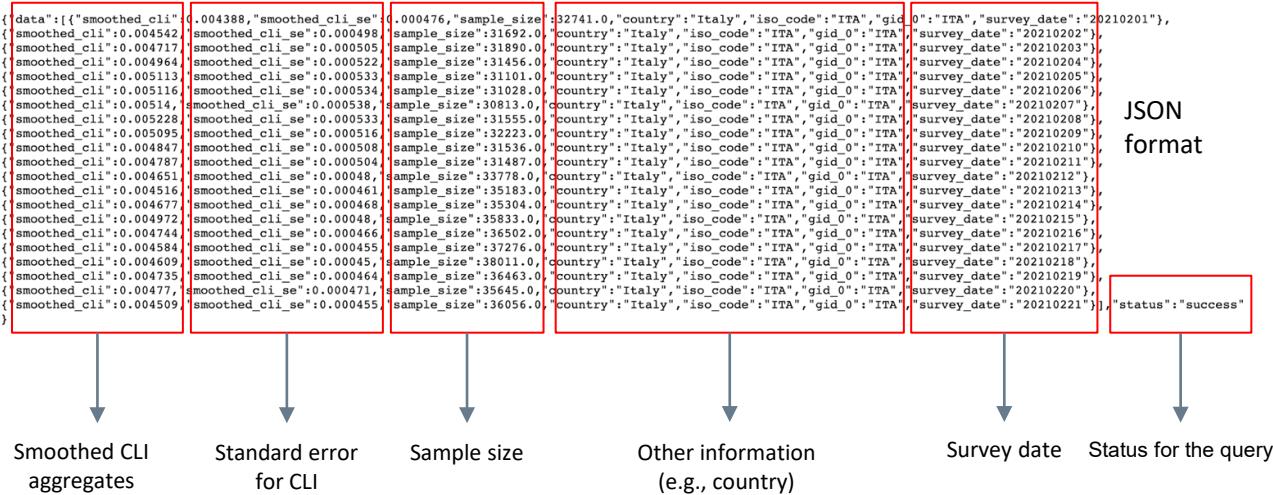
The Open Data API is accessed through a GET method. We provide scripts in both Python and R to access the data.

The address of the API includes four components:

- The base path. All URLs referenced in the documentation have the base path <https://covidmap.umd.edu/api/>
- The endpoint path. For example,
resources?indicator={indicator}&type={type}&country={country}&date={date}
- Required parameters. These parameters must be included in a request. In the example above, {indicator}, {type}, {country}, and {date} are required. In a request, you would replace the placeholders with real values.
- Optional parameters. These additional parameters can be included to customize a query to get regional aggregates, smoothed aggregates, all regions, etc.

An example:

