

Good data saves lives: But what are good data?

Samuel V. Scarpino, PhD

Vice President, Pathogen Surveillance

The Rockefeller Foundation & the Pandemic Prevention Institute

External Faculty

Santa Fe Institute

Vermont Complex Systems Center

Affiliate Faculty

Northeastern University

Co-Founder, Global.health

@svscarpino

scarpino.github.io



nature

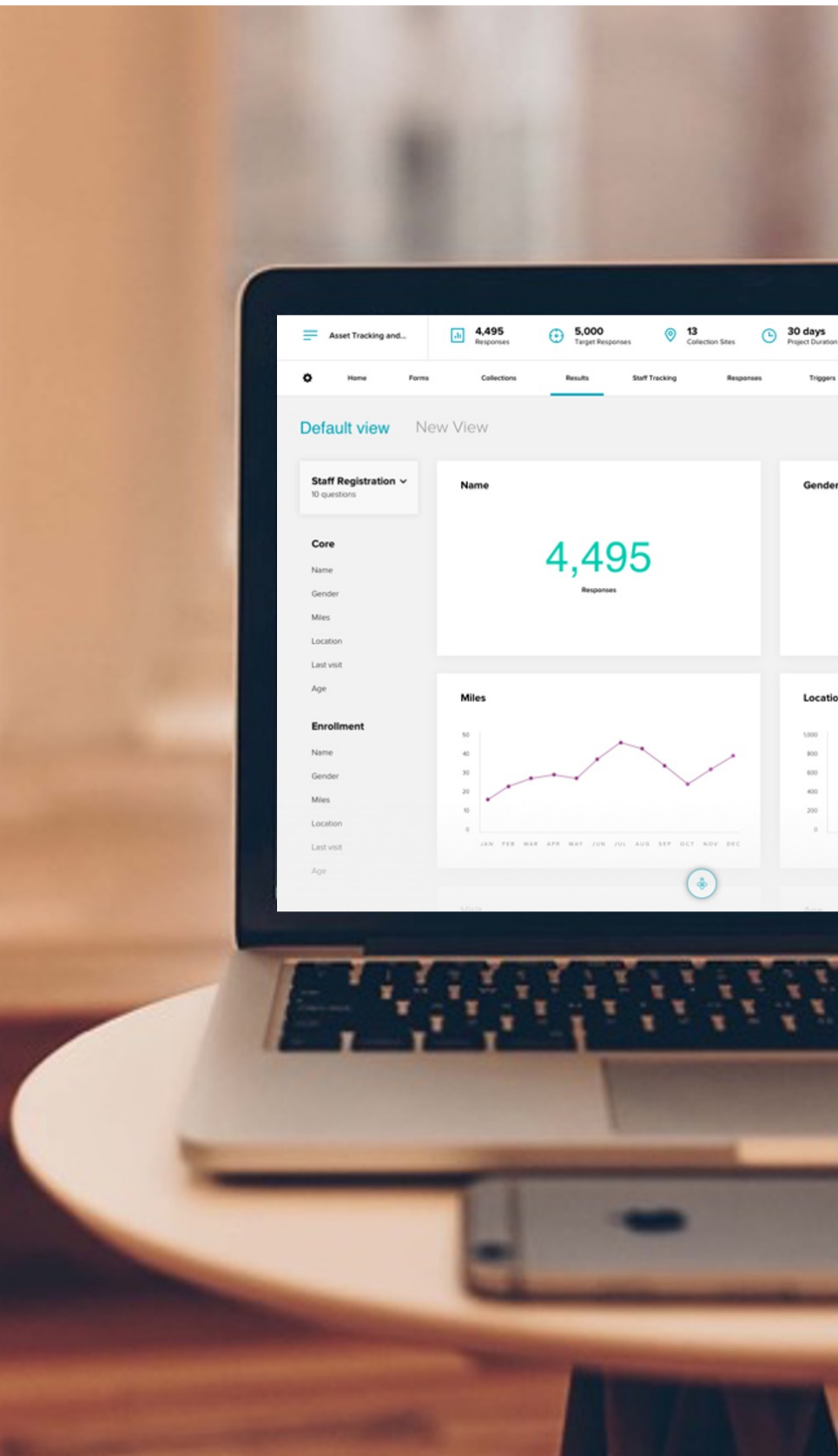
Out of the Syrian crisis, a data revolution takes shape

"I think Dharma is special because it has been developed by people who have worked in these chaotic situations, and it's been road-tested and improved in the midst of reality."

Jeremy Farrar

Director of Biomedical-Funding
Wellcome Trust





What are good data?

What are
good data?



Modified from Maslow's hierarchy of needs

Technology
can support
the whole
pyramid

IMPACT
ANALYTICS
SHARE
STORE
COLLECT

HIERARCHY OF DATA NEEDS

CONNECTIVITY & POWER

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped
II	Machine-readable, but proprietary (e.g., Excel)	Data can be analyzed using proprietary software	Data are still trapped. Stuck with expensive software

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped
II	Machine-readable, but proprietary (e.g., Excel)	Data can be analyzed using proprietary software	Data are still trapped. Stuck with expensive software
III	Machine-readable and non-proprietary (e.g., CSV)	Data can be analyzed using most any software	Data are open, but can be challenging to analyze/share

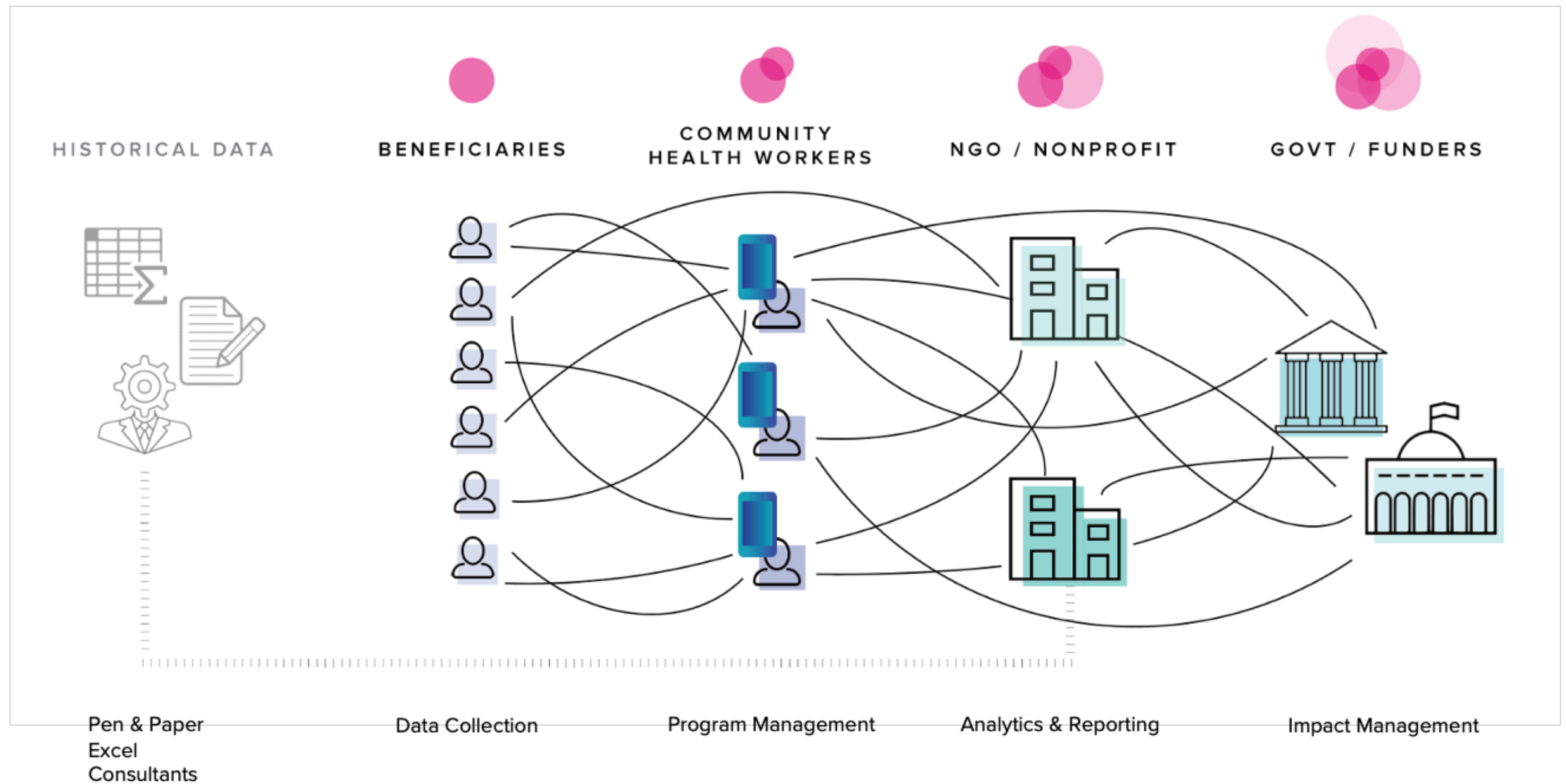
Adapted from The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences by Kitchin

