

Fogarty International Center

Search queries and social media for surveillance of influenza and COVID-19

Cécile Viboud

Division of International Epidemiology and Population Studies

CIW conference, Nov 2nd 2020



Fogarty International Center

Fogarty at
50 

Digital surveillance for influenza and emerging outbreaks (pre-COVID-19)

March 2014

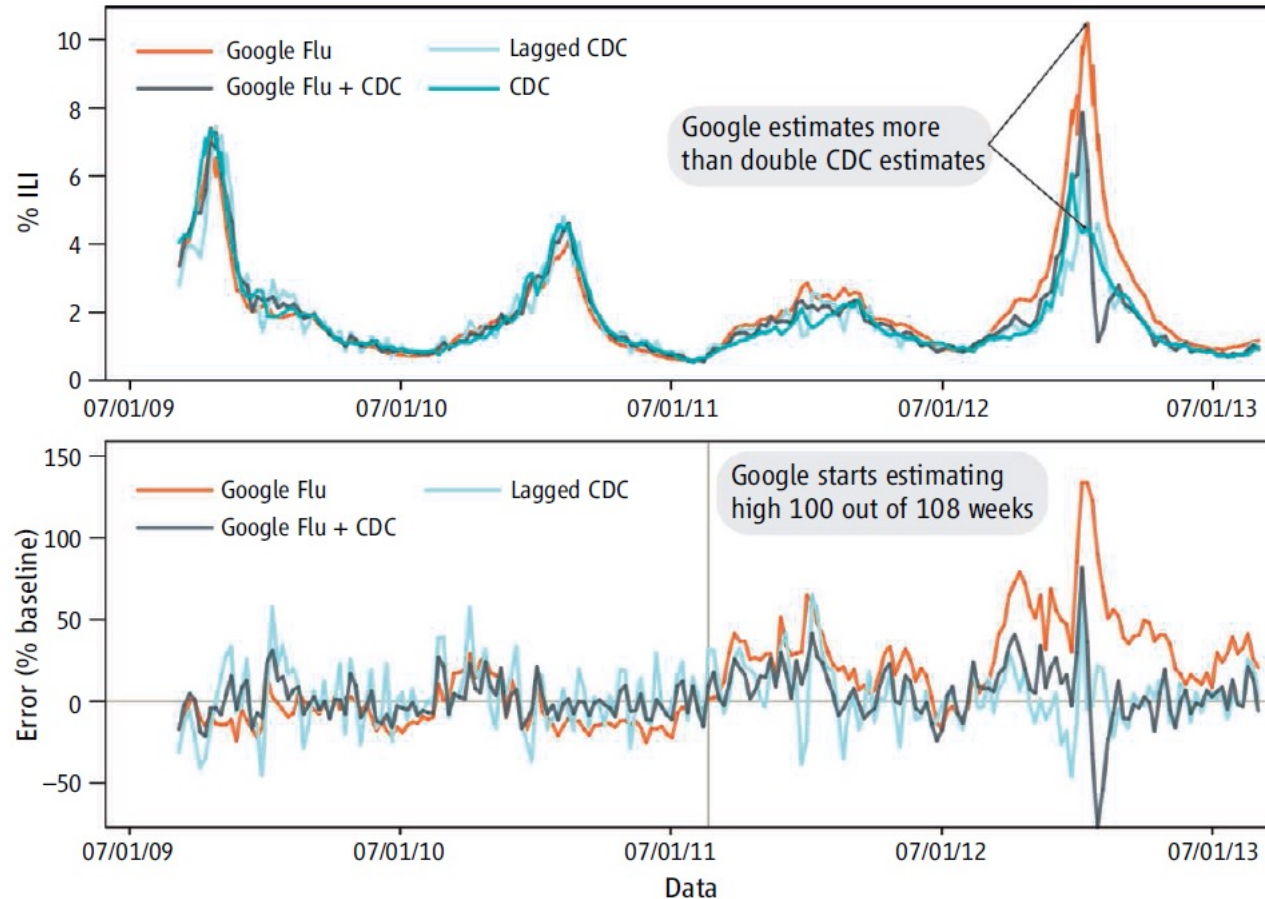
BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

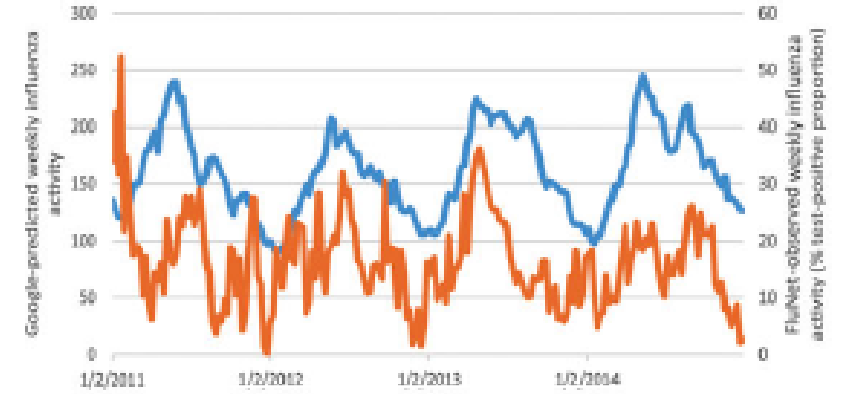
David Lazer,^{1,2*} Ryan Kennedy,^{1,3,4} Gary King,³ Alessandro Vespignani^{5,6,3}

The pitfalls of Big Data hubris

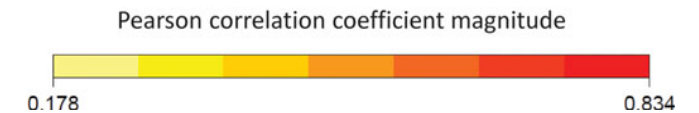
Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.



