Methods and Motives for Infectious Disease Models: The Tale of COVID-19

Nicholas G. Reich

ASTMH Committee on Global Health Pre-Meeting Course
"Modeling for Disease Outbreaks: Practical Approaches to Understanding and Using Models"
11 November 2020

reichlab.io
reichlab

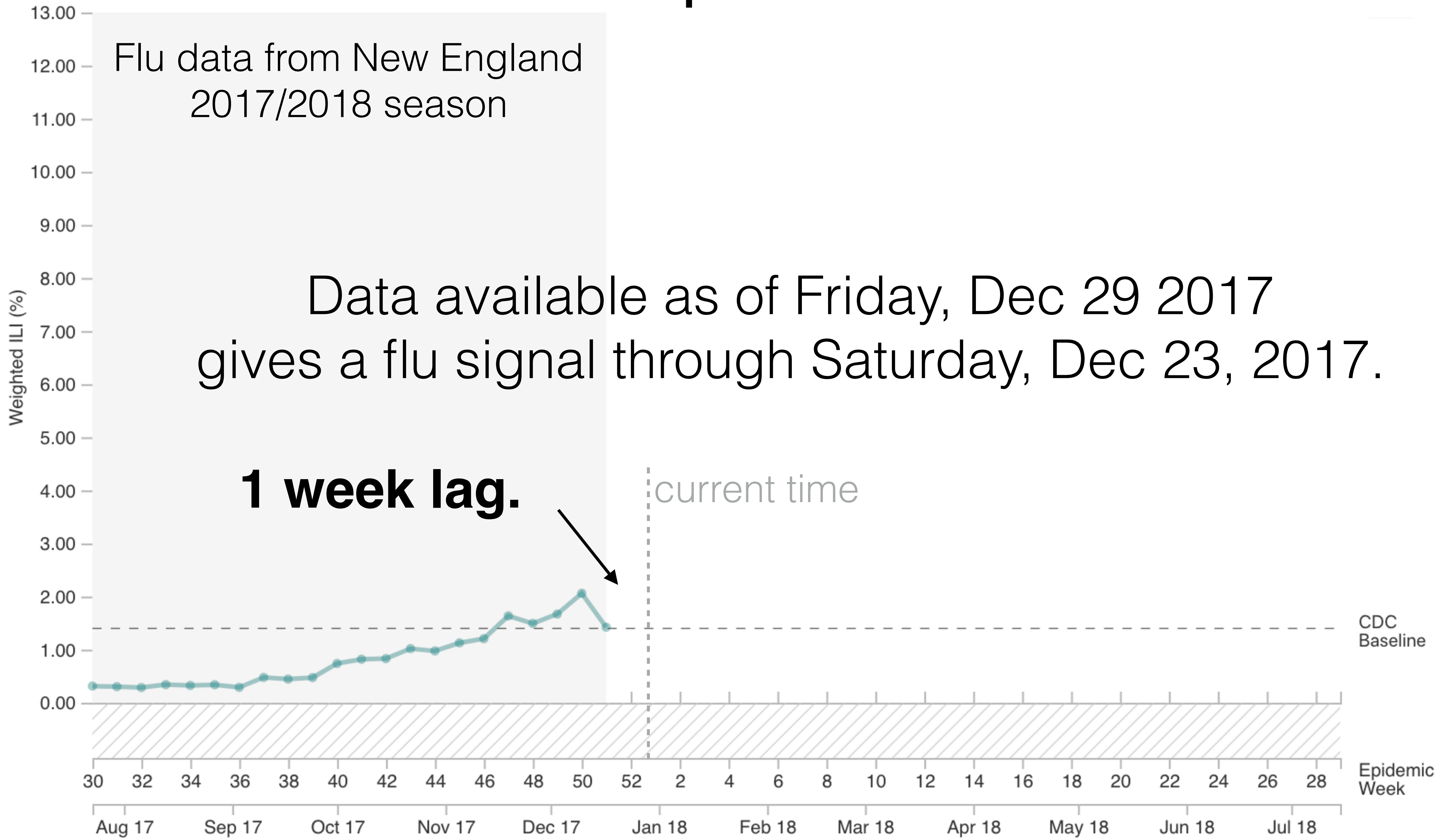
Reich Lab | AT UMASS AMHERST

These slides are available for download at:
https://covid19forecasthub.org/doc/talks/