API query return for smoothed estimates of
CLI for Italy
02/01/2021-02/21/2021



Select country

Norway

Select regions

Select

2nd country

Germany



45%

Map

Select variable

Direct Contact With Others

Totals

Behaviors

Mental Health

Beliefs

Symptoms

News and Information

Vaccine

Vaccine Barriers

Testing

Wave 10 and 11 cli Trend Break

Blog

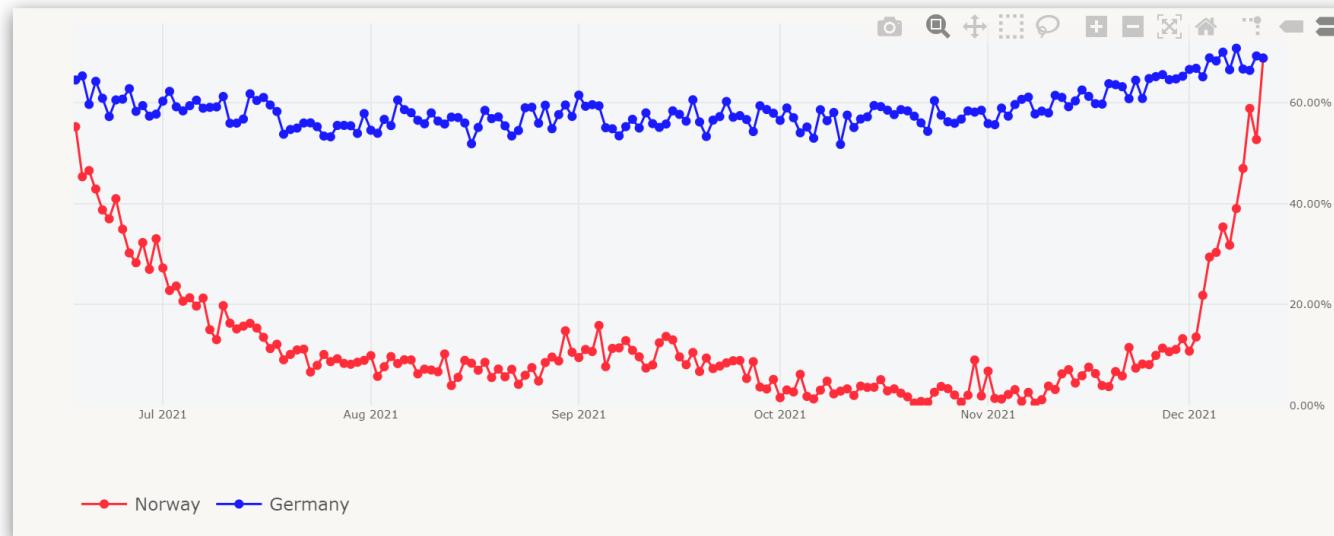
Note: Aggregates for the API, map, and survey results page are shown for weighted countries only. Please find the complete list of weighted countries that are weighted by Facebook [here](#).

Microdata files include both weighted and unweighted countries. For more information on how to gain access to the microdata go [here](#).

The Global COVID-19 Trends and Impact Survey (CTIS) is subject to several crucial limitations, many of which are common to web surveys. Anyone using the data to make policy decisions or answer research questions should be aware of these limitations, as outlined [here](#).

Saw Most Other People Wear Masks

Percentage of people who reported that all or most people they saw in public in the last 7 days wore a mask



Norway

Sample Size	Standard Error	Per Million	Latest Value
-	-	0	-

[Weekly Change](#) [Download CSV](#)


Germany

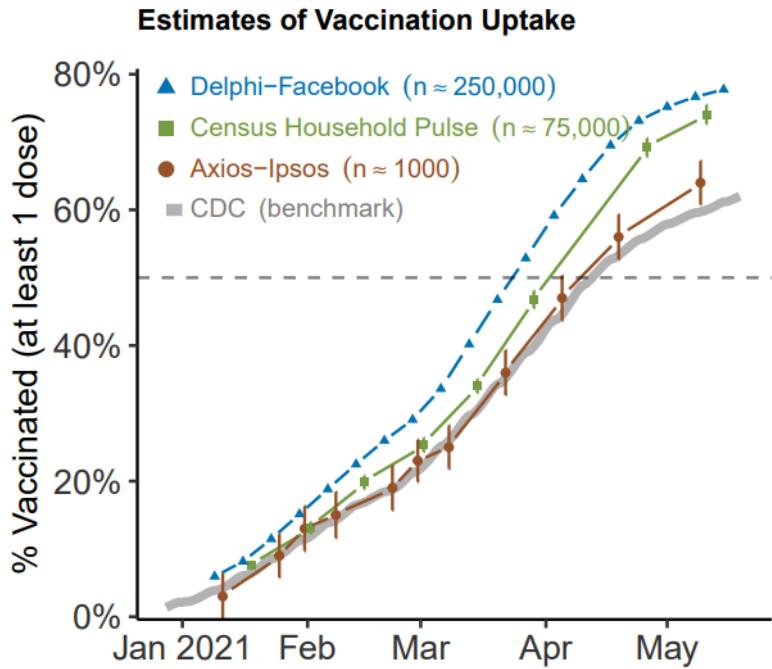
Sample Size	Standard Error	Per Million	Latest Value
1970	0.01424	0.688M	68.82%

[Weekly Change](#) [Download CSV](#)


Create Communication



Information gets processed with a goal in mind!