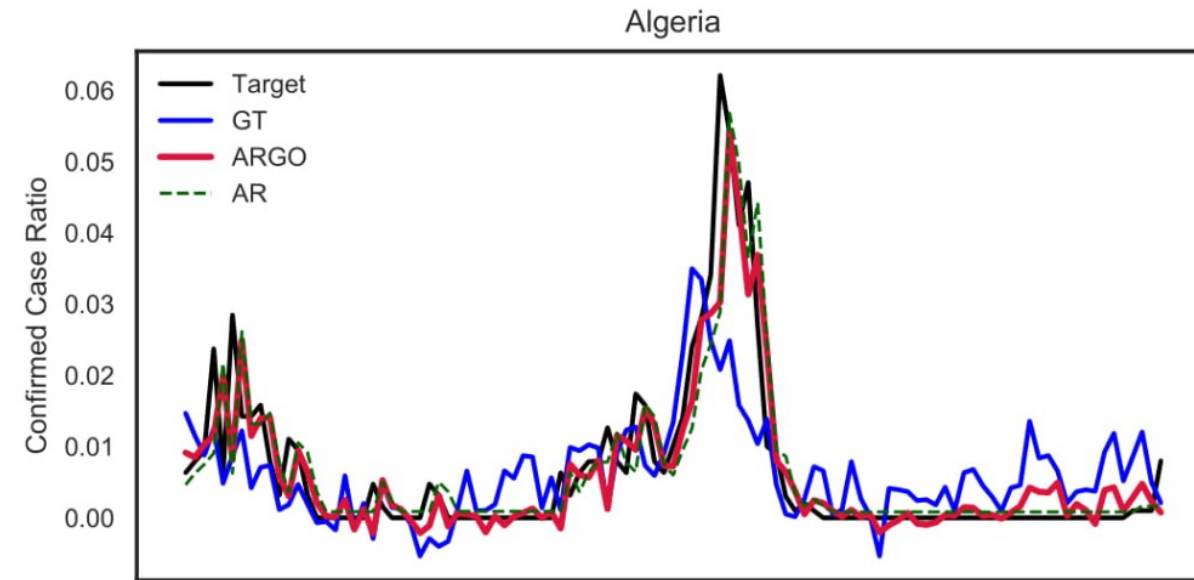
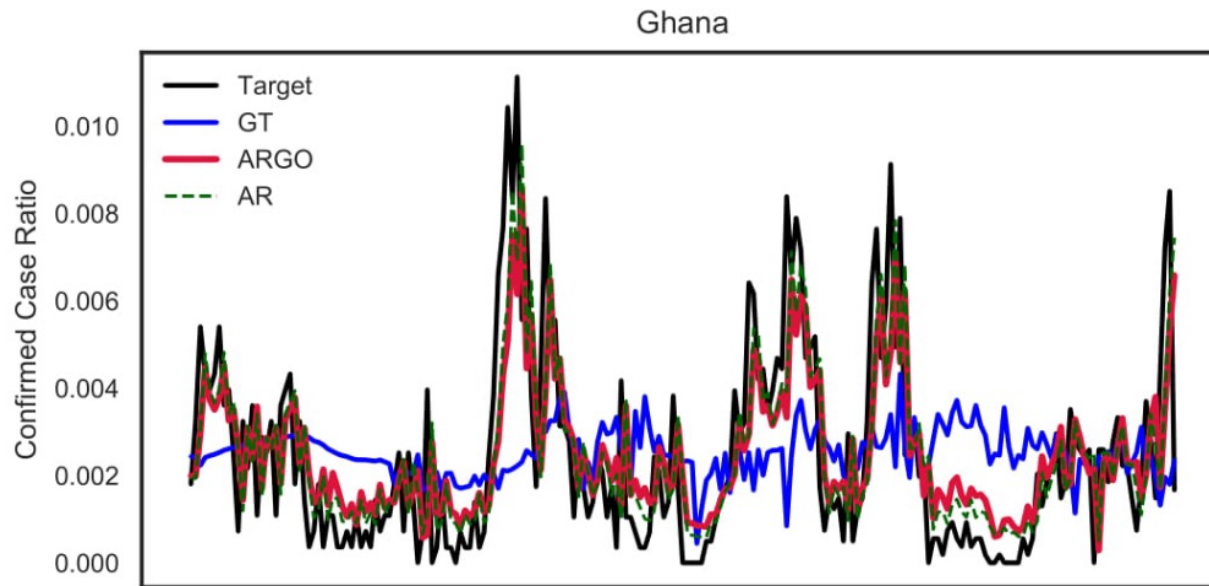
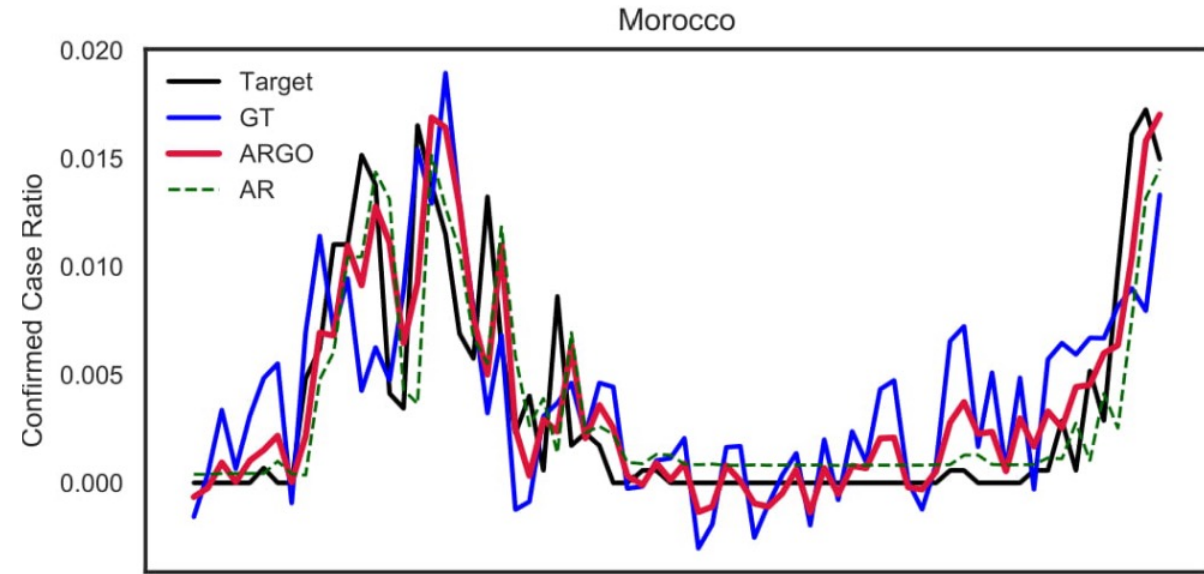
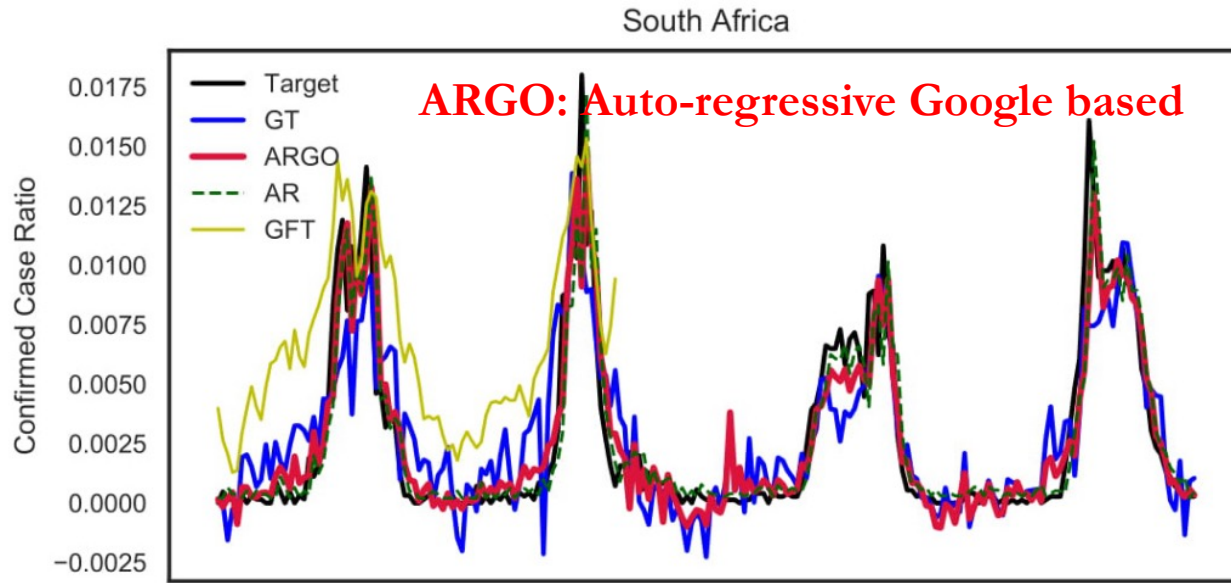
C Brazil