What are good data?

Level	Requirements	Benefits	Costs
I	Non-machine readable	Data are available	Data are trapped
II	Machine-readable, but proprietary (e.g., Excel)	Data can be analyzed using proprietary software	Data are still trapped. Stuck with expensive software
III	Machine-readable and non-proprietary (e.g., CSV)	Data can be analyzed using most any software	Data are open, but can be challenging to analyze/share
IV	Machine-readable, non-proprietary, and linked to metadata/schema/etc.	Data can be readily shared and are ready for analytics/ML/AI	Requires technology

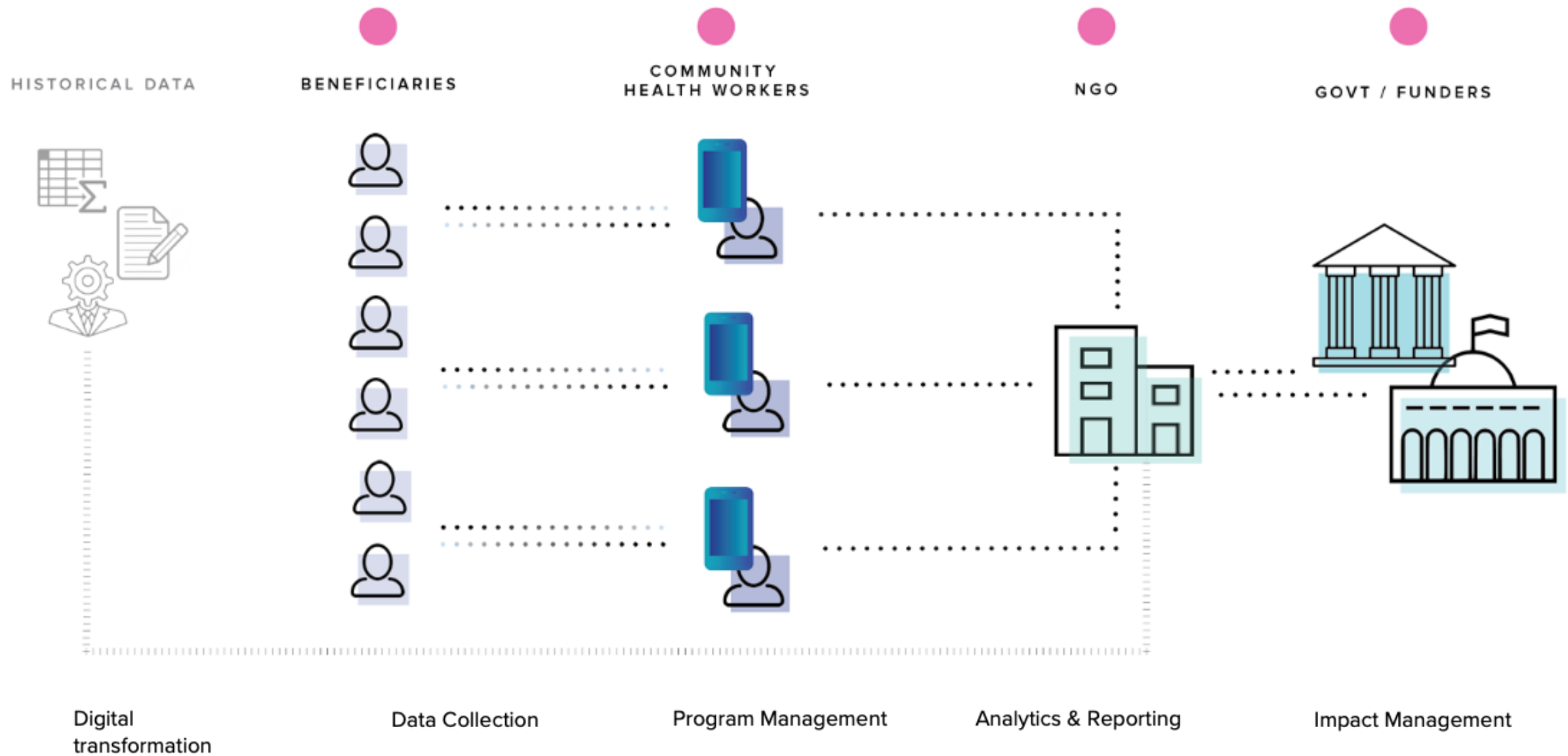
Adapted from The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences by Kitchen

