Estimating the early death toll of COVID-19 in the United States

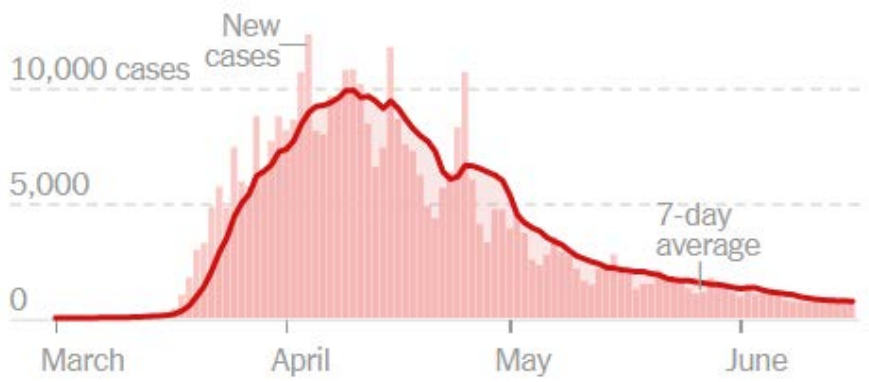
Dan Weinberger, Jenny Chen, Ted Cohen, Forrest W. Crawford, Farzad Mostashari, Don Olson, Virginia E Pitzer, Nicholas G Reich, Marcus Russi, Lone Simonsen,
Anne Watkins, Cecile Viboud

Yale SCHOOL OF PUBLIC HEALTH

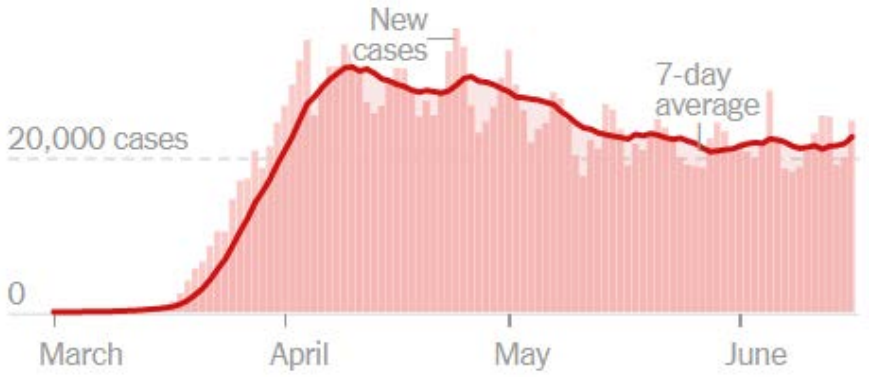
COVID-19 in the United States

- First reported death early February
- Explosive growth in the Northeast US
- Slower growth elsewhere, with recent upticks
- Testing was slow to start and geographically uneven

New York



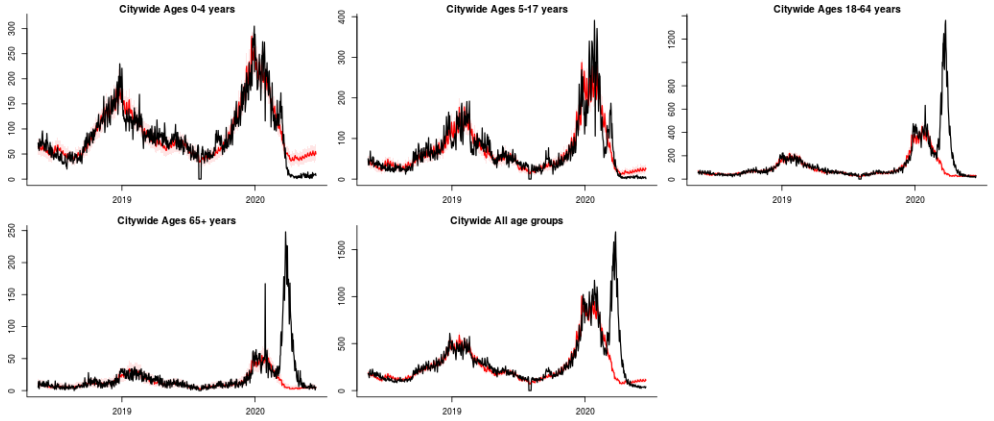
Entire US



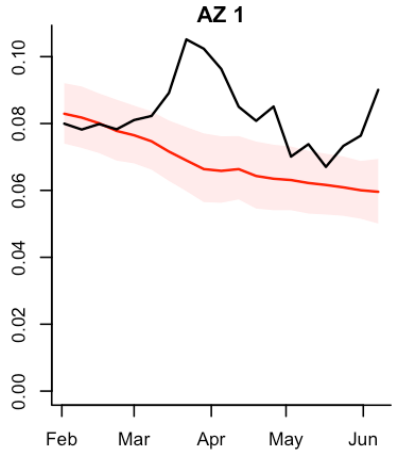
NYTimes

Tracking the epidemic in the absence of reliable virological data

ED visits for influenza-like illness



Ambulance calls for breathing difficulty



-Provides more timely data that 'official' statistics that rely on laboratory confirmation
- But can itself be biased by changes in behaviors

Tracking deaths

- Counting deaths might be less biased than counting cases
- Coding/diagnosis might be more reliable for deaths
- BUT…
  - Criteria for defining a death as COVID-19 related varies by area
  - Lack of testing still likely an impediment
  - Coding for cause of death takes a long time
- So: track changes in deaths due to <u>any cause</u> or due to broad outcomes like <u>pneumonia</u>

Cause of death statistics in the US

- Cause of death coded by physician/nurse (natural causes) or by medical examiner/coroner
- Typically assign an *underlying* cause and a *contributing* cause
  - Coronavirus as underlying cause; pneumonia, cardiac arrest as contributing causes
  - Assignment of cause of death is an art
- CDC recommends recording confirmed *and suspected* COVID deaths; but adherence varies by state
- Data registered by the state, reported to the CDC/NCHS

The politicization of death statistics

The New York Times

Opinion

How Many Covid Deaths? Don't Ask President Trump

The president has expressed skepticism about the fatality figures reported to him by his own experts.

FOX NEWS

Colorado amends coronavirus death count - says fewer have died of COVID-19 than previously reported

Questions raised over accuracy of US coronavirus death toll

POSTED 11:03 AM, MAY 16, 2020, BY FOX NEWS CHANNEL

THE INDEPENDENT

Trump warns US coronavirus death toll could reach 100,000 as Yale expert warns it already has

The Washington Post

Prices may vary in areas outside metropolitan Washington.

Mostly cloudy, shower 58/51 • Tomorrow: Mostly cloudy 76/60 B8

Democracy Dies in Darkness

TUESDAY, APRIL 28, 2020 · $2

U.S. testing 'blueprint' keeps onus on states

Health experts press for more robust federal strategy, greater funding

BY MIKE DEBONIS, CHRIS MOONEY AND JULIET EILPERIN

Pressure mounted Monday on the White House and Congress to

U.S. deaths surged early in pandemic

BY EMMA BROWN, ANDREW BA TRAN, BETH REINHARD AND MONICA ULMANU

FAR EXCEEDED TALLY ASCRIBED TO VIRUS

Over 5 weeks, data show unusually high mortality

In the early weeks of the coronavirus epidemic, the United States recorded an estimated 15,400 excess deaths, nearly two times as many as were publicly attributed to covid-19 at the time, according to an analysis of federal data conducted for The Washington Post by a research team led by the Yale School of Public Health.

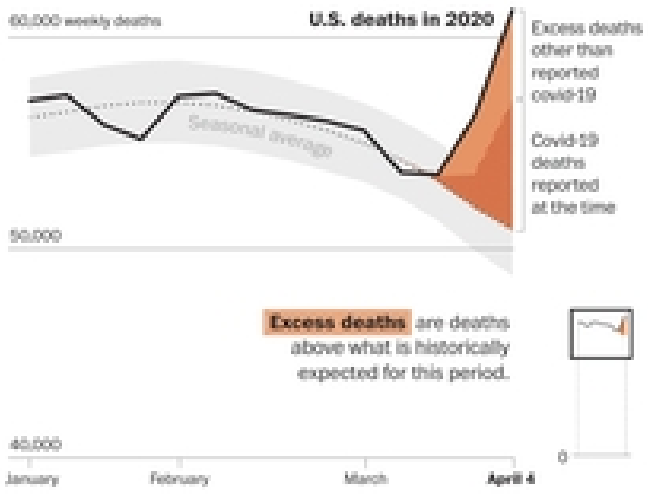
The excess deaths — the number beyond what would normally be expected for that time of year — occurred during March and through April 4, a time when 8,128 coronavirus deaths were reported.