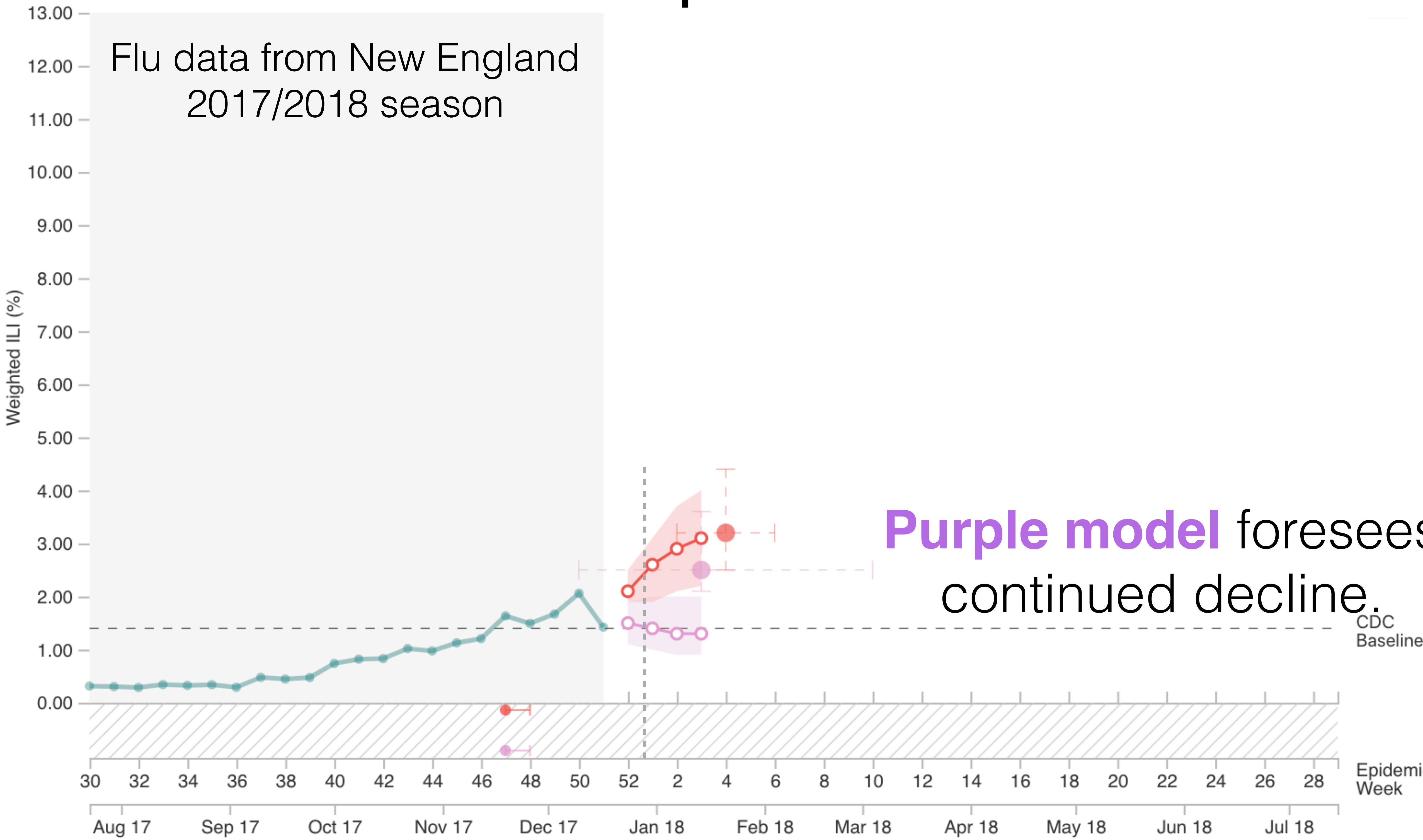
This work has been supported by the National Institutes of General Medical Sciences (R35GM119582) and the Centers for Disease Control and Prevention (1U01IP001122). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIGMS, the National Institutes of Health, or CDC.