Current state of data



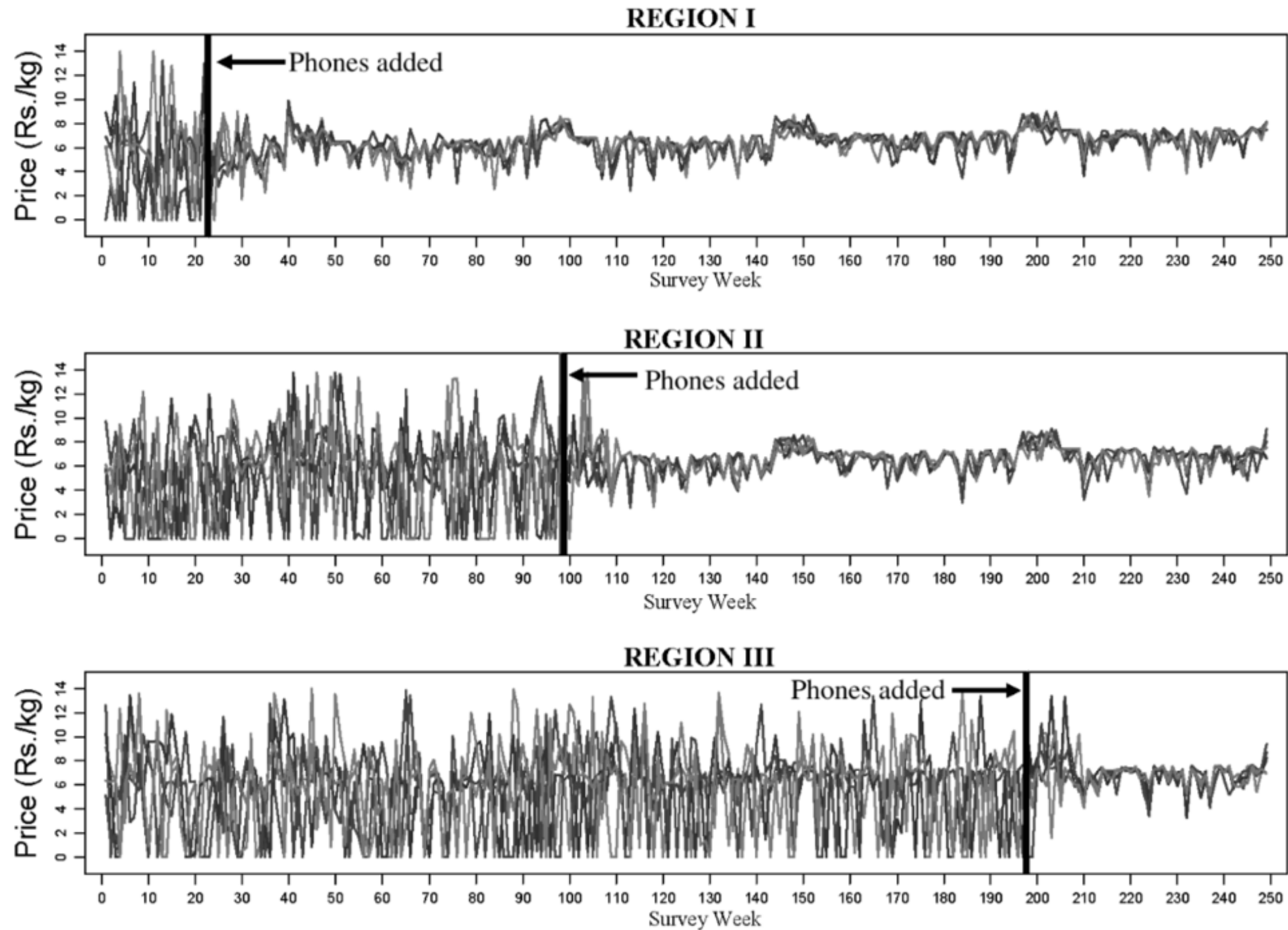
Kristen Clark, Dharma Platform 2019

Good data can improve impact measurement



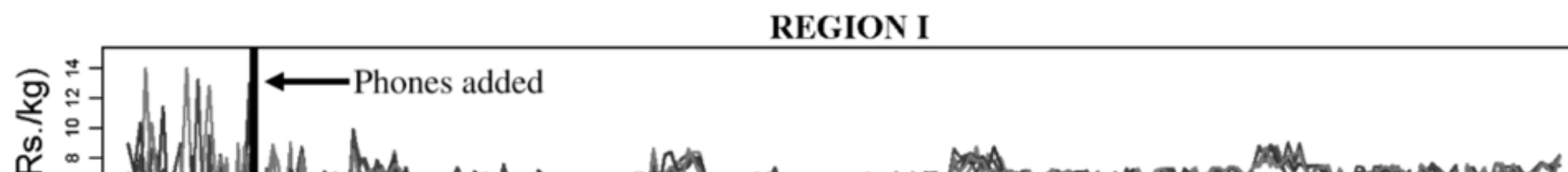
Kristen Clark, Dharma Platform 2019

We know the story of mobile phones & fishing



Jensen 2007

But why don't the results generalize?



Research Article

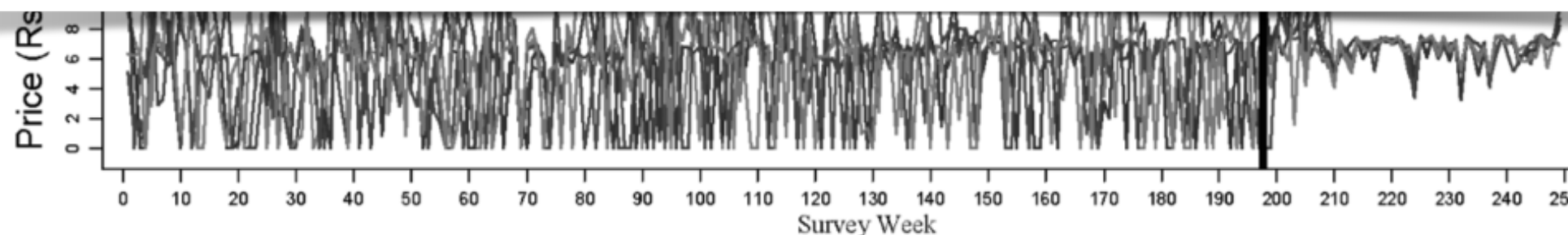
On the Importance of Price Information to Fishers and to Economists: Revisiting Mobile Phone Use Among Fishers in Kerala

Janaki Srinivasan

International Institute of Information Technology–Bangalore

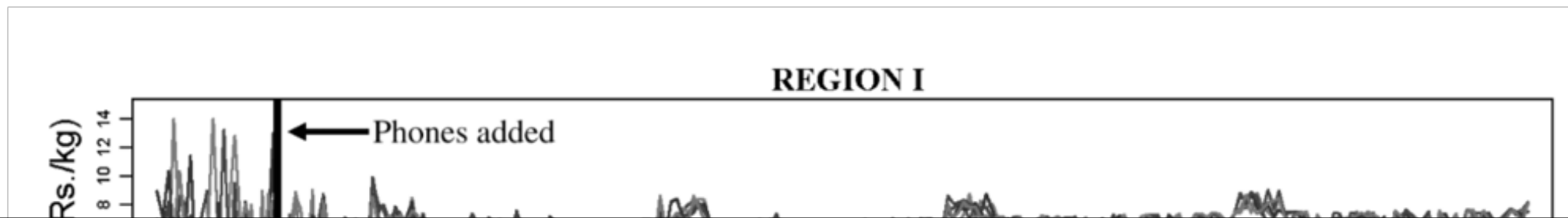
Jenna Burrell¹

University of California, Berkeley



Jensen 2007

Meta-data were under-valued



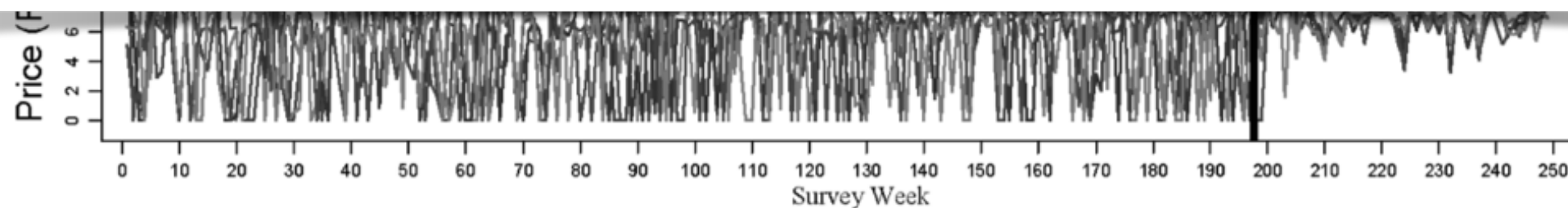
This is the extent to which the actual practice of mobile phone use among north Kerala's fishers is specified in the article. While the quantitative data that forms the substance of Jensen's argument about the reduction of price dispersion is collected systematically, details on exactly how fishers use phones are sparse in his account and lack the same kind of transparency about how such insights were acquired. Some of these details are deprioritized to footnotes. This is (broadly) a reflection of what counts as evidence in econometric analysis. Collected prices are data, but details on processes are merely background or explanatory material.