The excess deaths are not necessarily attributable directly to covid-19, the disease caused by the coronavirus. They could include people who died because of the epidemic but not from the disease, such as those who were afraid to seek medical treatment for unrelated illnesses, as well as some number of deaths that are part of the ordinary variation in the death rate. The count is also affected by increases or decreases in other categories of deaths, such as suicides, homicides and motor vehicle accidents.

But in any pandemic, higher-than-normal mortality is a starting point for scientists seeking to understand the full impact of the disease.

The Yale analysis for the first time estimates excess deaths, both nationally and in each state, in those five weeks. Relying on data that the National Center for Health Statistics (NCHS) released Friday, the analysis paints a picture of unusually high mortality that will come into sharper view as more data becomes available.

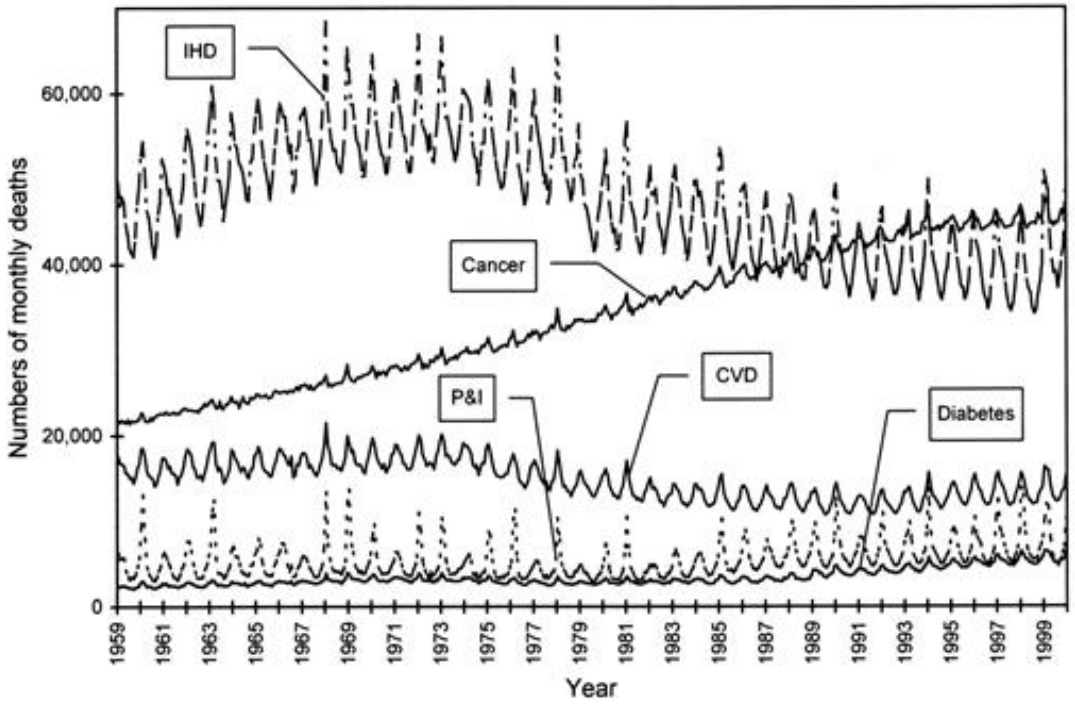
SEE DATA ON A12



Why are the reported number of deaths likely an undercount?

- Many deaths do not receive a pathogen-specific cause of death
- Testing for coronavirus is still sub-optimal
- Application of death coding criteria uneven (requirement for lab confirmation vs suspected)

Reichert et al 2004 AJE

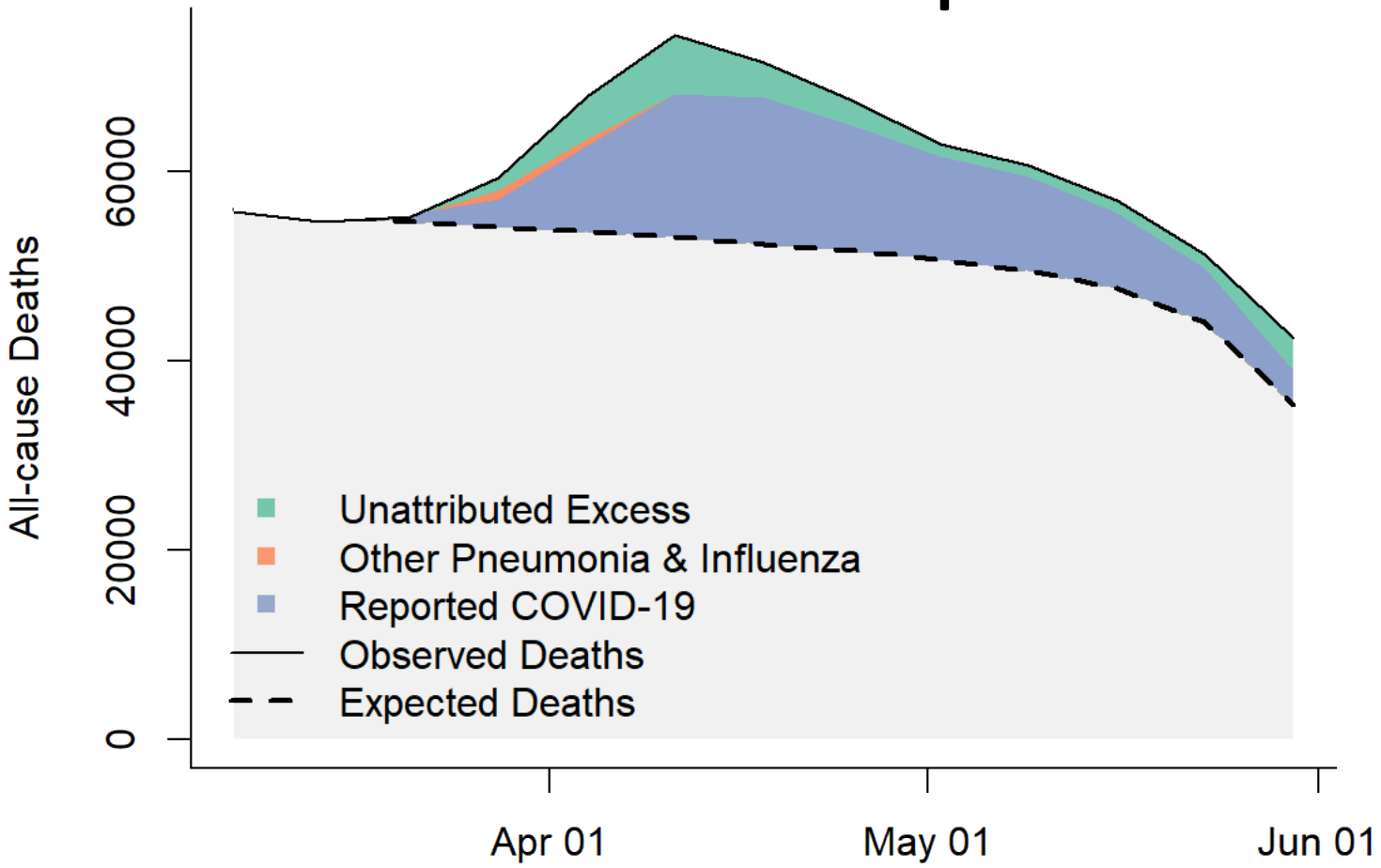


Analysis goals

- Quantify the excess burden of deaths due to any cause
- Quantify the excess burden of deaths recorded as to “pneumonia and influenza ” OR COVID-19
- Compare excess deaths and reported deaths
- Evaluate causes of the gap