Education	% Vaccinated			% Willing		
	AX	HP	FB	AX	HP	FB
High School	28%	39%	40%	32%	40%	35%
Some College	36	44	52	30	38	27
4-Year College	36	54	62	45	36	26
Post-Graduate	56	67	73	33	26	19

Bradley et al. 2021 arXiv:2106.05818

Thinking about estimands as a survey provider: Our recommended Don'ts

- **Make point estimates of population quantities** (such as the exact percentage of people who meet a certain criterion) without reference to other data sources.
- **Analyze very small or localized demographic subgroups.** Very small demographic groups may disproportionately include respondents who pick their demographics at random or attempt to disrupt the survey in other ways, even if those respondents are rare overall.

Thinking about estimands as a survey provider: Our recommended Do's

- **Track changes over time**, such as to monitor sudden increases in reported symptoms or changes in reported vaccination attitudes.
- **Make comparisons across space**, such as to identify regions in a country with much higher or lower values.
- **Make comparisons between groups**, such as between occupational or age groups, keeping in mind any sample limitations that might affect these comparisons.
- **Augment data collected from other sources**, such as more rigorously controlled surveys with high response rates.

Train the Team

DOMAIN EXPERT

User, analyst, or leaders with deep subject matter expertise related to the data, its appropriate use, and its limitations

SYS ADMIN

Team member responsible for defining and maintaining a computation infrastructure that enables large scale computation



METHODOLOGIST

Team member with experience applying formal research methods, including survey methodology and statistics

COMPUTER SCIENTIST

Technically skilled team member with education in computer programming and data processing technology



SOCIAL DATA SCIENCE CENTER

<https://socialdatascience.umd.edu/training/>

SoDa

HOME ABOUT PEOPLE RESEARCH TRAINING NEWS & EVENTS

Search

A blurred background image showing a close-up of a computer keyboard and a laptop screen displaying a pie chart with green and orange segments. Overlaid on the left side is the word "Training" in a large, semi-transparent white font.

Training

Educating the next generation of researchers and data scientists onsite and through long distance education for careers such as survey methodologists, political analysts, quantitative sociologists, applied economists, computational criminologists, social media analysts, data journalists, and smart city planners.



Degree and non degree offerings



Onsite and Online Programs



Graduate and Undergraduate Degrees

[IPSDS](#)[Events](#)[Project](#)[Newsletter](#)[Contact](#)[Search](#)

OPEN COURSES & CERTIFICATES

Enhance your skills in one specific area with our graduate-level Open Courses & Certificates!

[MORE](#)

<https://survey-data-science.net/>





BERD@NFDI: NFDI Consortium for Business, Economic and Related Data

BERD@NFDI is an initiative to build a powerful platform for collecting, processing, analyzing and preserving **Business**, **Economic** and **Related Data** – all in one place. We will facilitate the integrated management of algorithms and data along the whole research cycle, with a special focus on unstructured (big) data such as video, image, audio, text or mobile data.

BERD@NFDI will provide infrastructures to the challenges of the expanded empirical research. We will not only foster community building, offer publicly available and online accessible data sets, and enhance data documentation and preservation guided by the FAIR principles. We will also provide an algorithm repository and benchmarks, computing and storage power to analyze (big) data as well as a broad set of APIs to interact with external systems.

Learn more about us in our "[BERD@NFDI in a Nutshell](#)" presentation.



news

[BERD@NFDI InfraTalk postponed](#)
November 28, 2022

[Current Vacancy @ZBW](#)
October 28, 2022

[Student Assistant \(m/f/d\) Wanted!](#)
October 19, 2022

[→ all news](#)

follow us



Tweets from @BERD_NFDI

BERD as an Open Platform for Analysis



Domain Specialist

- Define task in accordance with theory
- Refine theory based on results

BERD

- Interactive notebooks
- Similarity search on studies
- AutoML removes drudge work
- ...

Data Scientist

- Map task to analysis
- Refine and optimize analysis pipelines

BERD facilitates optimal collaboration between domain specialists and data scientists



Structure and
Data Collection

Instruments and
Weights

Validation and
Access



Structure and
Data Collection

Participation
and Privacy

Validation



Prepare
Pipelines

Create
Communication

Train the Team

Empirical Research with Large Datasets

Workshop

BPLIM



BANCO DE
PORTUGAL
EUROSISTEM



Thank you!

frauke.kreuter@lmu.de

PORTO | 19 - 20 DEC. 2022