Janaki Srinivasan

International Institute of Information Technology–Bangalore

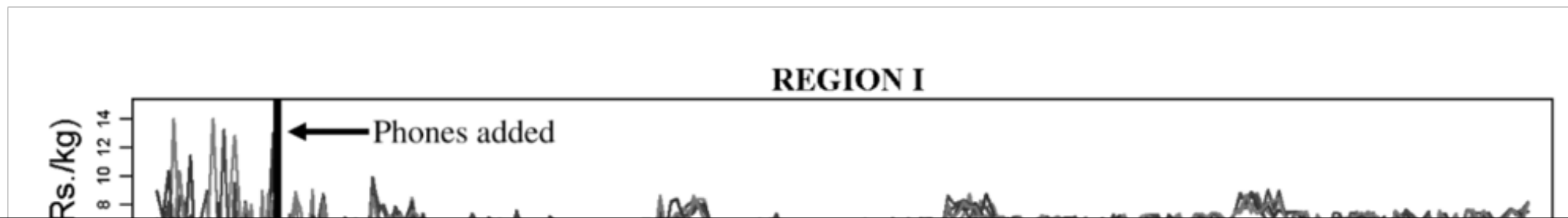
Jenna Burrell¹

University of California, Berkeley



Jensen 2007

Meta-data were under-valued



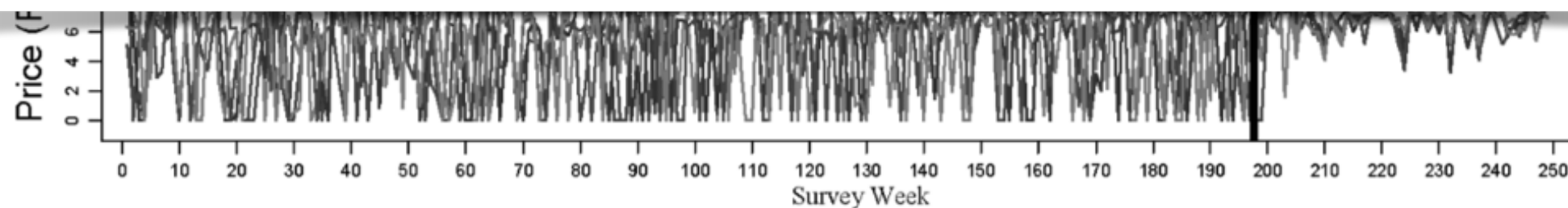
This is the extent to which the actual practice of mobile phone use among north Kerala's fishers is specified in the article. While the quantitative data that forms the substance of Jensen's argument about the reduction of price dispersion is collected systematically, details on exactly how fishers use phones are sparse in his account and lack the same kind of transparency about how such insights were acquired. Some of these details are deprioritized to footnotes. This is (broadly) a reflection of what counts as evidence in econometric analysis. Collected prices are data, but details on processes are merely background or explanatory material.

Janaki Srinivasan

International Institute of Information Technology–Bangalore

Jenna Burrell¹

University of California, Berkeley



Jensen 2007

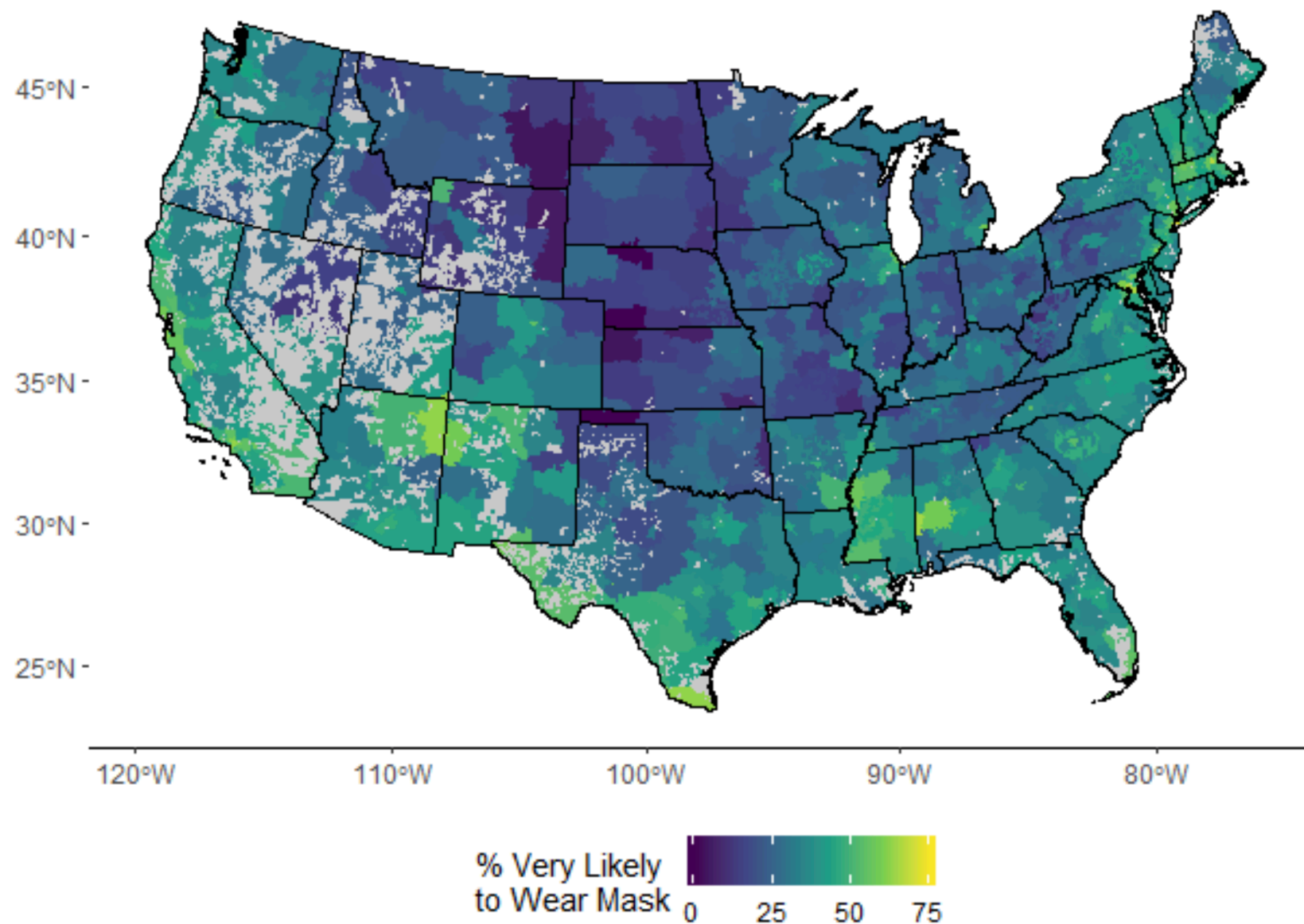
A similar effect was seen with mask mandates

Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study

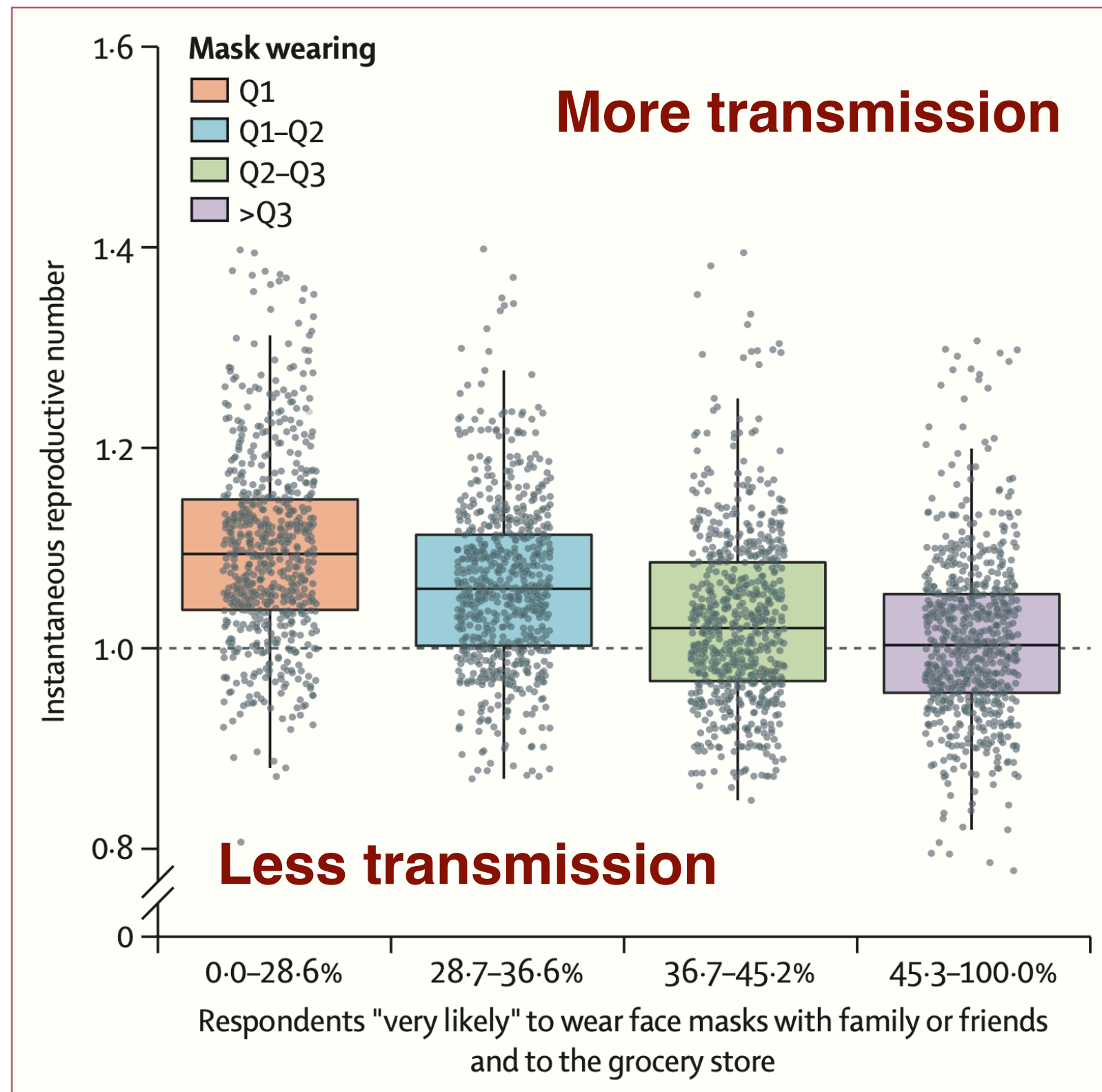
Benjamin Rader, Laura F White, Michael R Burns, Jack Chen, Joseph Brilliant, Jon Cohen, Jeffrey Shaman, Larry Brilliant, Moritz U G Kraemer, Jared B Hawkins, Samuel V Scarpino, Christina M Astley, John S Brownstein



~400k responses across the USA



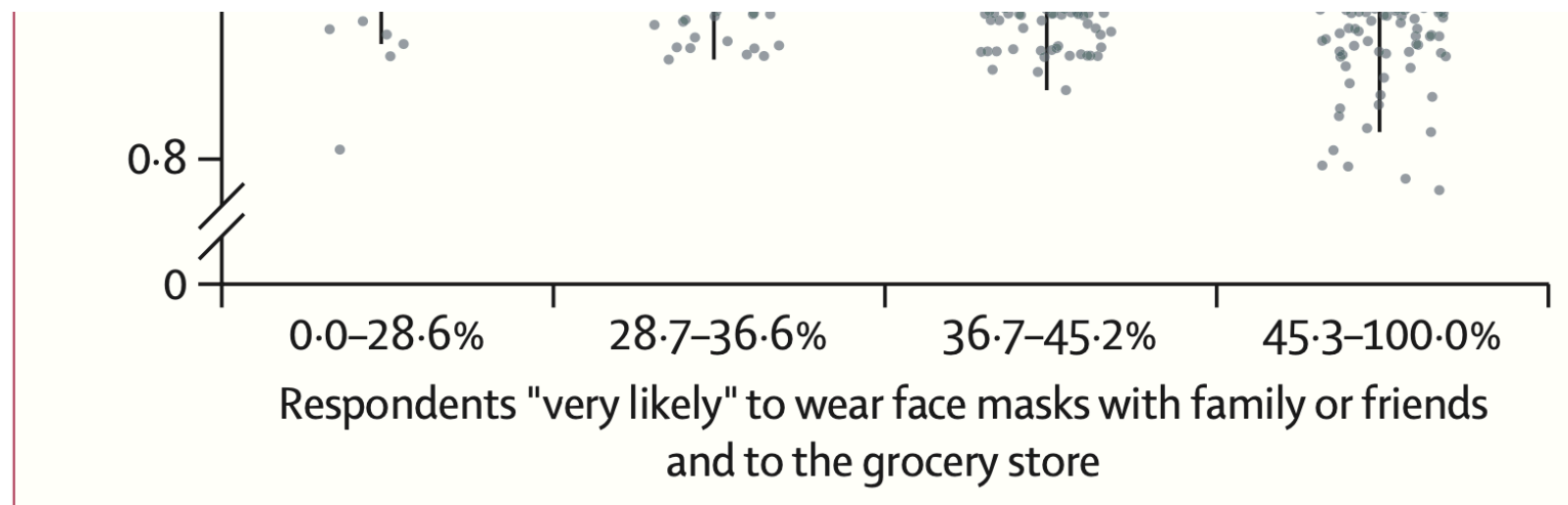
Mask wearing slowed COVID-19 transmission



But wasn't strongly influenced by mandates



Segmented regression analysis of reported mask-wearing showed no statistically significant change in the slope after mandates were introduced;



And with vaccine mandates

PNAS

RESEARCH ARTICLE

SOCIAL SCIENCE

 OPEN ACCESS



Opposition to voluntary and mandated COVID-19 vaccination as a dynamic process: Evidence and policy implications of changing beliefs

Katrin Schmelz^{a,b,c,1}  and Samuel Bowles^c 

Edited by Margaret Levi, Stanford University, Stanford, CA; received October 12, 2021; accepted January 24, 2022