- Adjust for reporting delays to obtain timely estimates…

Top line finding



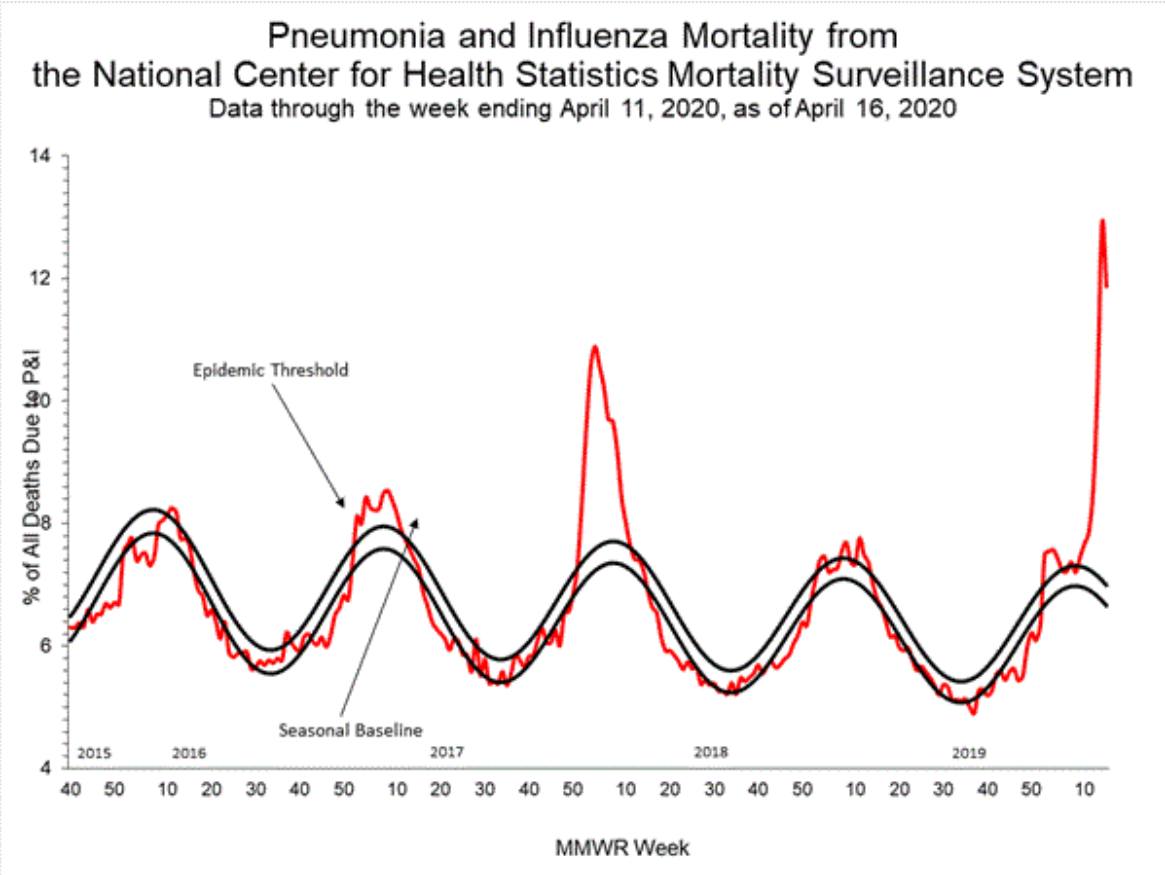
From March 1-May 30:

95,235 reported COVID-19 deaths

122,000 excess deaths

Gap between reported an excess deaths shrank over time

Data



- Weekly deaths due to all causes or due to pneumonia, influenza, or COVID-19 from NCHS multiple cause data for each state
- Testing and confirmed deaths from covidtracking.com

https://www.cdc.gov/flu/weekly/index.htm

General Approach: all-cause data

- Seasonal regression, adjusting for flu activity in the previous week:

$$All_Deaths_{i,t} \sim Poisson(\lambda_{i,t})$$

$$\log(\lambda_{i,t}) = \beta_0 +$$

$$\beta_1 * \sin(\Theta_t) + \beta_2 * \cos(\Theta_t) + \beta_3 * \sin(\Theta_t/2) + \beta_4 * \cos(\Theta_t/2) +$$

Seasonality

$$\beta_6 * \log(Flu_Epidemic_{i,t-1}) +$$

Flu percent positive in previous week

$$\gamma_i + \alpha_i * \log(Flu_Epidemic_{i,t-1}) +$$

Baseline and flu effect varies year-to-year

$$\log(proportion_complete_t)$$

Reporting delay, estimated by Bayesian nowcasting (NoBbs)

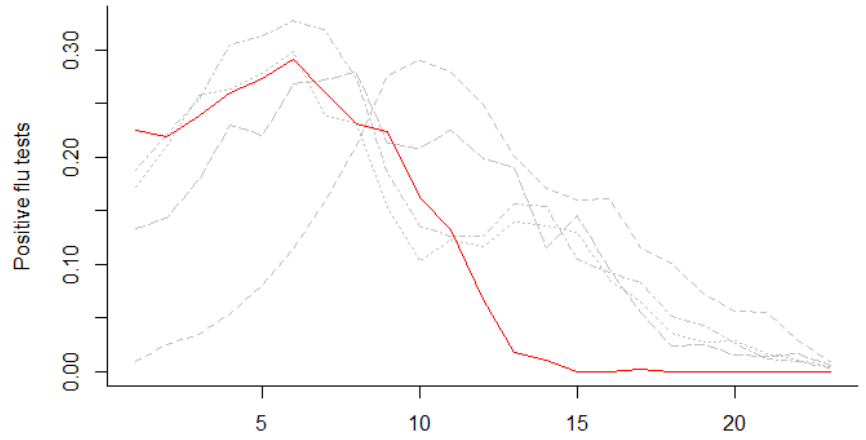
General approach

- Fit to data through beginning of February
- Extrapolate seasonality and baseline forward for Feb-May, assuming seasonality and baseline are consistent
- Generate prediction intervals by resampling:
  - Stage 1: Sample from MVN to capture parameter uncertainty
  - Stage 2: Sample from Poisson (mean=stage 1 estimates) to capture prediction uncertainty
  - (Working on Bayesian alternative)

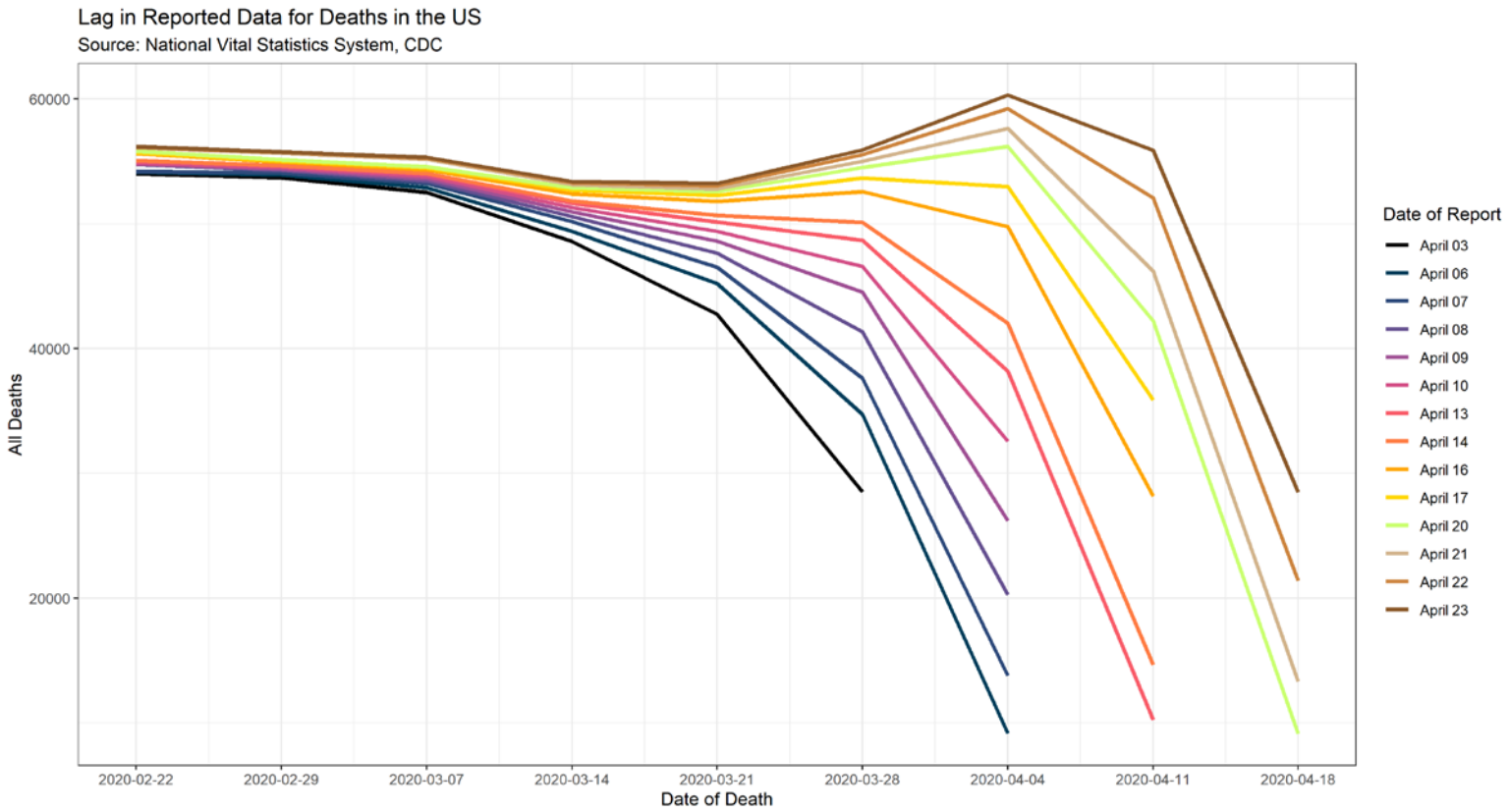
Challenge #1: Flu activity

- We want to adjust for flu activity;
  - Could help to avoid attributing excess deaths in March to flu
  - BUT: flu fell to historically low levels; adjustment for flu would bias the baseline
  - Solution: de-seasonalize flu, and use