Predicted
Observed



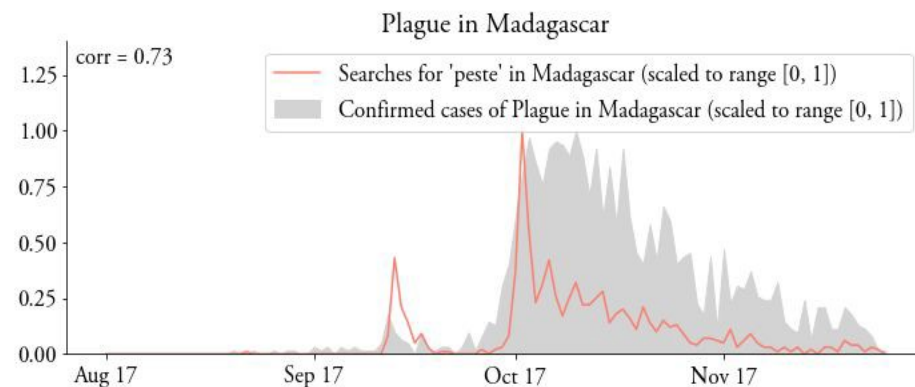
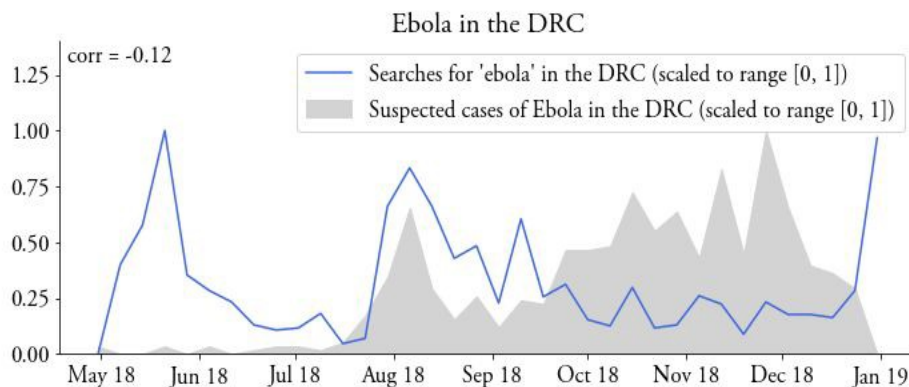
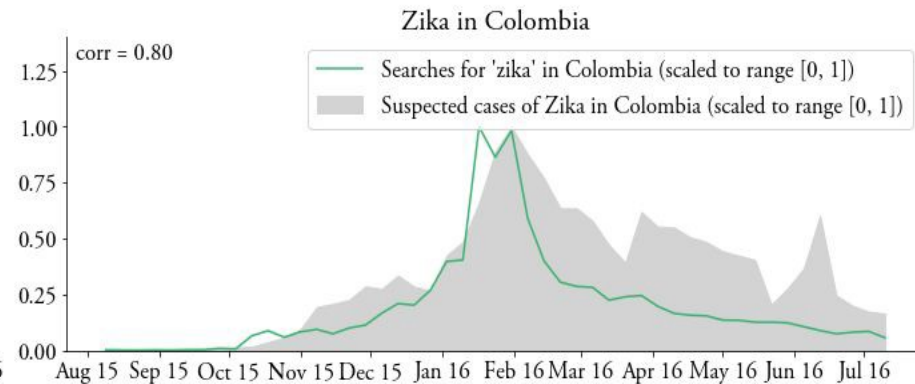
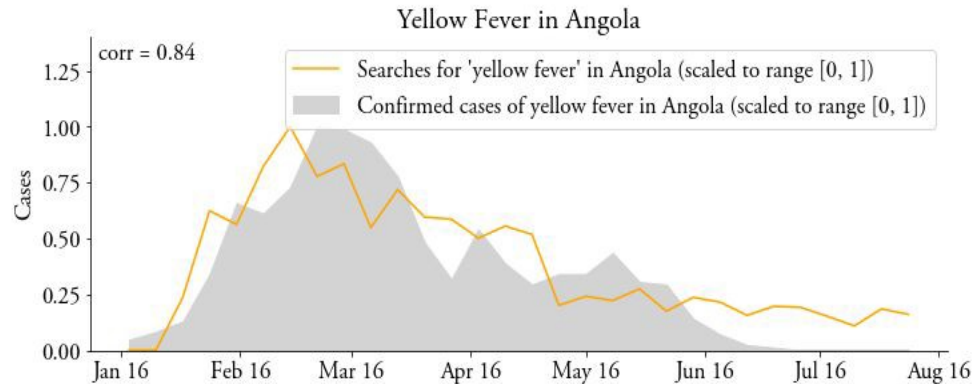
Google-search based surveillance for influenza in South Africa



What makes a population conducive to digital surveillance?

	Characteristics	South Africa	Algeria	Morocco	Ghana
Disease	ARGO correlation	0.91	0.87	0.84	0.70
	Median yearly case count	7,388	632	1,337	3,207
	Seasonality	Yes	Yes	Yes	No
Population	Internet penetration	54%	43%	58%	35%
	Google market share	95%	97%	97%	89%
	Literacy rate	99%	80%	69%	77%
	Average population (millions)	57.5	42.2	35.7	28.3
	Country size (10 ³ mi ²)	471	919.6	274.5	92.5
	Population (millions / mi ²)	47	15.9	50	262.9
	GDP per capita	\$13,840	\$4,669	\$8,959	\$2,081
	Poverty headcount ratio	56%	6%	15%	24%
	Conflict zone	No	Yes	No	No

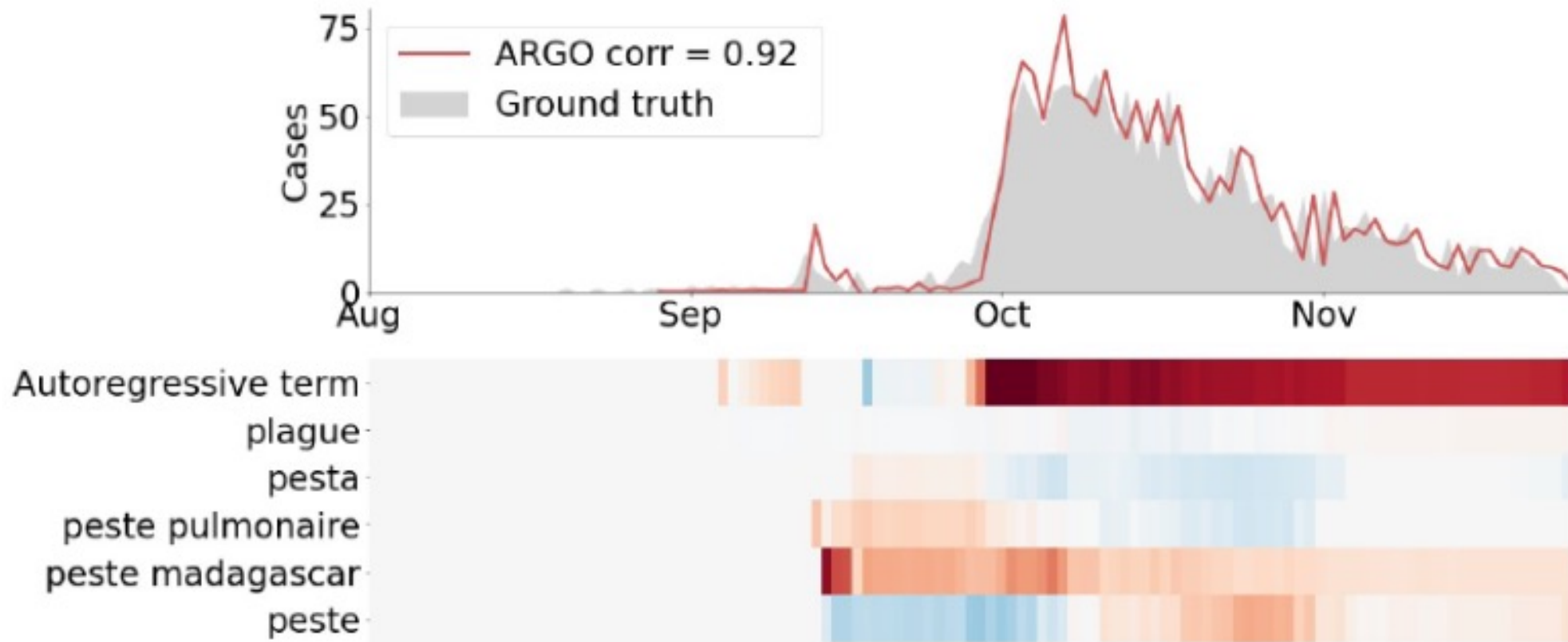
Digital epidemiological models for emerging infections