What mattered most was trust in public institutions

PNAS

RESEARCH ARTICLE

SOCIAL SCIENCE

 OPEN ACCESS



Opposition
a dynamic
changing

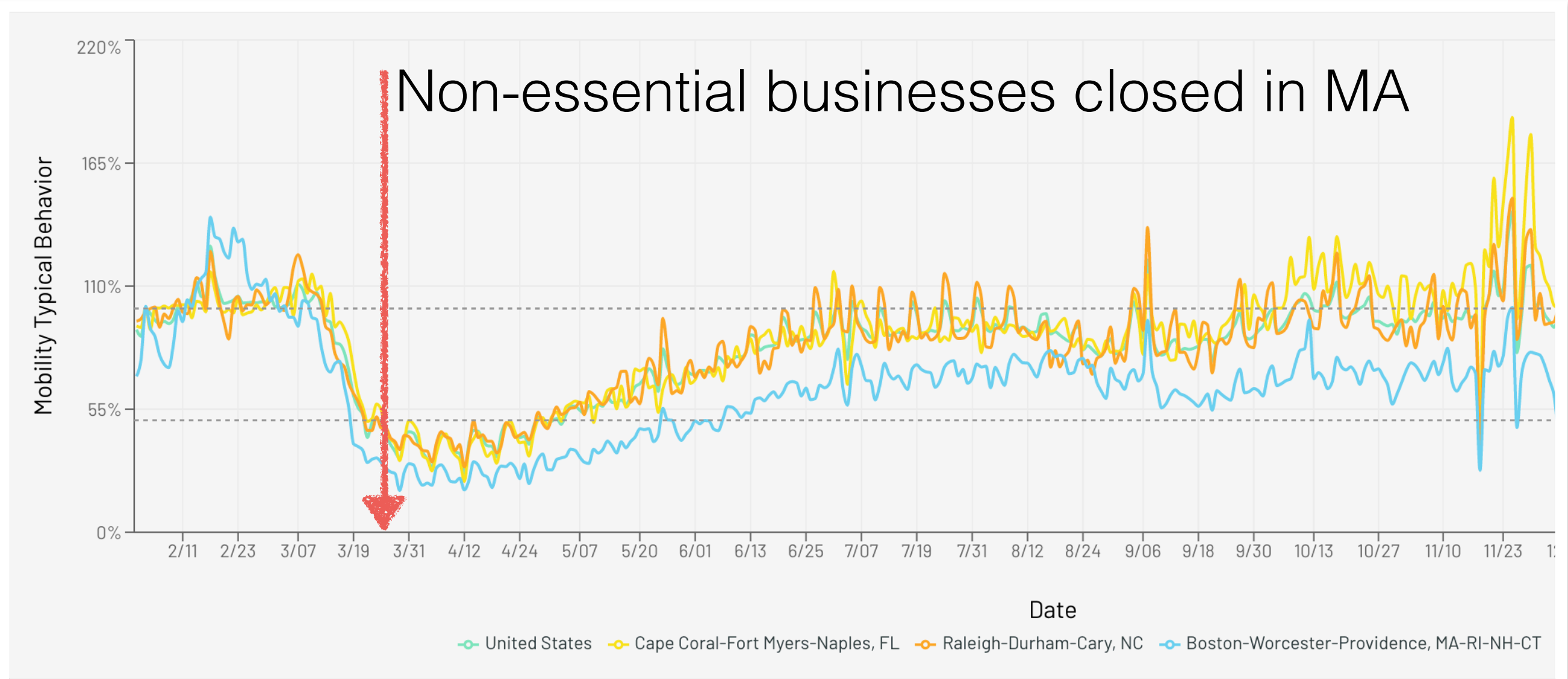
Vaccine attitudes may apply generally across societal groups. What differentiates them from others are their beliefs about vaccination effectiveness, their trust in public institutions, and whether they perceive enforced vaccination as a restriction on their freedom. We find that changing these beliefs is both possible and necessary to increase vaccine willingness, even in the case of mandates. An inference is that well-designed policies of persuasion and enforcement will be complementary, not alternatives.

on as

Katrin Schmelz^{a,b}

Edited by Margaret Levi, Stanford University, Stanford, CA; received October 12, 2021; accepted January 24, 2022

And with mobility



So how do we get to good data?



So how do we get to good data?





Process

People

Tech

Good data

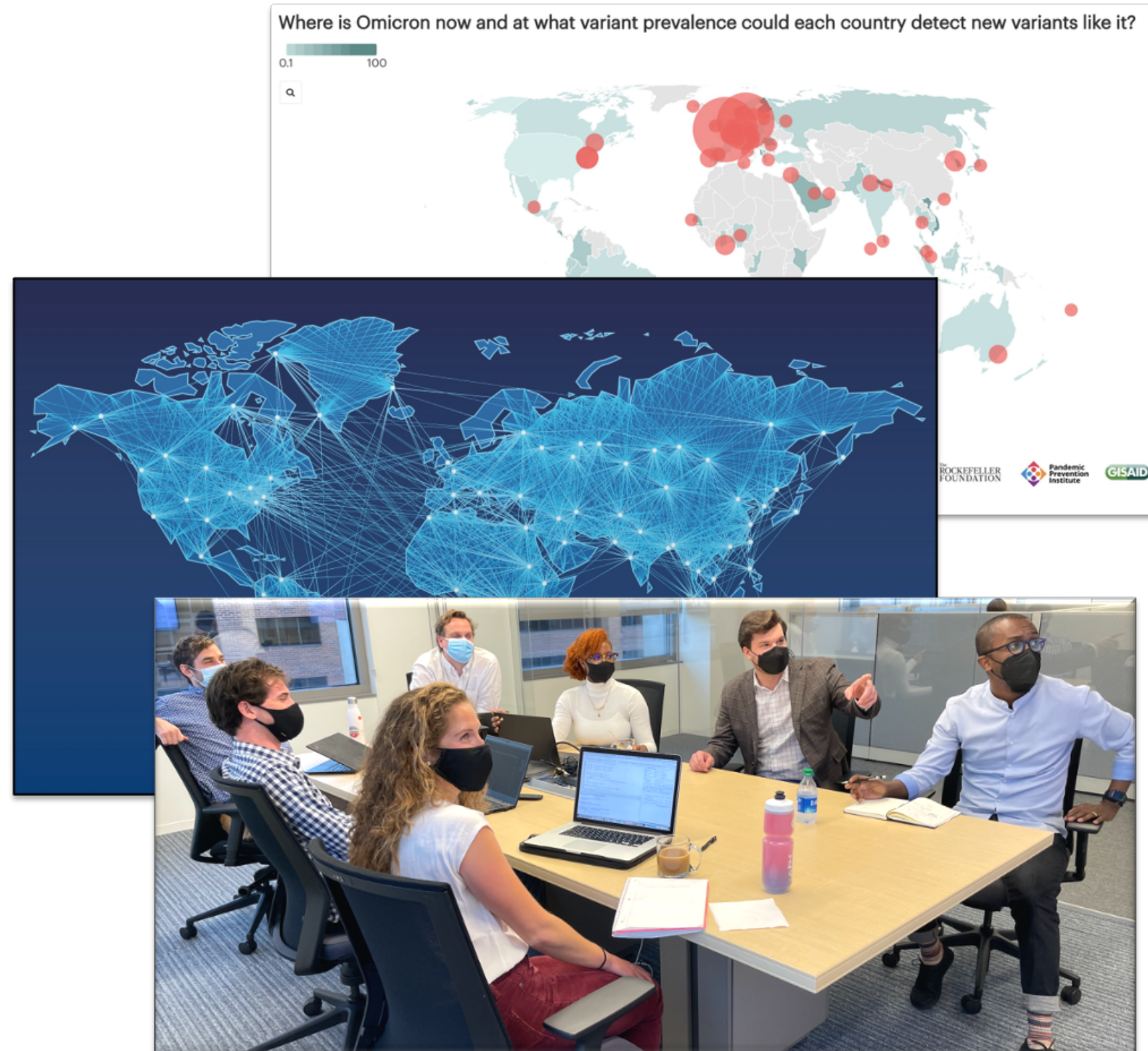


Process

People

Tech

How might this apply to pandemics prevention?