flu-above-baseline to adjust death counts

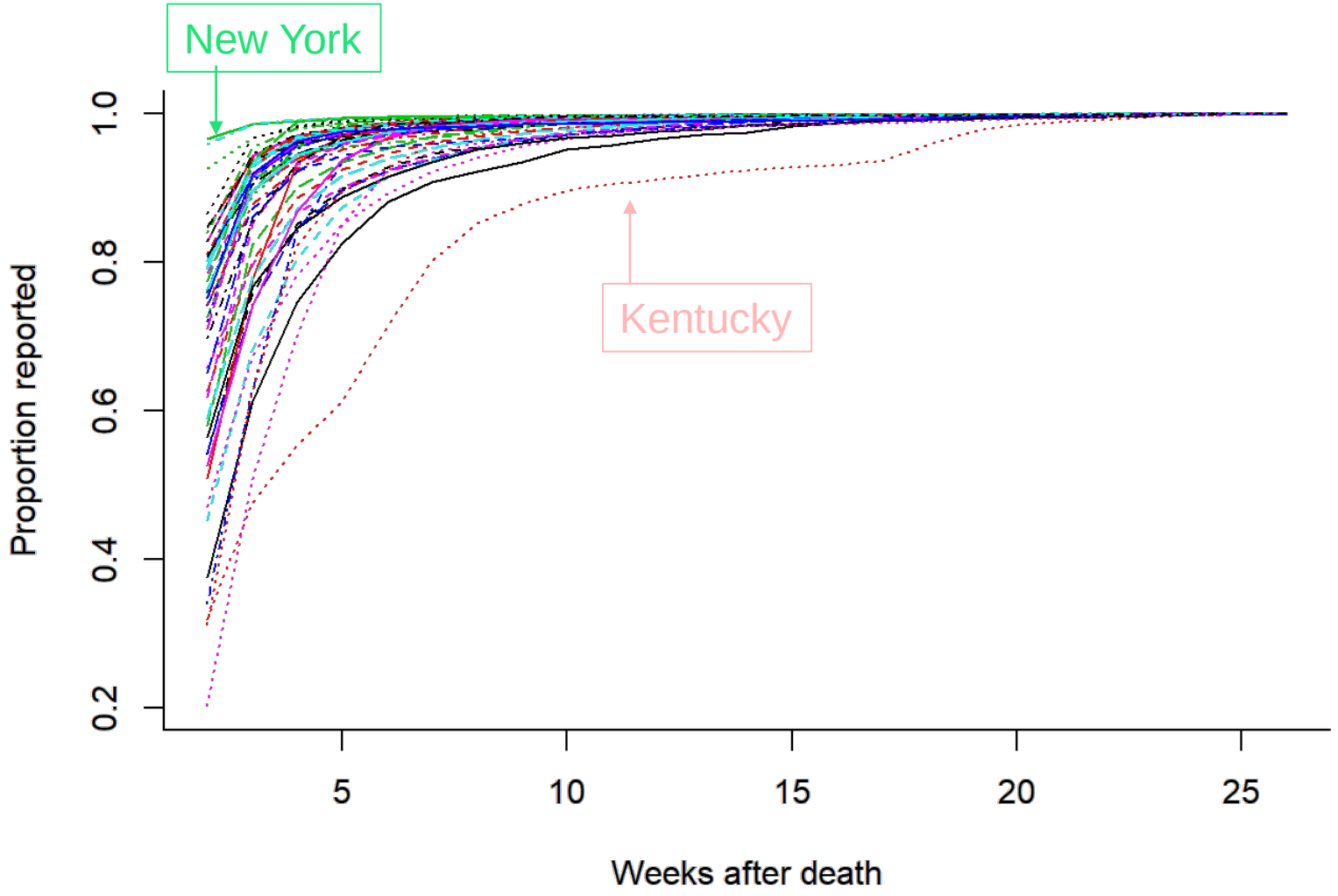


Challenge #2: Lags in reporting of deaths



-Provisional counts stabilize after ~4 weeks

Lags in reporting vary by state



Correction for reporting delays

- Use Nowcasting by Bayesian Smoothing approach of McGough et al (PLOS Comp Biol 2020)
- Obtain 'provisional' reports of number of deaths that occurred in each week , update weekly

Number of death in week *t* that were reported *d* weeks after the death

$N_{t,d} \sim \text{Poisson}(\text{lambda}_{t,d})$

$\log(\text{lambda}_{t,d}) = \text{alpha}_t + \log(\text{Beta}_d)$ Alpha is a random walk

$\text{sum}(\text{Beta}_d)$ → the proportion of deaths that have been reported for a given delay

Beta gives the proportion of the deaths reported in week d

General Approach: all-cause data

- Seasonal regression, adjusting for flu activity in the previous week:

$$All_Deaths_{i,t} \sim Poisson(\lambda_{i,t})$$

$$\log(\lambda_{i,t}) = \beta_0 +$$

$$\beta_1 * \sin(\Theta_t) + \beta_2 * \cos(\Theta_t) + \beta_3 * \sin(\Theta_t/2) + \beta_4 * \cos(\Theta_t/2) +$$

Seasonality

$$\beta_6 * \log(Flu_Epidemic_{i,t-1}) +$$

Flu percent positive in previous week

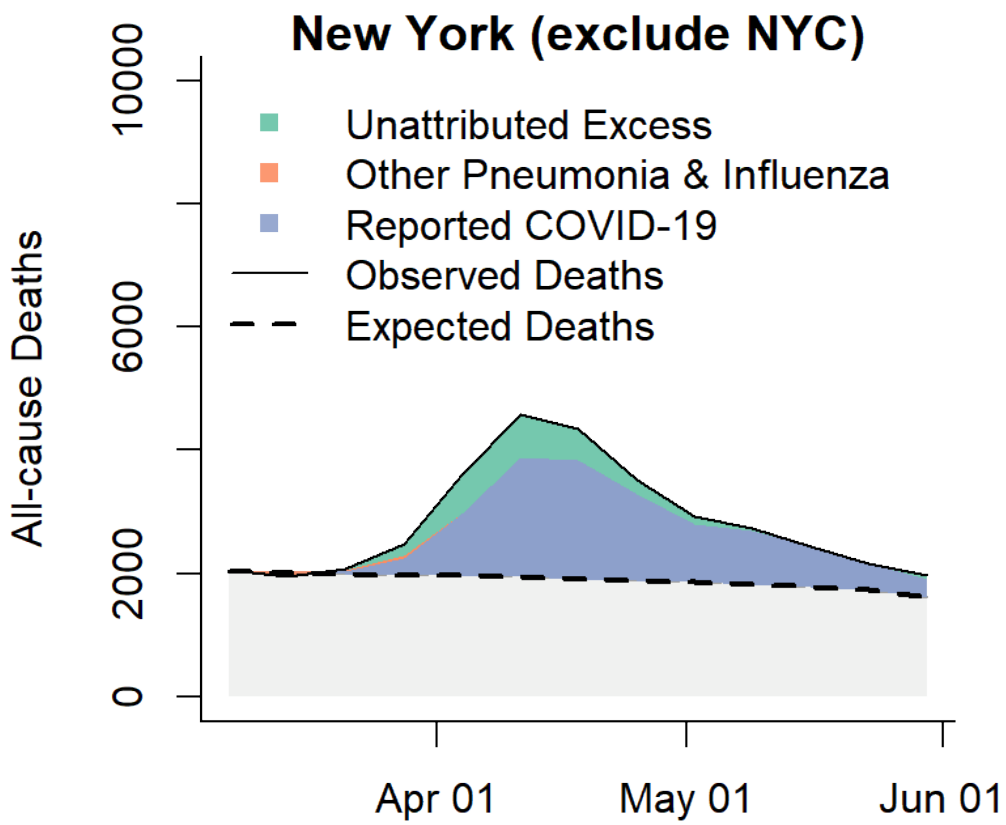
$$\gamma_i + \alpha_i * \log(Flu_Epidemic_{i,t-1}) +$$