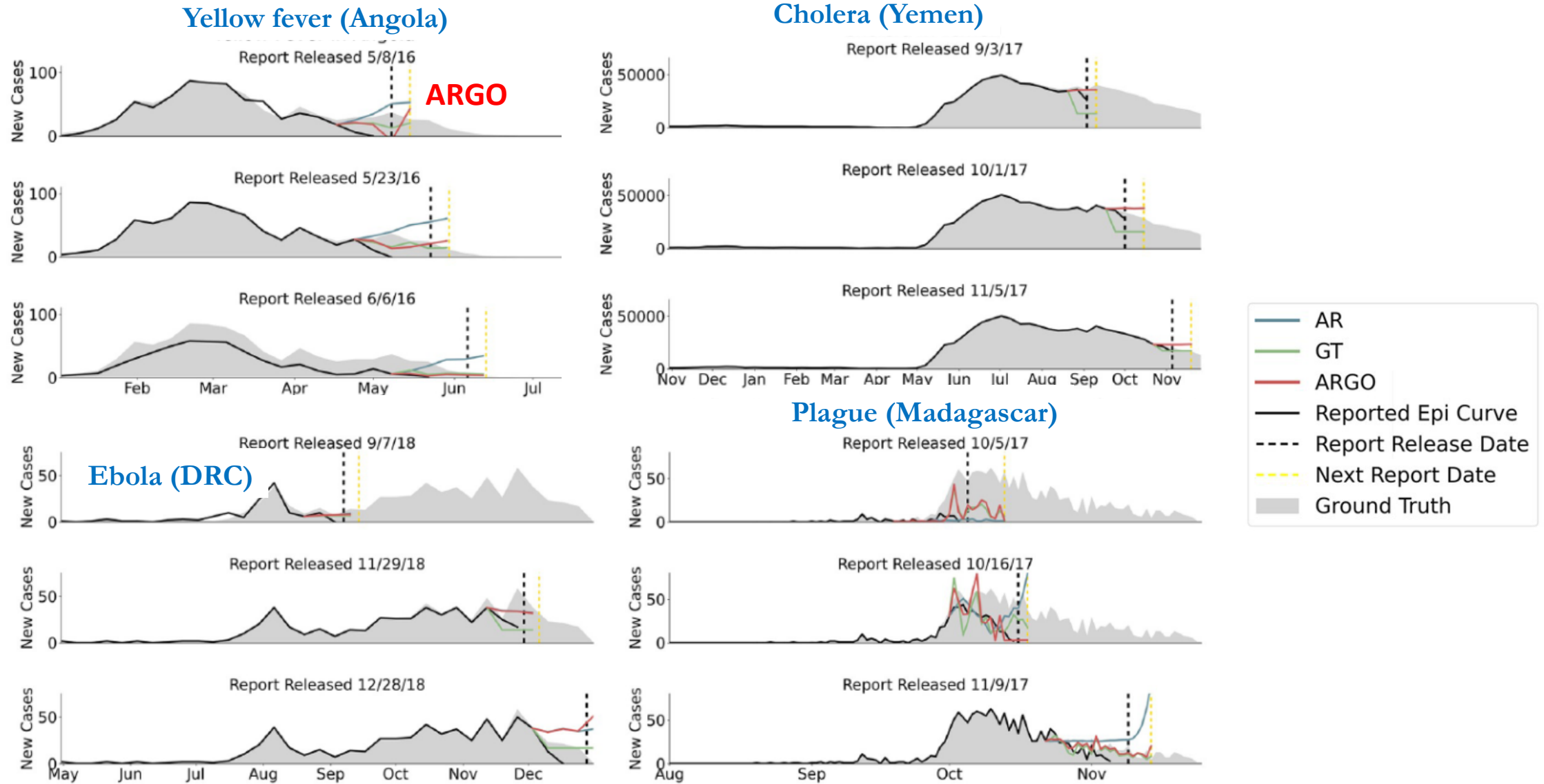
- Limited ground-truth surveillance data
- Reporting issues expected
- Coverage and use of Google searches
- Diversity of languages
- Search terms not necessarily well-established

E. Aiken, S F. et al. *Real-time Estimation of Disease Activity in Emerging Outbreaks using Internet Search Information*. 2020. *PLoS Comp Bio*.

ARGO model for plague in Madagascar (2017)



Four case studies of ARGO predictions for emerging infections



Digital surveillance and the COVID-19 pandemic

Real-time broadcast

Case data scraping from social media DXY.cn

2 hours ago
2-04 10:22**Lates 2 new cases in Gansu, cumulative 57 cases**

As of 20:00 on February 4, two new cases of pneumonia confirmed by a new coronavirus infection in Gansu. The newly confirmed cases were in Lanzhou. In Gansu, 57 confirmed cases of pneumonitis with new coronavirus infection were reported. 4 were discharged, and 53 were treated in isolation at designated hospitals. At present, a total of 1,784 close contacts are receiving medical observations.



People's Network 🏆

4 hours ago from Weibo weibo.com

3 hours ago
2-04 08:51**4 new cases in Liaoning, a total of 81 cases**

From 4 pm to 21:30 on February 4, 2020, four new cases of pneumonia confirmed by a new coronavirus infection in Liaoning. On February 4, 2020, at 21:30, Liaoning Province had reported 60 imported cases and 19 local infections. 60 have been lifted, and 1109 people are currently receiving medical observations.

4 hours ago
2-04 07:44**1 new case in Tianjin, a total of 67 cases**

It was learned from the Tianjin Centers for Disease Control and Prevention that a new case of pneumonia confirmed by a new coronavirus infection was confirmed in Tianjin, with a total of 67 confirmed cases, including 37 males and 30 females; 4 critically ill, 17 severely ill, and 1109 people are currently receiving medical observations.