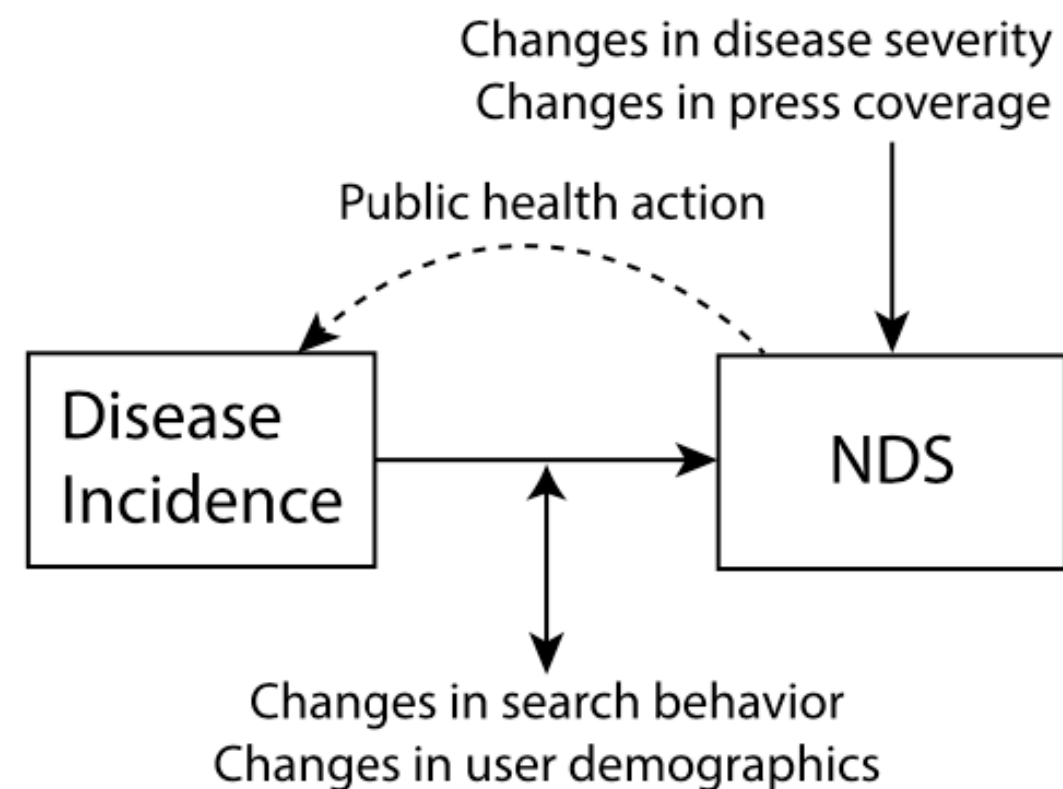




Enhancing disease surveillance with novel data streams: challenges and opportunities

Benjamin M Althouse^{1*†}, Samuel V Scarpino^{1*†}, Lauren Ancel Meyers^{1,2}, John W Ayers³, Marisa Bargsten⁴, Joan Baumbach⁴, John S Brownstein^{5,6,7}, Lauren Castro⁸, Hannah Clapham⁹, Derek AT Cummings⁹, Sara Del Valle⁸, Stephen Eubank¹⁰, Geoffrey Fairchild⁸, Lyn Finelli¹¹, Nicholas Generous⁸, Dylan George¹², David R Harper¹³, Laurent Hébert-Dufresne¹, Michael A Johansson¹⁴, Kevin Konty¹⁵, Marc Lipsitch¹⁶, Gabriel Milinovich¹⁷, Joseph D Miller¹⁸, Elaine O Nsoesie^{5,6}, Donald R Olson¹⁵, Michael Paul¹⁹, Philip M Polgreen²⁰, Reid Priedhorsky⁸, Jonathan M Read^{21,22}, Isabel Rodríguez-Barraquer⁹, Derek J Smith²³, Christian Stefansen²⁴, David L Swerdlow²⁵, Deborah Thompson⁴, Alessandro Vespignani²⁶ and Amy Wesolowski¹⁶

Figure 1 The link between public health problems and NDS is modified by user behavior (i.e., propensity to search, what terms are chosen to search, etc.), user demographics, external forces on user behavior (i.e., changing disease severity, changing press coverage, etc.), and finally by public health interventions, which by design aim to modify the public health problem creating feedback loops on the link to NDS.



Benjamin
Joan B
Sara D
David
Gabrie
Philip
Derek

Alessandro Vespignani²⁶ and Amy Wesolowski¹⁶

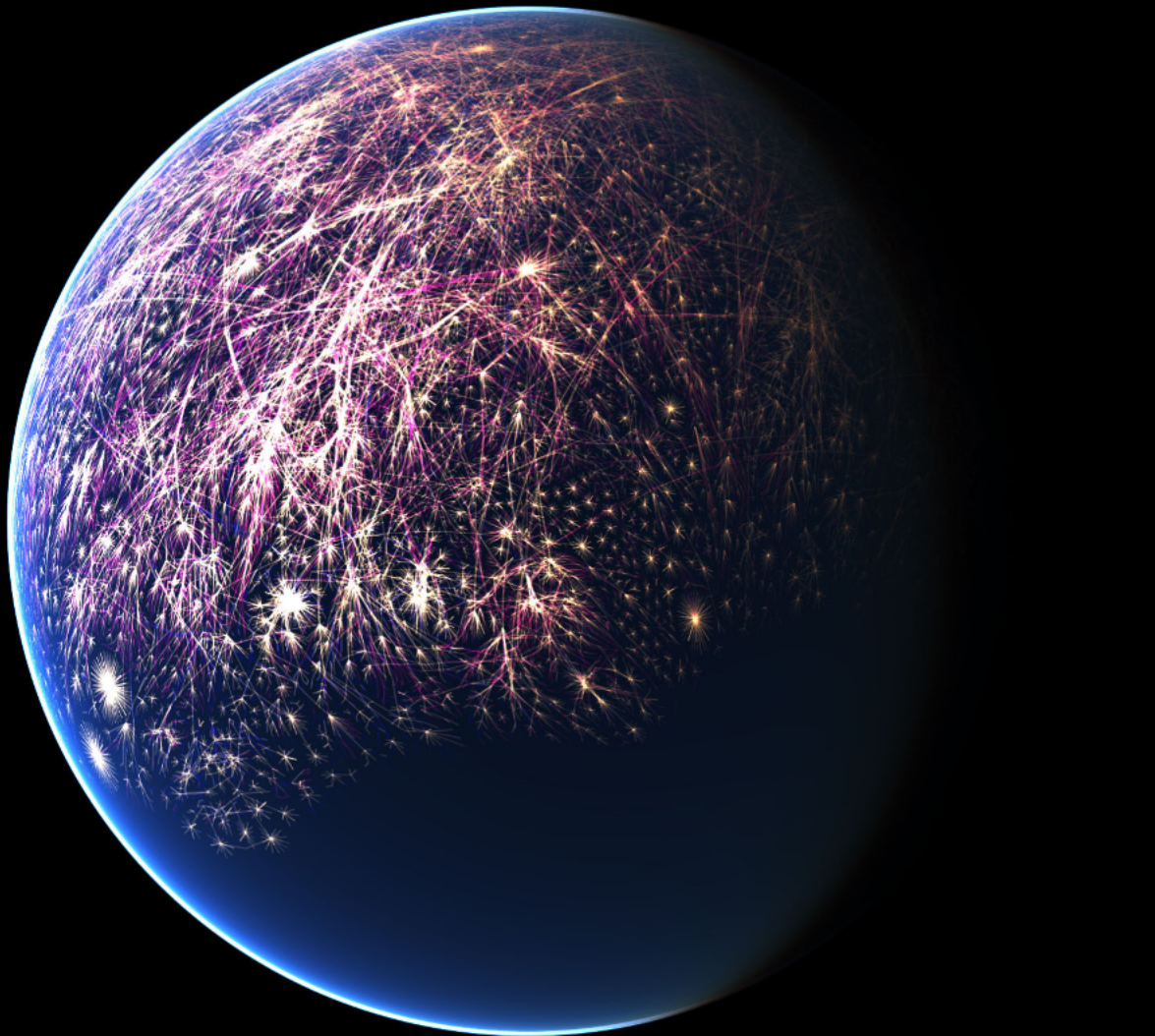
**We envision a world
where pandemics are
contained.**

We Believe in a Future Where Pandemics are Contained Because Information Empowers Action

We envision a future where...