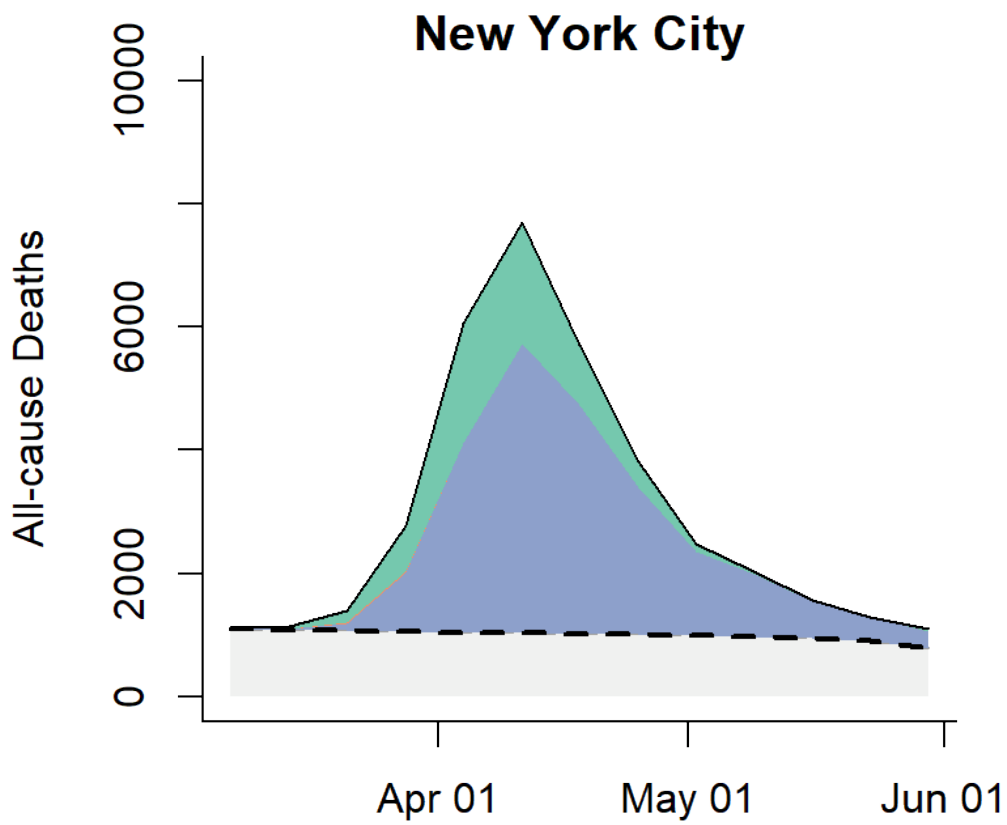
Baseline and flu effect varies year-to-year

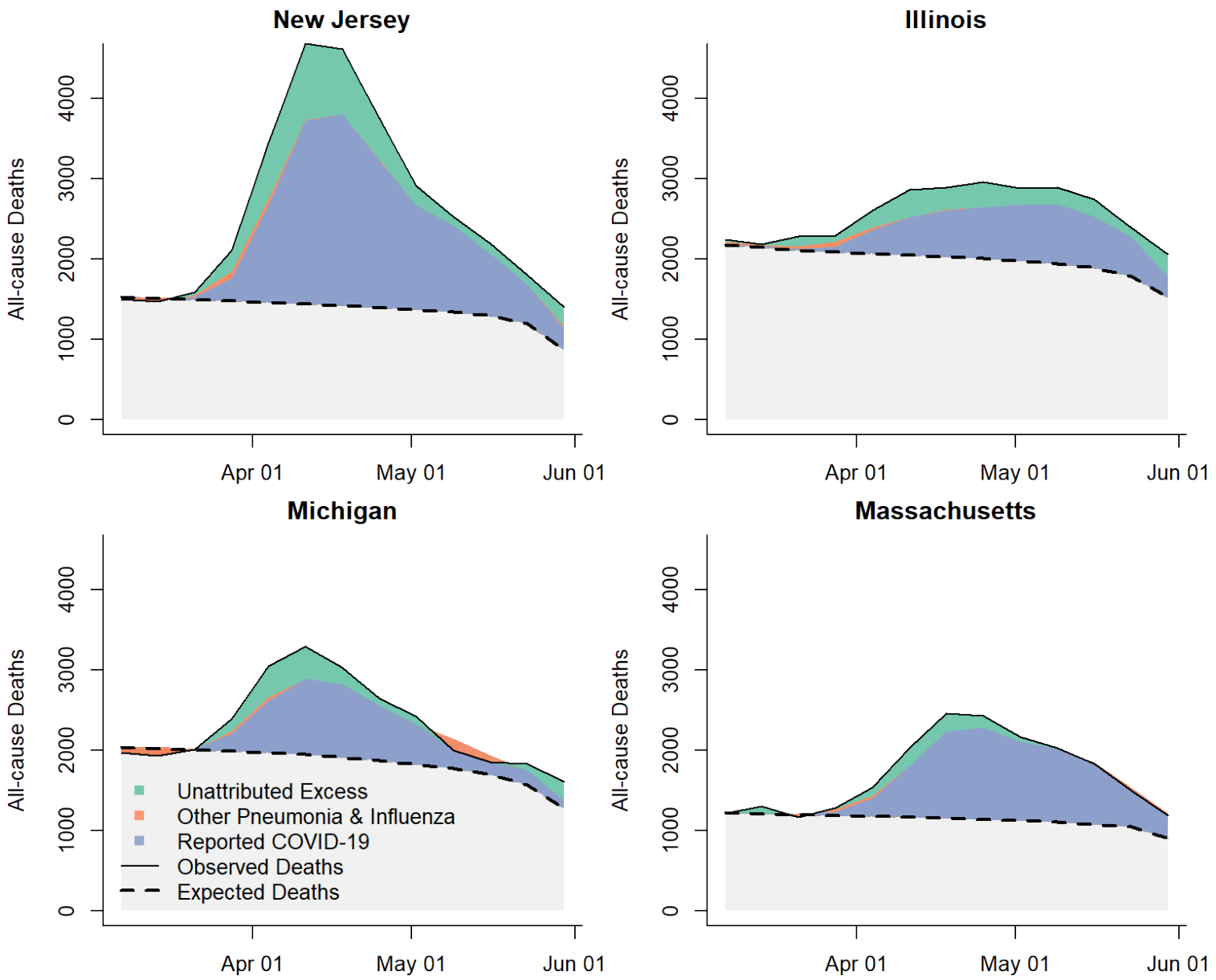
$$\log(proportion_complete_t)$$

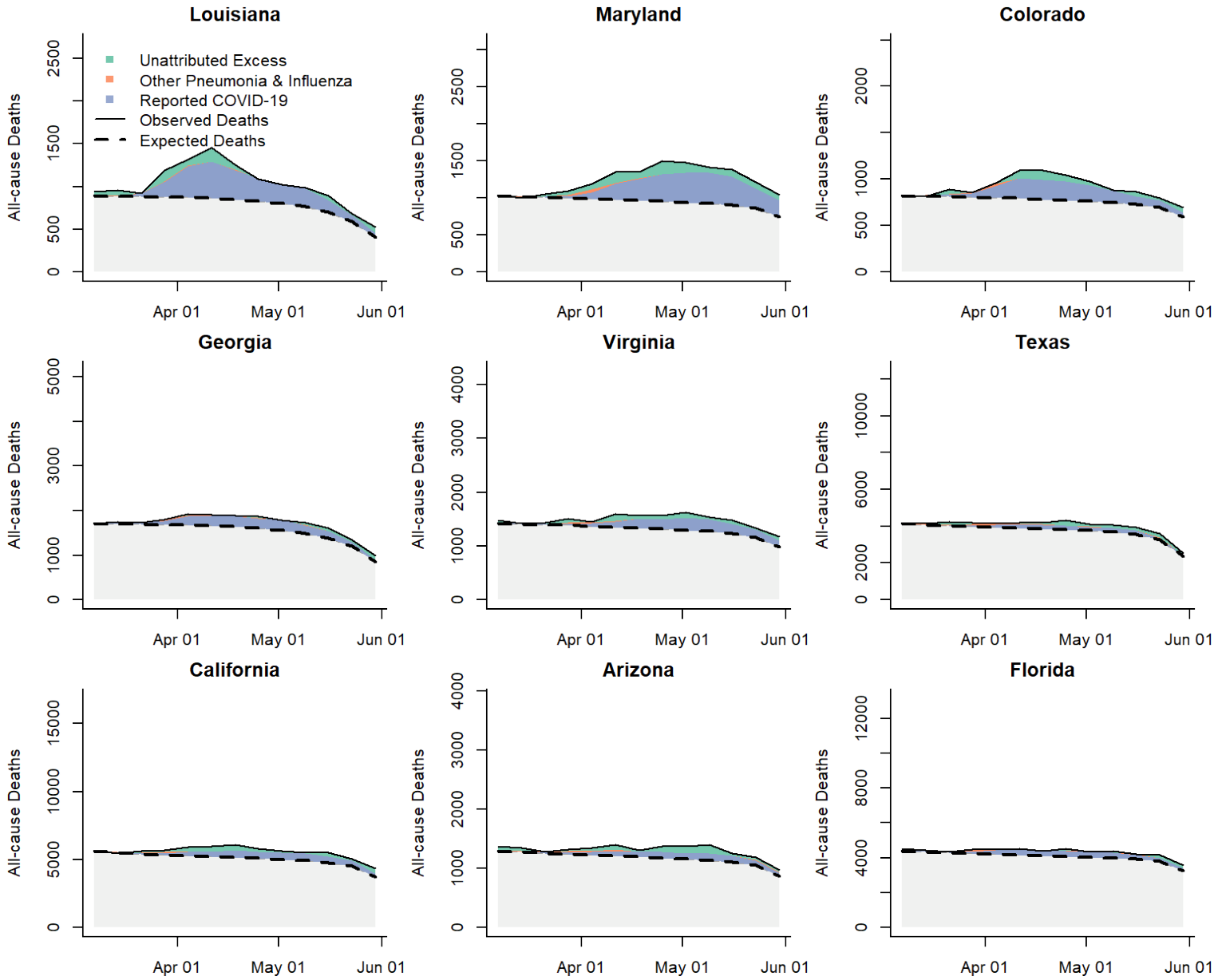
Reporting delay, estimated by Bayesian nowcasting (NoBbs)