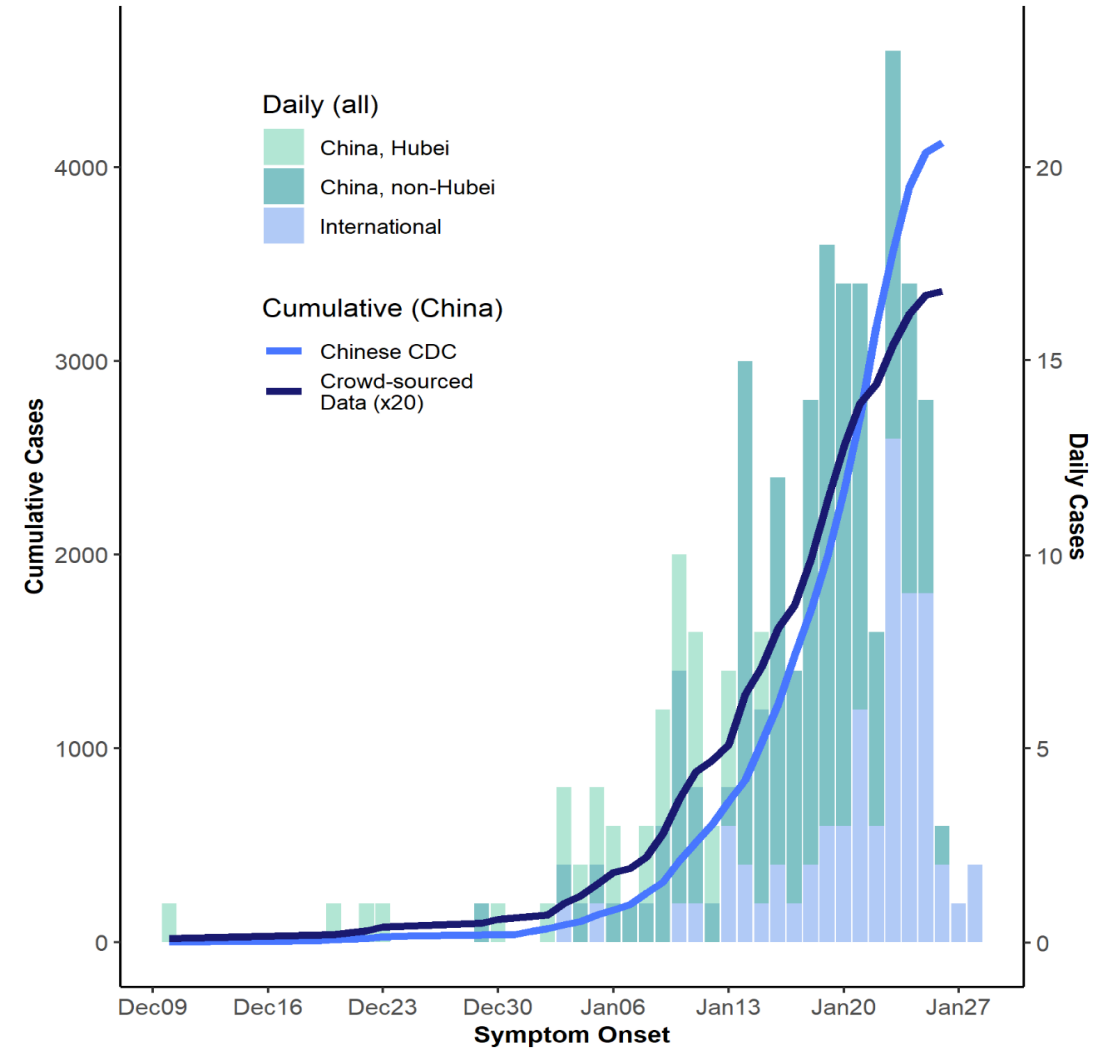
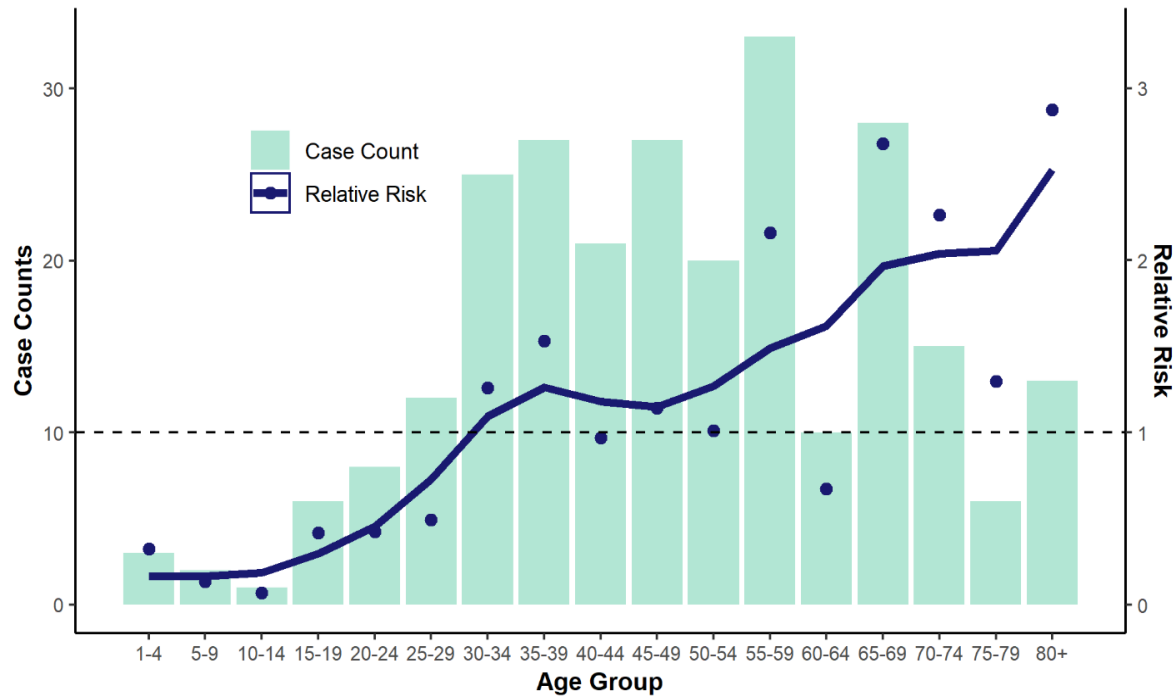
新冠肺炎更新动态 # [Tianjin newly added 1 case with a total of 67 cases] It was learned from the Tianjin Centers for Disease Control and Prevention that as of 18:00 on February 4, 1 new case of new coronavirus pneumonia was confirmed in Tianjin, and a total of confirmed cases were confirmed 67 cases.

The 67th patient, female, 33 years old, residing in Jinnan District, Tianjin, drove to the Chinese New Year's Eve in Chibi, Hubei with her husband and two children on January 20, consciously developed fever, took medication on her own, and did not visit a medical institution. I returned to Tianjin on January 31, and fever reappeared on February 2 and took the medicine on my own. On February 3, his husband drove him to the First Affiliated Hospital of Tianjin University of Traditional Chinese Medicine. On February 4th, the city expert group confirmed the 67th case in our city, which is a common type. He has been transferred to Haihe Hospital for treatment, and his vital signs are stable. At present, five close contacts have been determined, and other close contacts are continuing to be tracked down, and their homes are being disinfected at the end.