- ➡ SEEN: Information is shared rapidly, equitably and transparently across the globe
- ➡ VERIFIED: A global early warning system ensures pathogens with pandemic potential are quickly detected and verified
- ➡ MITIGATED: Institutions and individuals alike are informed and empowered to act swiftly to contain outbreaks

...any outbreak is contained before it becomes established



Our goal is to **contain** outbreaks before they become established by building a distributed, global network



Empowering Decision-makers: from nations to neighborhoods



Individuals



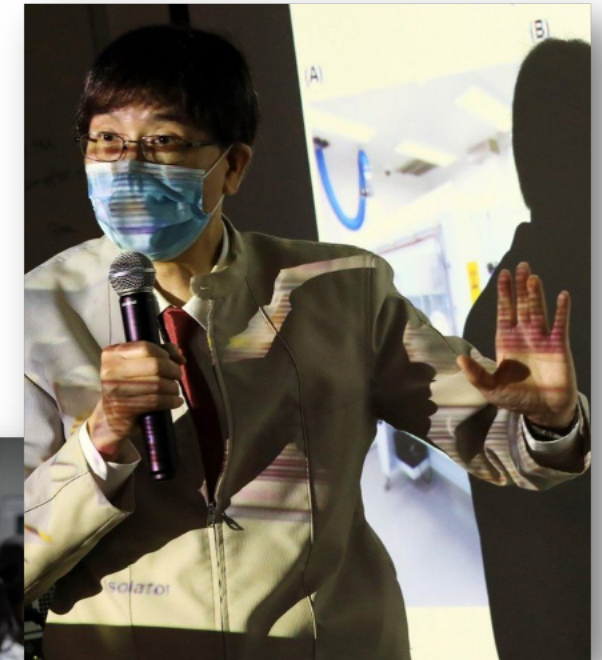
Government



***Business /
Private Sector***



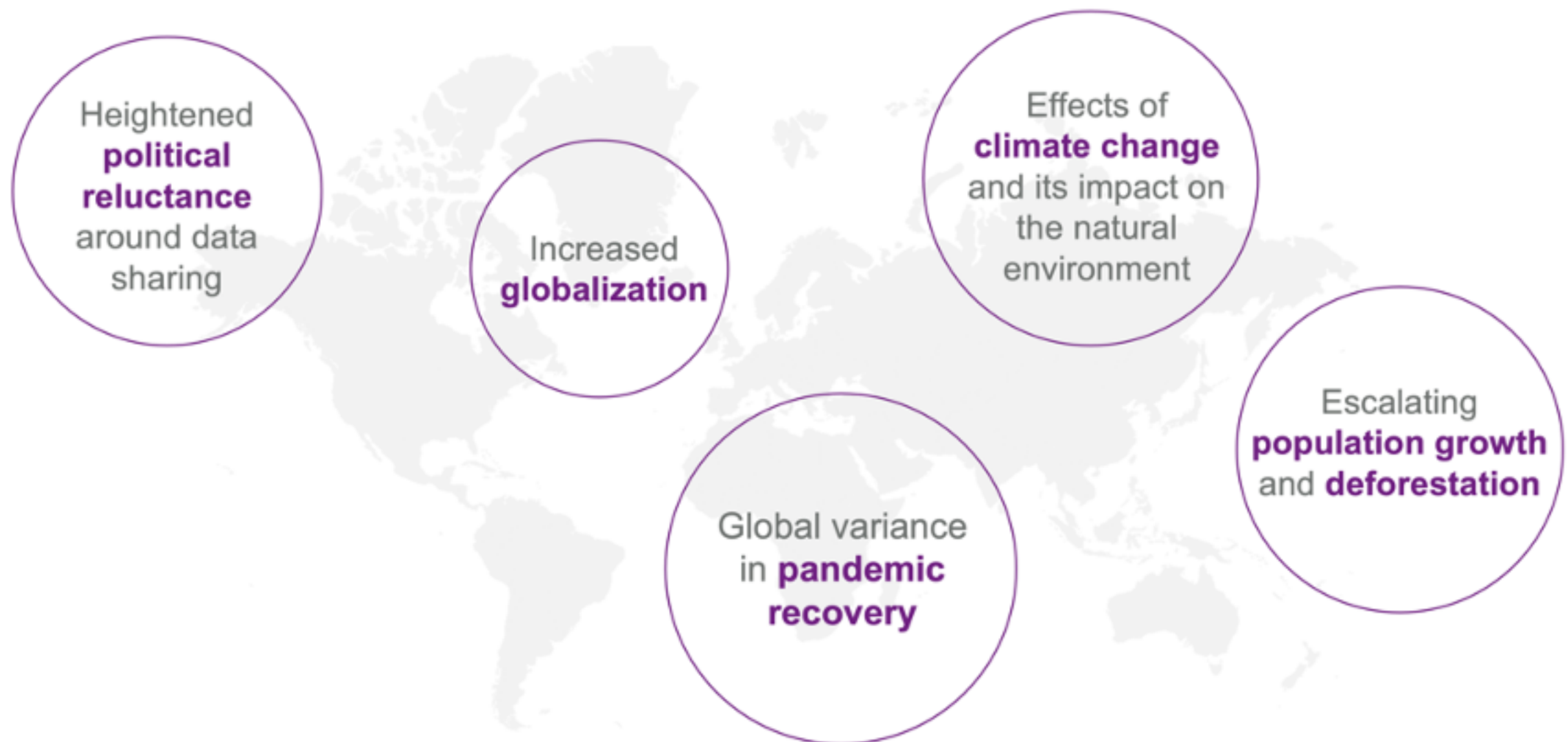
***Medical Countermeasure
Developers***



Academia

Pandemic-potential outbreaks will continue to threaten the world and the risk of future crises is only growing

The world faces an average of 200 epidemics each year. This is a direct result of population growth, human encroachment on animal habitats, globalization, and climate change.



Pandemic-potential outbreaks will continue to threaten the world and the risk of future crises is only growing

The world faces an average of 200 epidemics each year. This is a direct result of population growth, human encroachment on animal habitats, and climate change.

Heightened
**political
reluctance**
around data
sharing



The Washington Post @washingtonpost · Feb 1

Researchers are asking why some countries were better prepared for covid.

One surprising answer: **Trust.**



washingtonpost.com

Researchers are asking why some countries were better prepared for ...

A new study of pandemic preparedness across 177 countries and territories appears to show that trust in government plays a major role.

ge
on
t

Escalating
**population growth
and deforestation**

recovery

Never bring data to a story fight

-Prof. Peter Dodds



Questions?

Samuel V. Scarpino, PhD

Vice President, Pathogen Surveillance

The Rockefeller Foundation & the Pandemic Prevention Institute

External Faculty

Santa Fe Institute

Vermont Complex Systems Center

Affiliate Assistant Professor

Northeastern University

Co-Founder, Global.health

@svscarpino

scarpino.github.io

sscarpin@rockfound.org



**SANTA FE
INSTITUTE**



**Pandemic
Prevention
Institute**



The
**ROCKEFELLER
FOUNDATION**