Fit of the models to pre-pandemic data


New York (exclude NYC)
Unattributed Excess
Other Pneumonia & Influenza
Reported COVID-19
Observed Deaths
Expected Deaths
All-cause Deaths
0
2000
6000
10000
Apr 01
May 01
Jun 01


New York City
All-cause Deaths
0
2000
6000
10000
Apr 01
May 01
Jun 01

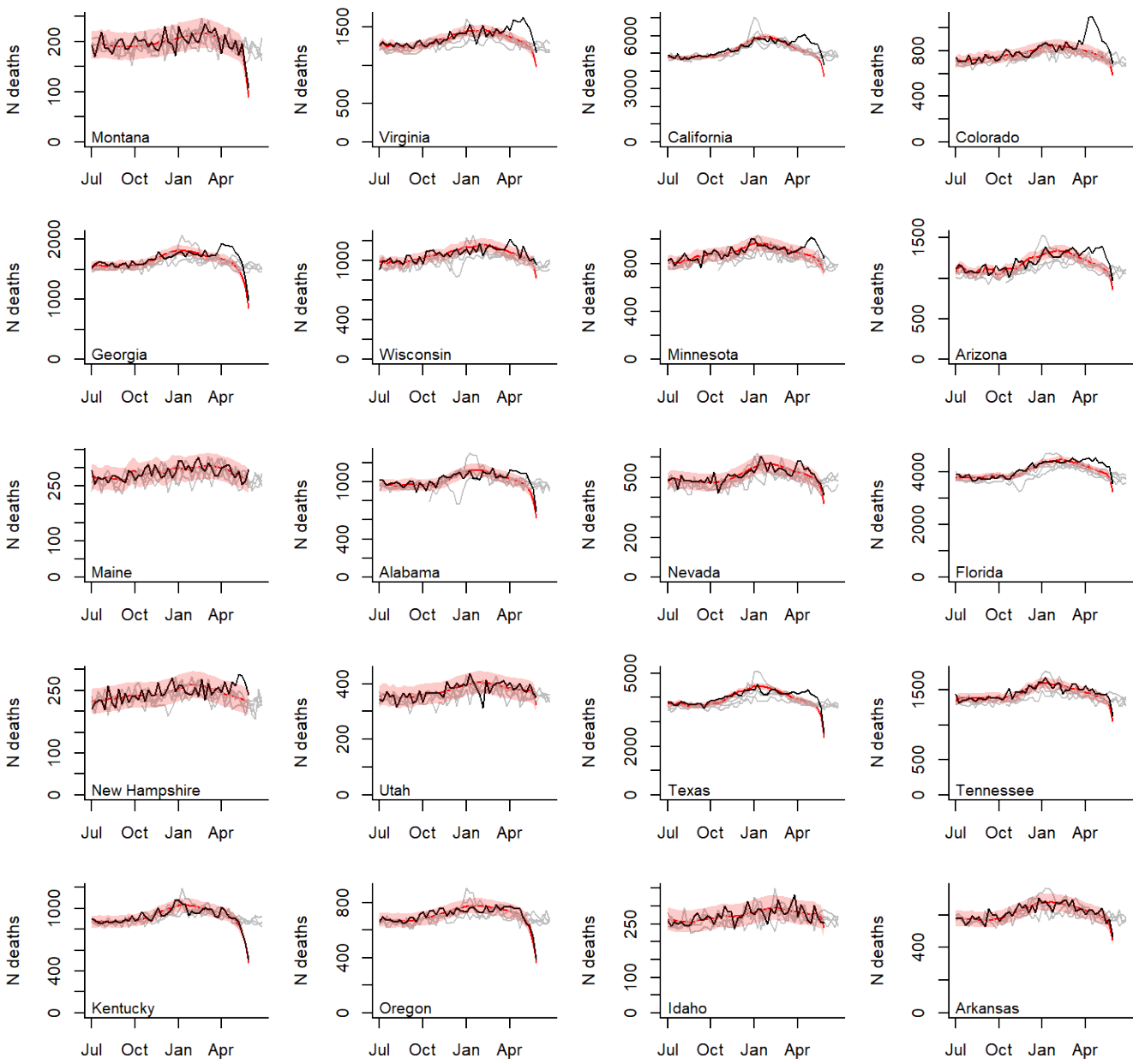

New Jersey
Illinois
Michigan
Massachusetts
All-cause Deaths
0
1000
2000
3000
4000
Apr 01
May 01
Jun 01
Unattributed Excess
Other Pneumonia & Influenza
Reported COVID-19
Observed Deaths
Expected Deaths


Louisiana
Maryland
Colorado
Georgia
Virginia
Texas
California
Arizona
Florida
Unattributed Excess
Other Pneumonia & Influenza
Reported COVID-19
Observed Deaths
Expected Deaths
All-cause Deaths
Apr 01
May 01
Jun 01

Observed and Excess deaths due to COVID-19, pneumonia/influenza/Covid-19, and all-causes COVID-19, from March 1, 2020 through May 30, 2020