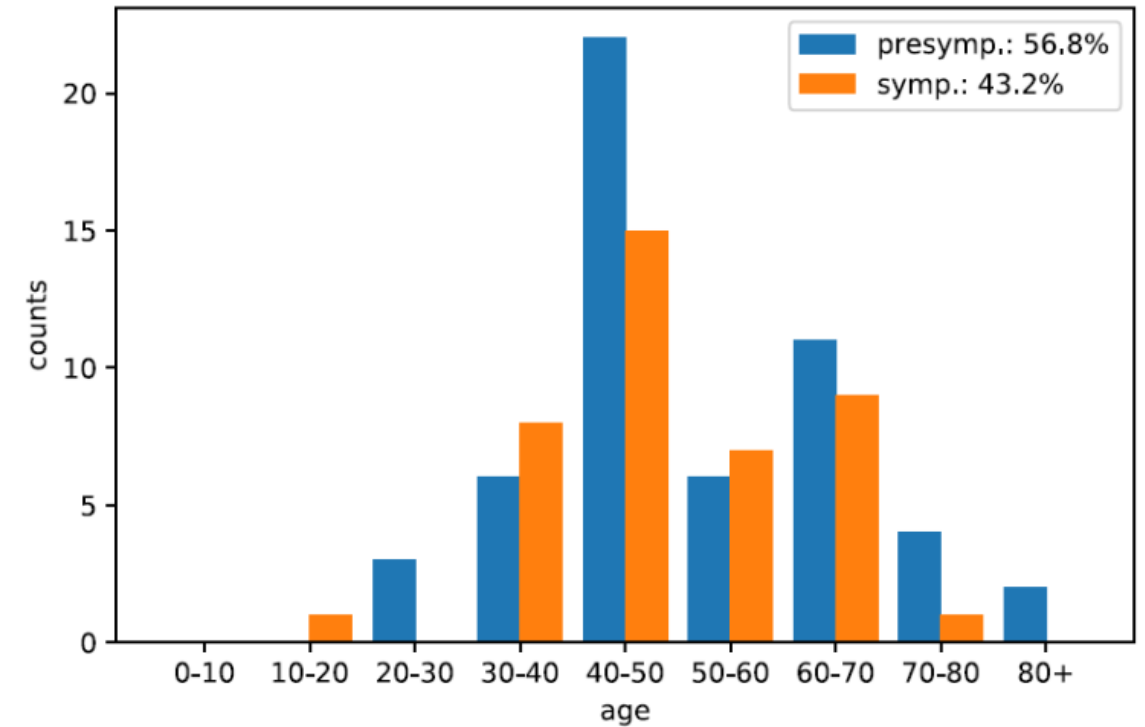
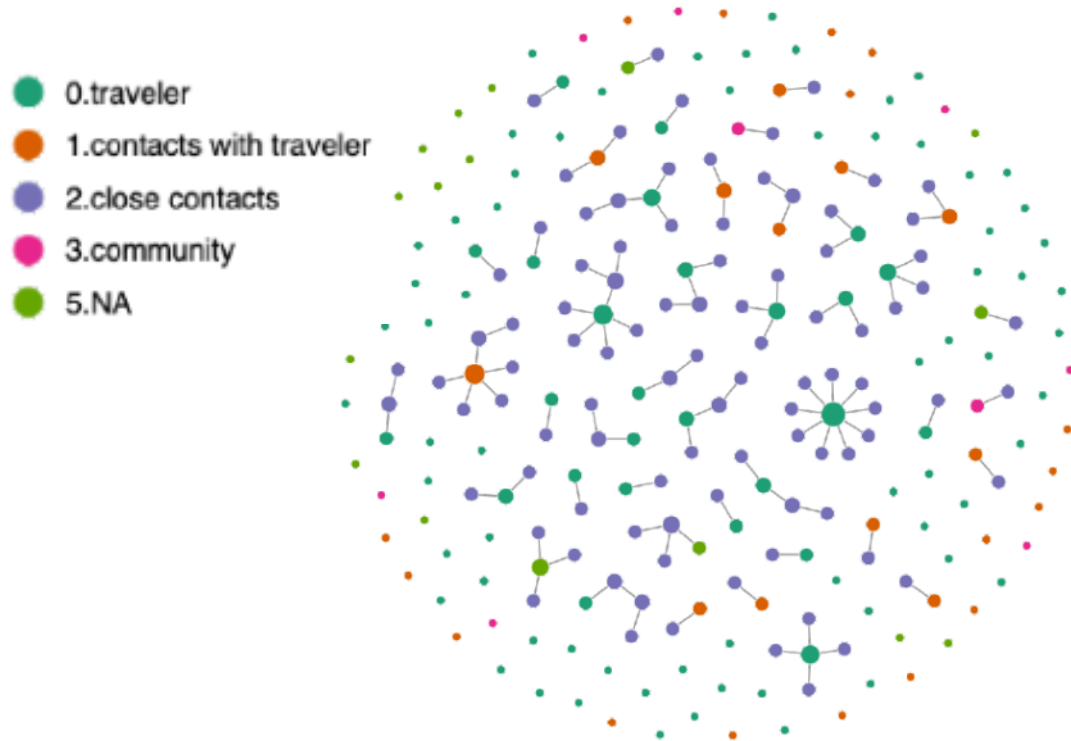
5 hours ago
2-04 07:26**The central government made clear: to accelerate the construction of a group of hotels and other hospitals for treating patients with pneumonia**

On the 4th, the Central Leading Group for New Coronavirus pneumonia treatment approved emergency cabin hospitals, and local governments should take over the system to take over the intensive care hospitals. The system should be consolidated, and companies such as medical protective clothing and masks should be accelerated to resume work and expand production capacity.

Age profile and progression of the outbreak in China based on news reports



Reconstructing transmission chains from news reports (Shaanxi)

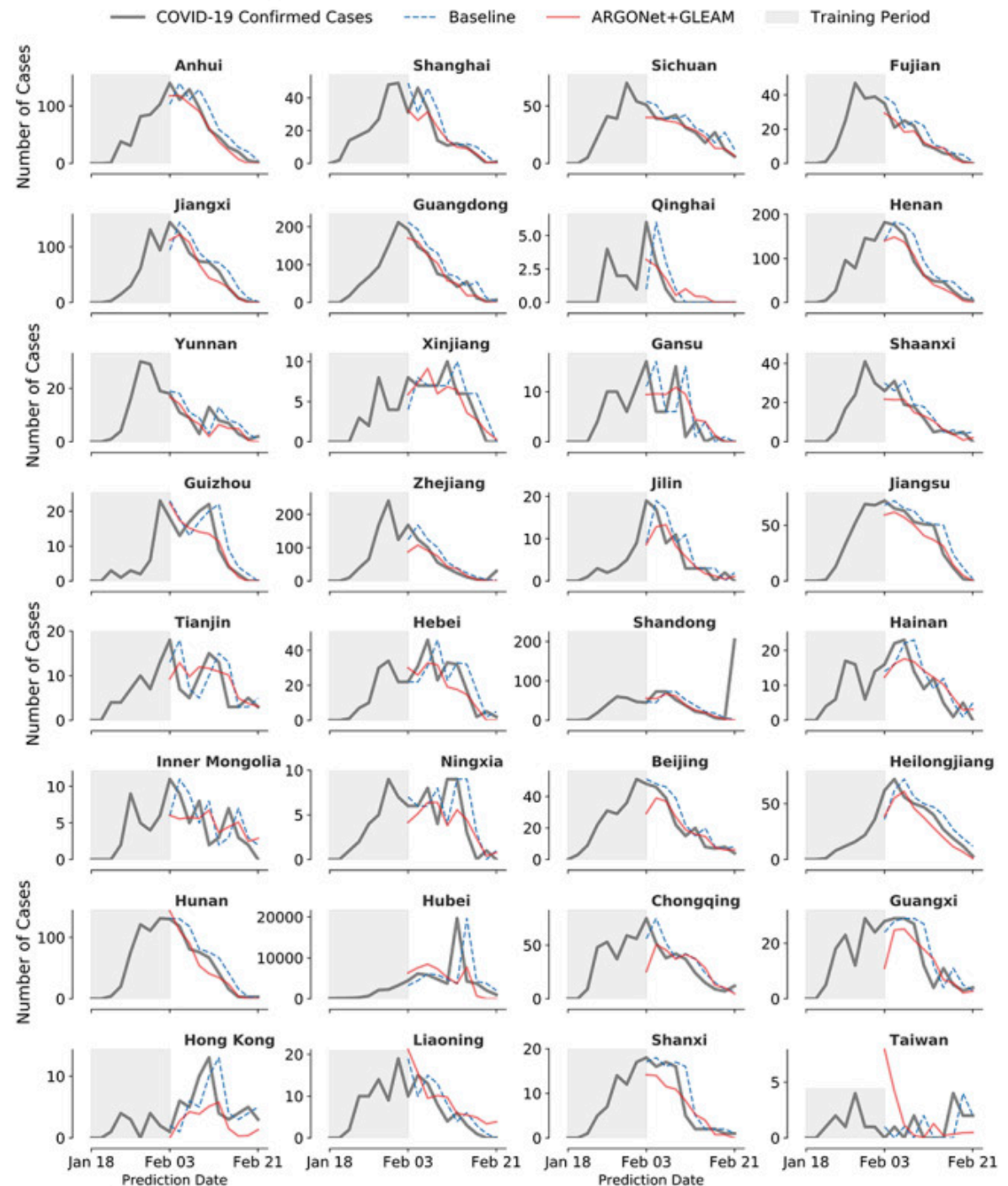


Intense contact tracing: 77 contacts/cases; mean onset to isolation 3 dys.
Overall $R=0.3$, highly dispersed

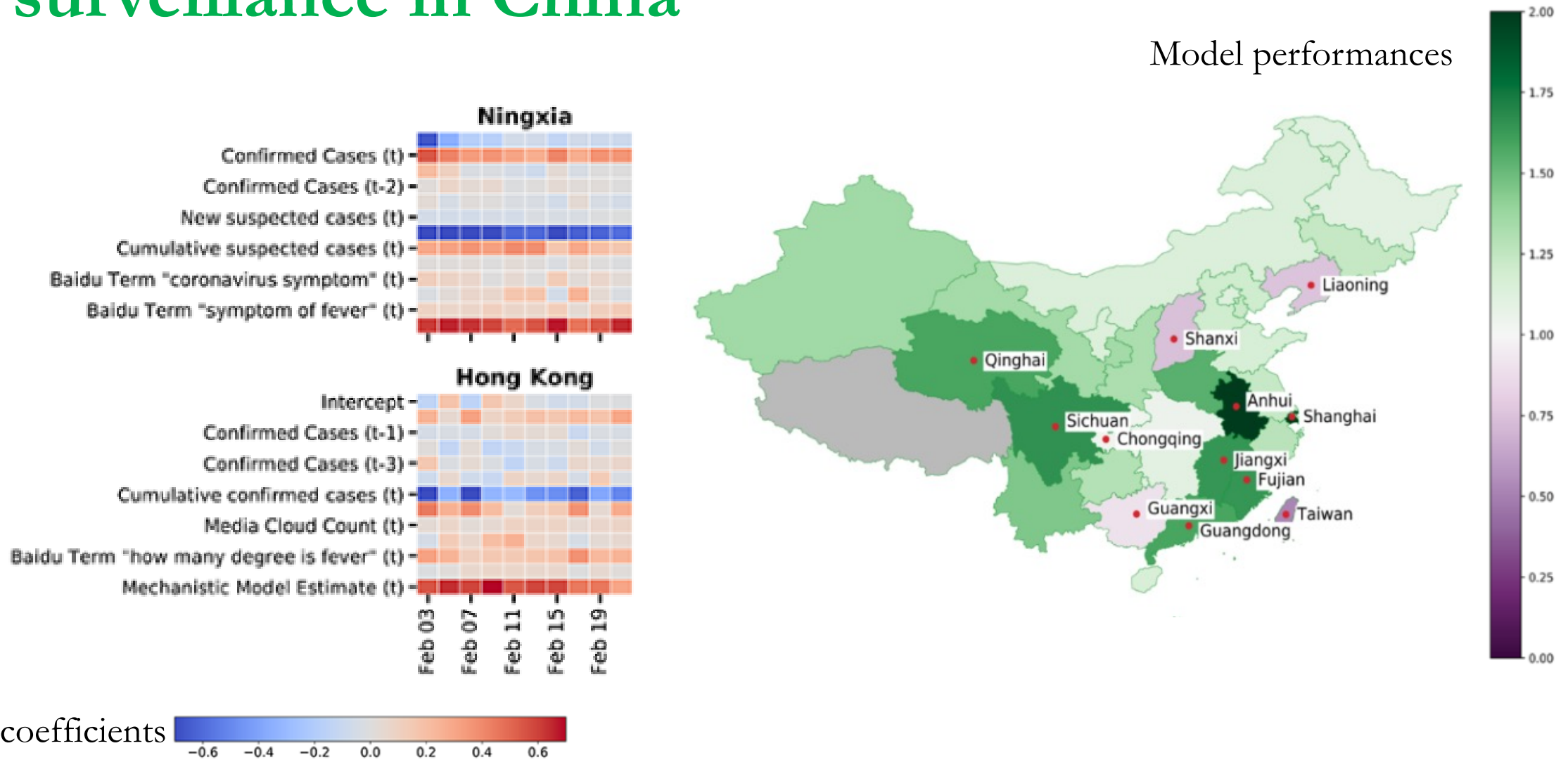
Digital surveillance of COVID-19 in China

Regression approach (Lasso ARGO) to predict cases 2-days ahead:

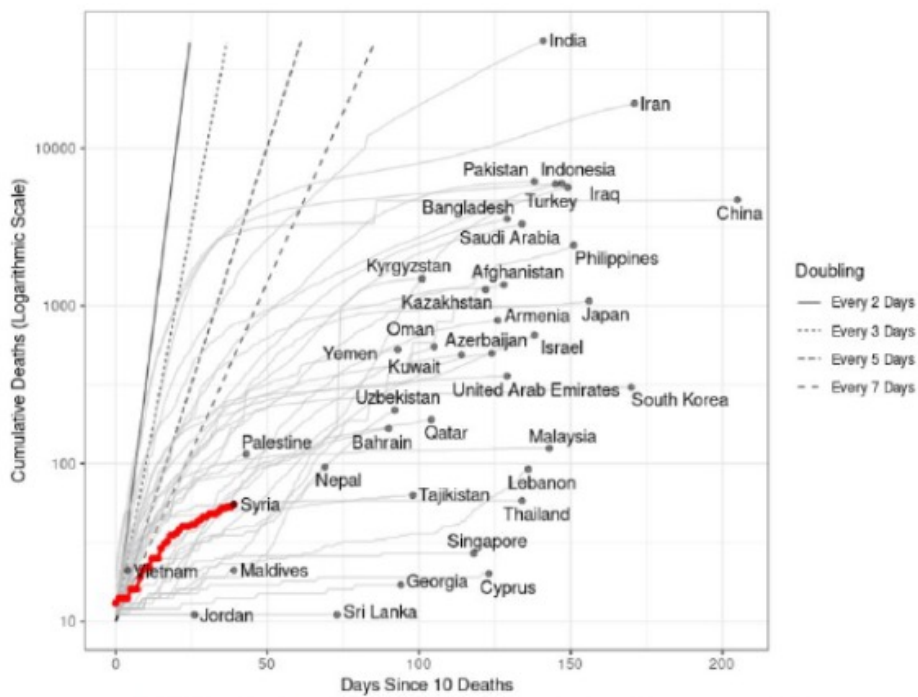
- Laboratory-confirmed COVID19 cases
- Baidu searches for COVID-related terms and symptoms
- Online news reports
- Incidence estimates from a mechanistic transmission model
- Spatial interactions between provinces
- January calibration period.