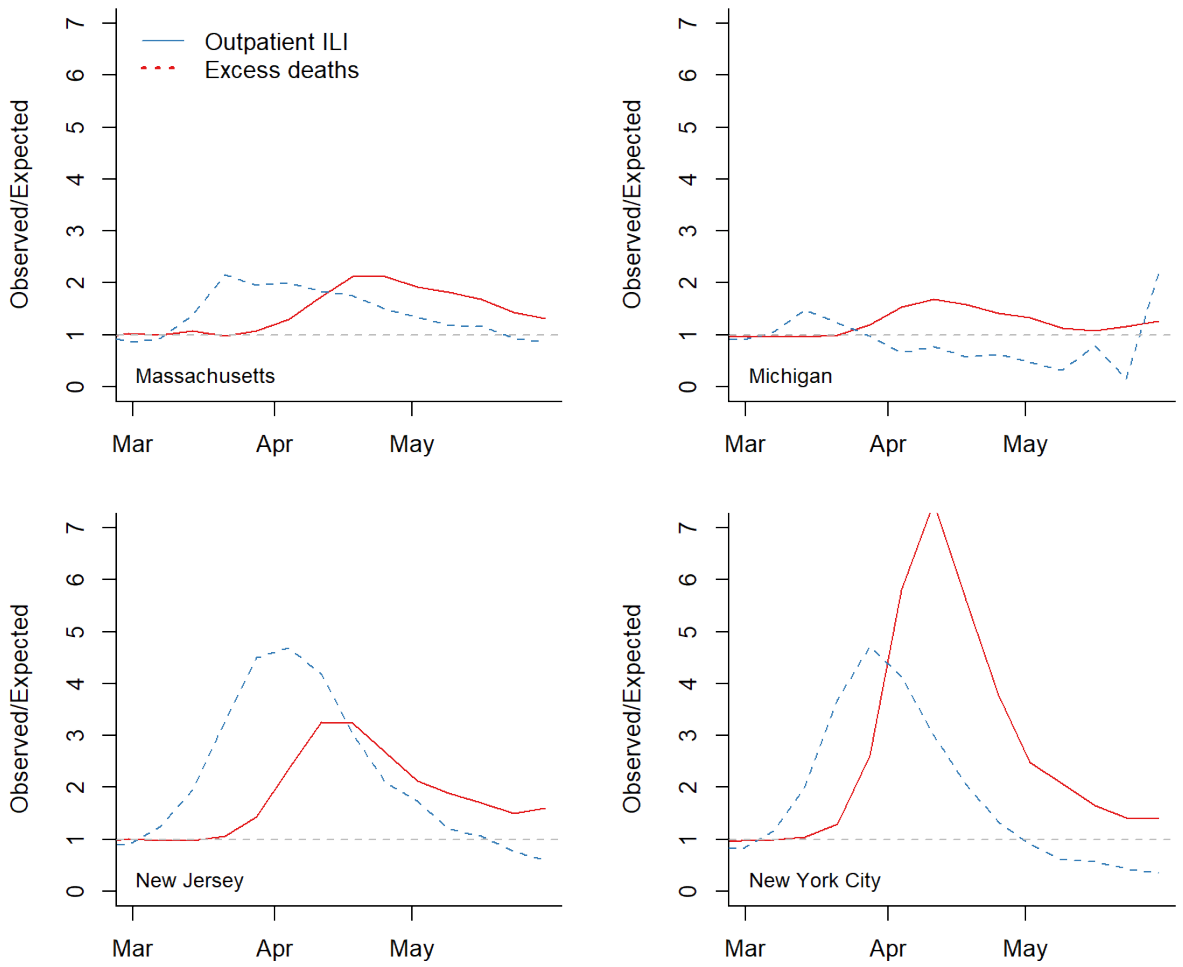
State	Observed Deaths	Expected deaths	Excess all-cause deaths	COVID-19 deaths (U07.1)	Excess Pneumonia/Influenza/Covid-19 deaths	Excess all-cause deaths/100,000
US	780975	658700	122300 (116800,127000)	95235	94130 (92350,95690)	37.3 (35.6,38.7)
NYC	38170	13000	25100 (24800,25400)	18603	17500 (17280,17670)	299.1 (295.3,302.7)
NJ	34013	17800	16200 (15800,16500)	11952	12060 (11910,12180)	182.3 (178.1,186.2)
NY	36871	24600	12300 (11900,12700)	9773	9220 (9040,9380)	111.4 (107.4,114.9)
MA	22146	14700	7400 (7100,7700)	6627	6400 (6280,6510)	107.6 (103,112.2)
LA	13256	10200	3000 (2700,3300)	2310	2340 (2270,2400)	64.7 (59.1,70.3)
MI	30044	23900	6100 (5700,6600)	4839	4740 (4620,4870)	61.4 (56.7,65.6)
IL	33296	25800	7500 (7100,7900)	4911	4920 (4800,5050)	59.2 (55.7,62.3)
SC	13601	12200	1400 (1100,1700)	528	480 (400,550)	27.9 (21.8,33.6)
AZ	16929	15200	1800 (1400,2100)	846	930 (840,1010)	24 (19.6,28.5)
GA	22105	19900	2200 (1800,2600)	1686	1800 (1710,1880)	21 (17.2,24.4)
WA	15367	14000	1300 (1000,1700)	1021	1010 (930,1090)	17.5 (13.4,21.7)
CA	72407	65600	6800 (6100,7500)	4046	4130 (3940,4330)	17.2 (15.5,19)
FL	56462	52900	3500 (2900,4100)	2457	2940 (2780,3080)	16.4 (13.4,19.3)

-Excess death rate varies by >10x between states
-Gap between excess and COVID death varies by states
-Few unexplained pneumonia deaths

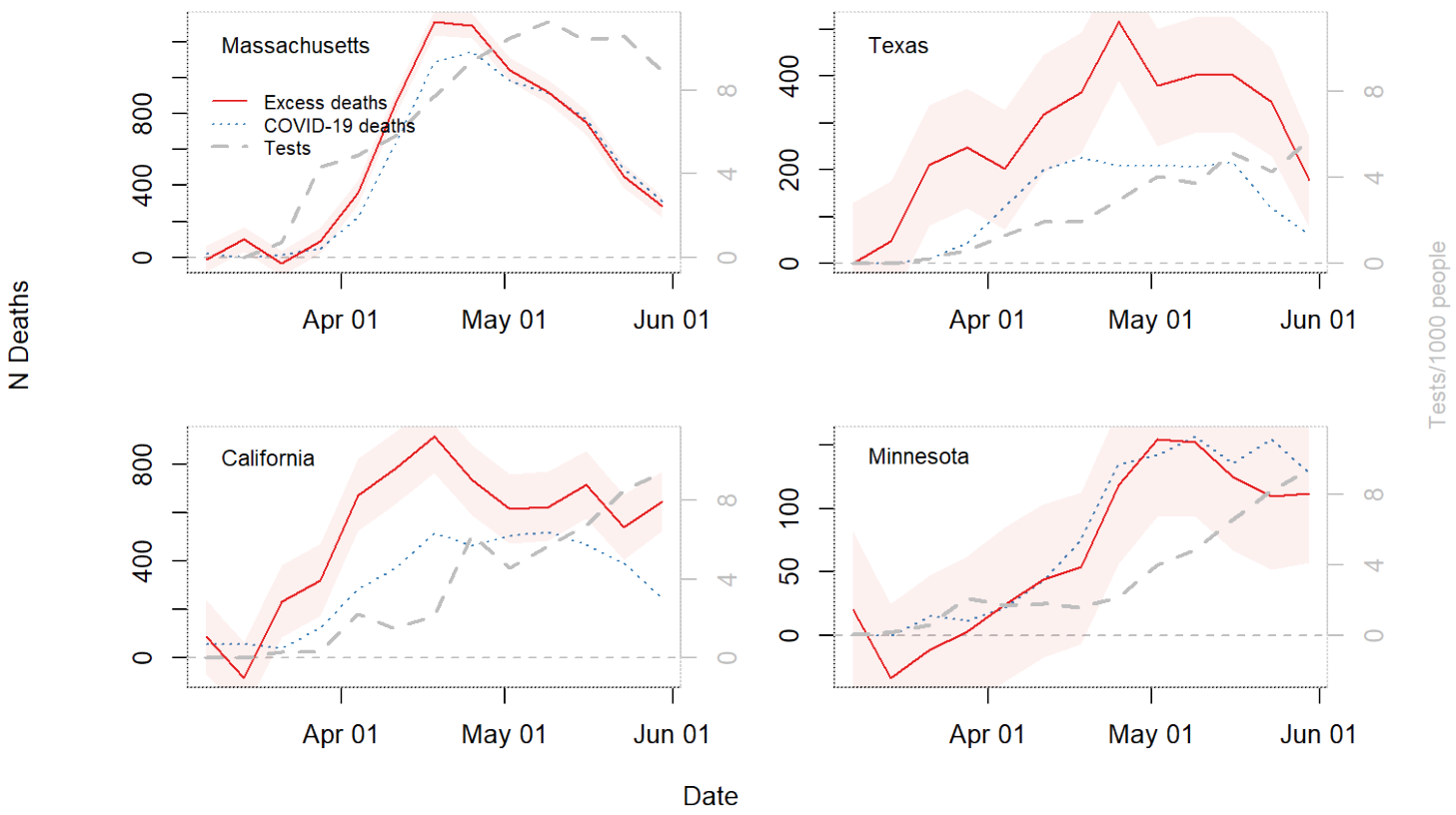


Observed and expected proportion of deaths due to P&I, by state,
through May 30, compared to previous years

Recent dip in observed and baseline is due to reporting delays



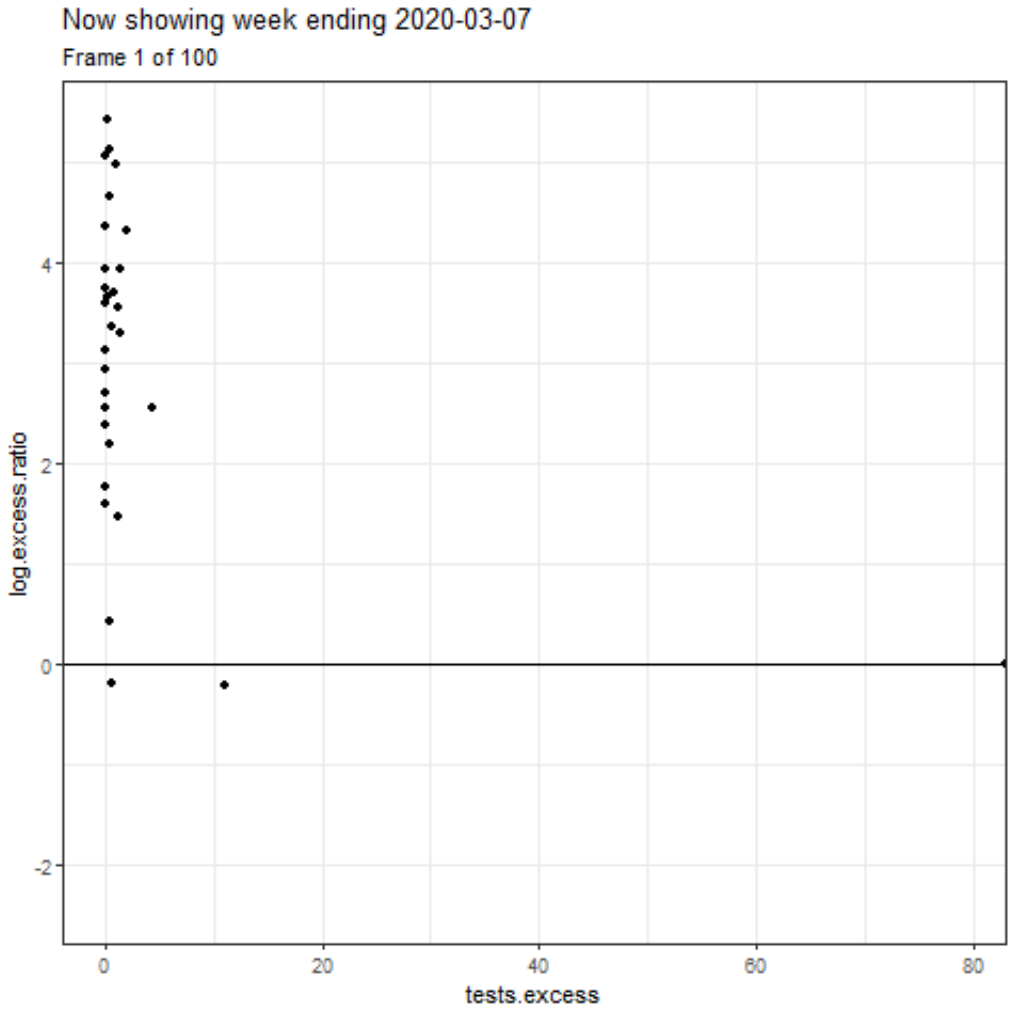
The increase in excess deaths lagged behind spikes in influenza-like illness by several weeks



Excess deaths increased earlier and more intensely than reported COVID deaths

This difference was more pronounced where testing ramped up later

Relationship between testing and unattributed COVID deaths



As testing increased, the gap between excess deaths and COVID-19 deaths narrowed

A comment on the excess gap

- When we started monitoring this, excess deaths were 60% higher than reported deaths
- This has now narrowed to <30% (nationally)

- This is because of: backfilling of cause of death data; changes in criteria for attributing deaths to COVID-19, increased testing…

What is causing the unattributed excess?

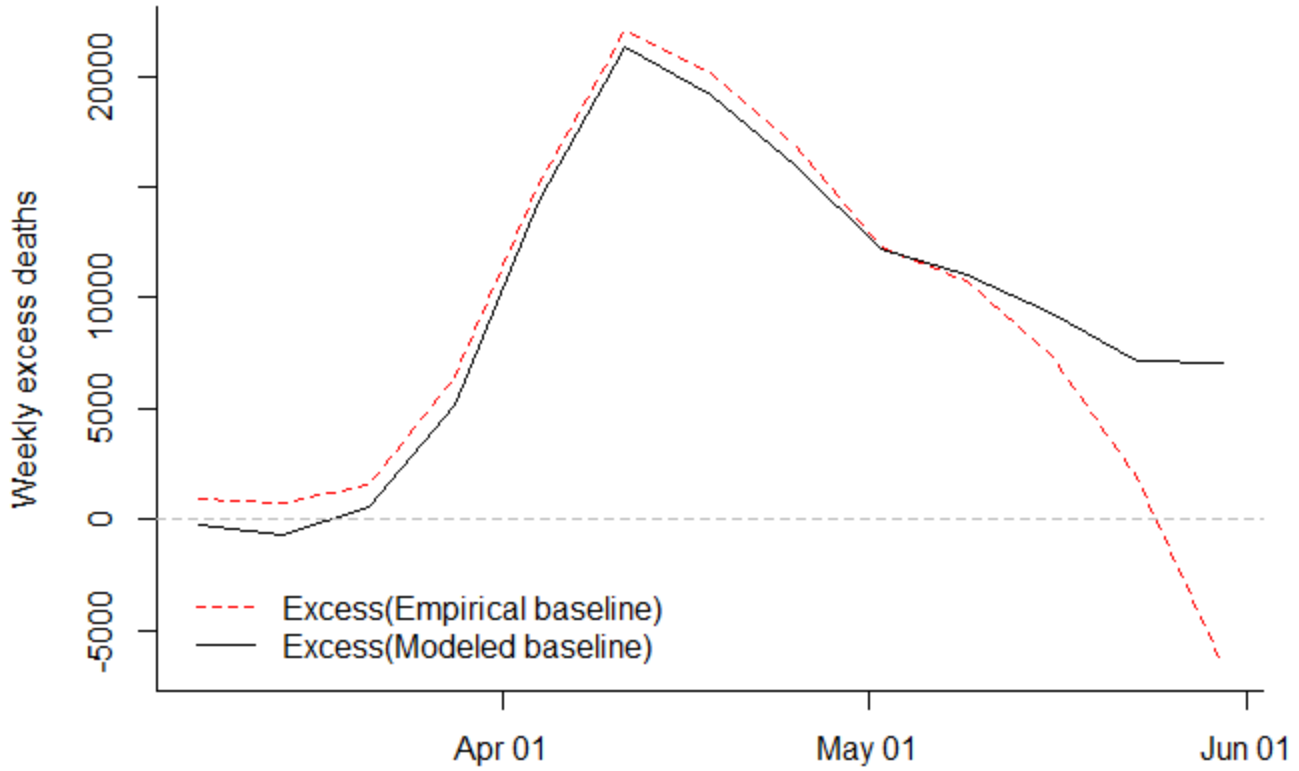
Total number of deaths above average since 2/1/2020, by cause of death

Respiratory diseases	Influenza and pneumonia	2,916
	Chronic lower respiratory disease	517
	Other diseases of the respiratory system	1,403
Circulatory diseases	Hypertensive dieases	8,334
	Ischemic heart disease	5,678
	Heart failure	2,274
	Cerebrovascular diseases	3,784
	Other diseases of the circulatory system	167
Malignant neoplasms	Malignant neoplasms	1,852
Alzheimer disease and dementia	Alzheimer disease and dementia	15,024
Other select causes	Diabetes	5,138
	Renal failure	495
	Sepsis	258

Alzheimers/dementia might reflect that a large proportion of deaths occur in nursing homes

https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm

Sensitivity analysis: empirical baseline



- Calculate average cases/week in previous years
- Good agreement with modeled baseline until early May (reporting delays)

Sensitivity analysis: exclude influenza adjustment

	Excess deaths, adjusted for influenza	Excess deaths, unadjusted for influenza
AK	-100 (-200,0)	-70 (-160,0)
AL	900 (600,1200)	710 (420,1000)
AR	100 (-200,300)	80 (-170,300)
AZ	1800 (1400,2100)	1250 (900,1570)
CA	6800 (6100,7500)	6190 (5500,6850)
CO	1900 (1600,2100)	1730 (1450,1980)
DC	500 (400,600)	490 (400,590)
DE	400 (300,600)	440 (320,550)
FL	3500 (2900,4100)	2770 (2140,3370)
GA	2200 (1800,2600)	1960 (1580,2310)
HI	-100 (-200,100)	-80 (-230,50)
IA	100 (-100,400)	150 (-90,380)
ID	0 (-100,200)	10 (-160,160)
IL	7500 (7100,7900)	7370 (6910,7760)
IN	2100 (1800,2500)	2120 (1780,2440)
KS	0 (-200,200)	10 (-200,220)

Adjustment for influenza pulls baseline down slightly, which increases excess estimates in some states

Conclusions

- Official statistics likely undercount the burden of deaths due to COVID-19
- There is no evidence that official tolls *over* count COVID deaths
- The degree of undercounting varies by state and over time and might be partially related to testing intensity

Possible methodological extensions

- Integrate with case and testing data into a Bayesian nowcasting model
- Incorporate reporting delay model directly into main model of the baseline
- Develop more parsimonious hierarchical mode with the states

- **Continue to monitor as states reopen**

Other work on Covid-19

- Developed an R package for estimating and visualizing increases above seasonal baselines (ExcessILI)
  - Now being used by Partners in Health, city of Chicago, AZ DPH, TX DPH… to monitor syndromic surveillance signals
- Collaborations on estimating effective reproductive number from imperfect data (Ted Cohen, Ginny Pitzer et al)

Resources

Estimating the early death toll of COVID-19 in the United States

https://doi.org/10.1101/2020.04.15.20066431

https://github.com/weinbergerlab/ExcessILI package

https://github.com/weinbergerlab/excess_pi_covid death data and code

Full paper coming very soon