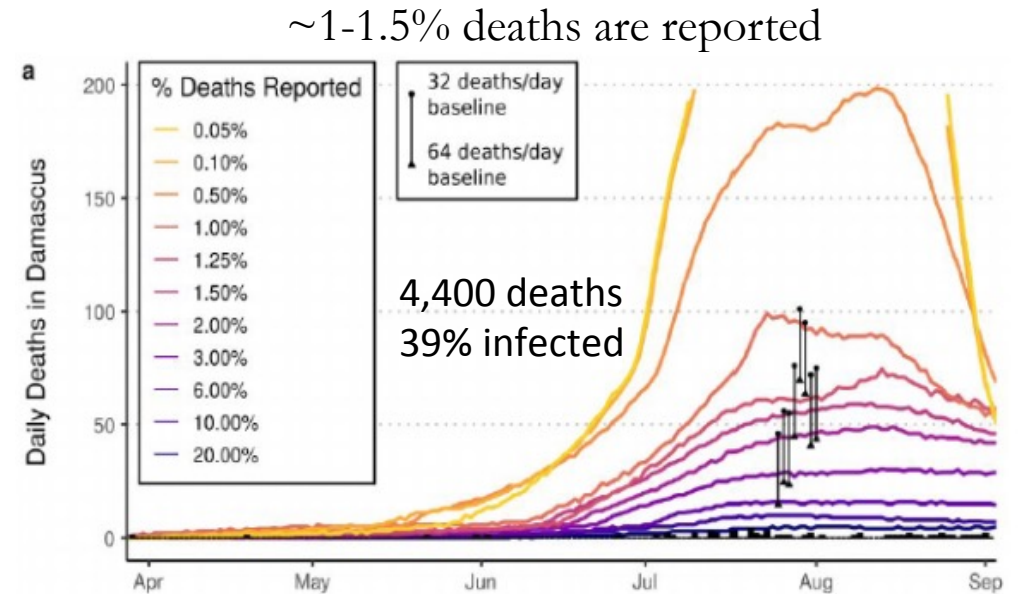
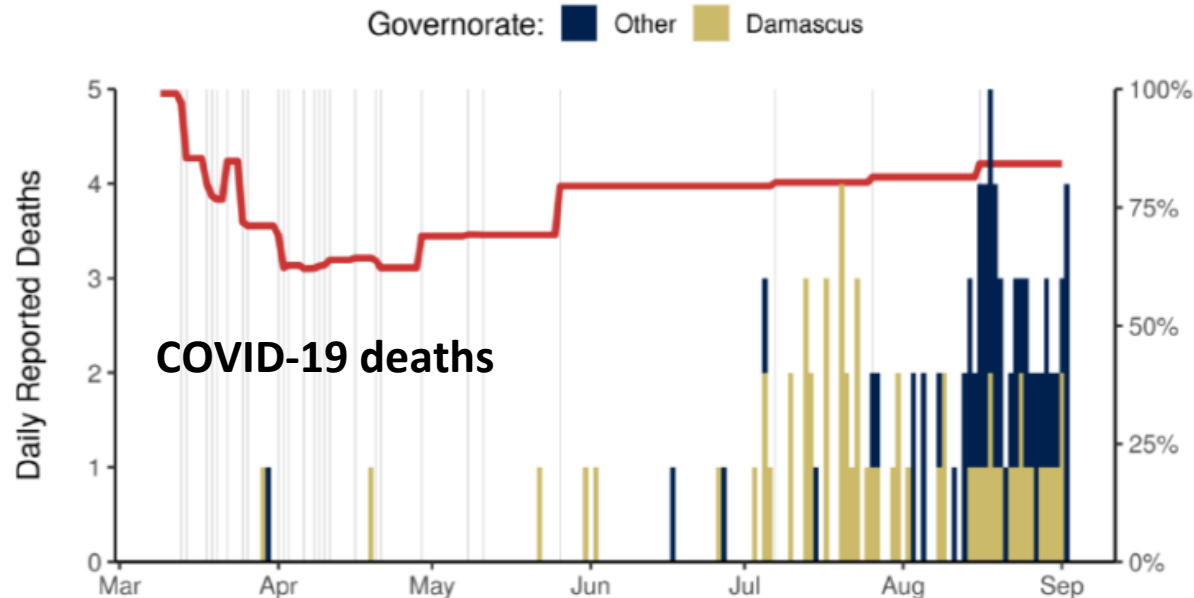
Predictors and performances of COVID-19 surveillance in China



COVID-19 dynamics and mortality in Damascus, Syria



- First death reported Mar 29th
- Mar-Apr curfews; interventions relaxed since May
- Only 120 COVID-19 confirmed deaths reported by Sep, 60 in Damascus
- Hospitals at capacity by late July/early Aug
- What is the true size of the epidemic?

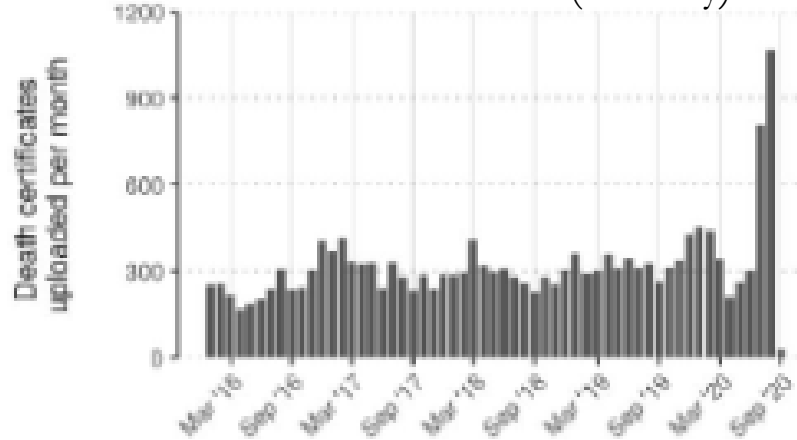


Calibrating excess mortality against Facebook obituaries in Damascus

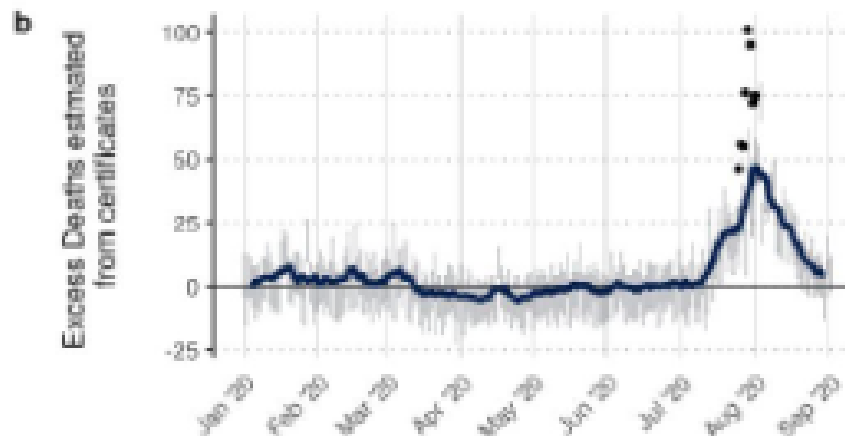


Printed obituary
(now on Facebook)

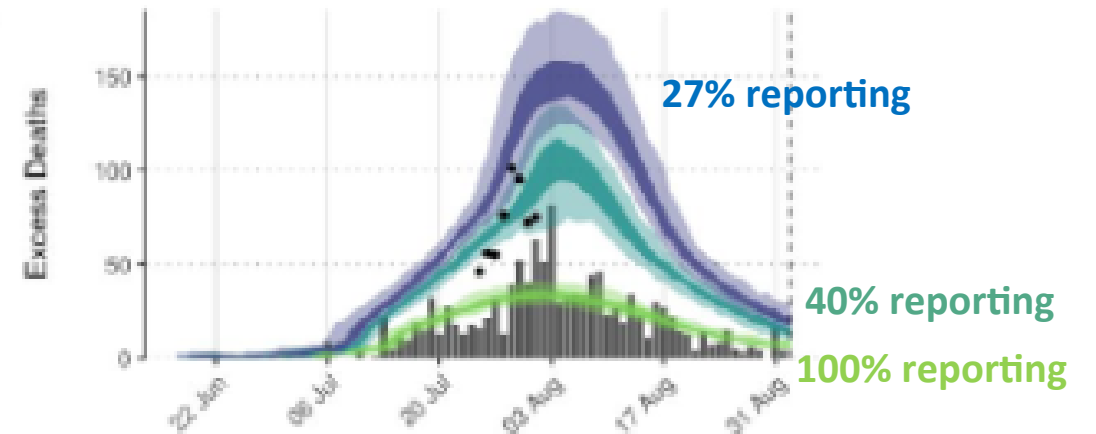
Facebook obituaries (monthly)



Facebook excess deaths (daily)



3,820 COVID-19 deaths
35% infected



Watson et al. *MedRxiv*. 2020

Conclusions

- Use of social media and digital data streams can be useful for short-term forecasts:
 - Lags in surveillance reports
 - Backfilling
- Ground-truth data essential
- More case studies needed, especially in LMIC to assess:
 - Population characteristics
 - Coverage of social media
 - Disparities (age, socio-economic status, languages, etc)
 - Disease characteristics
 - Seasonality and historic patterns
 - Specificity of symptoms and search queries
 - Rare/common
 - Transmission mode