Why model?

Real-time public health data is imperfect

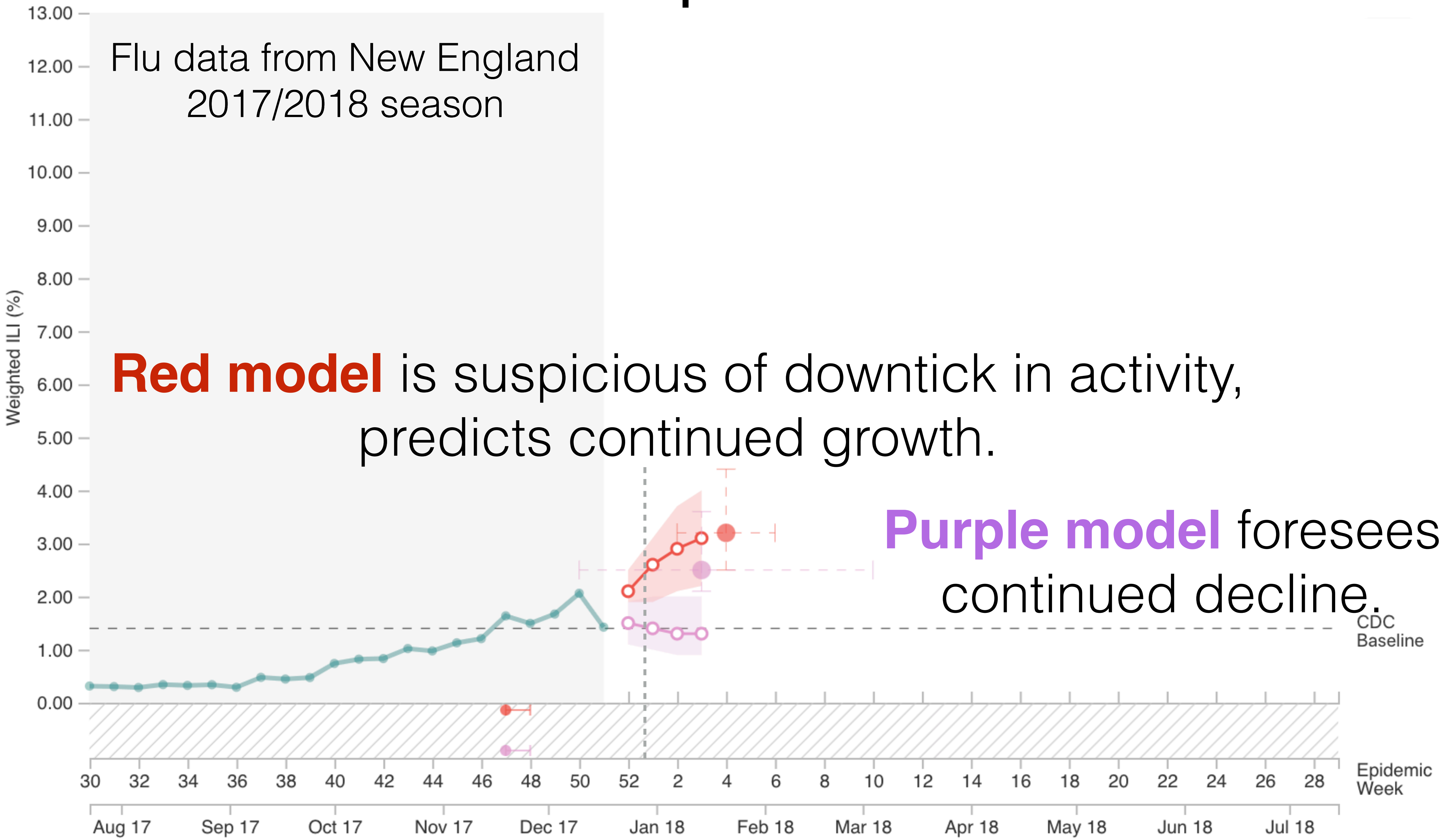


Real-time public health data is imperfect

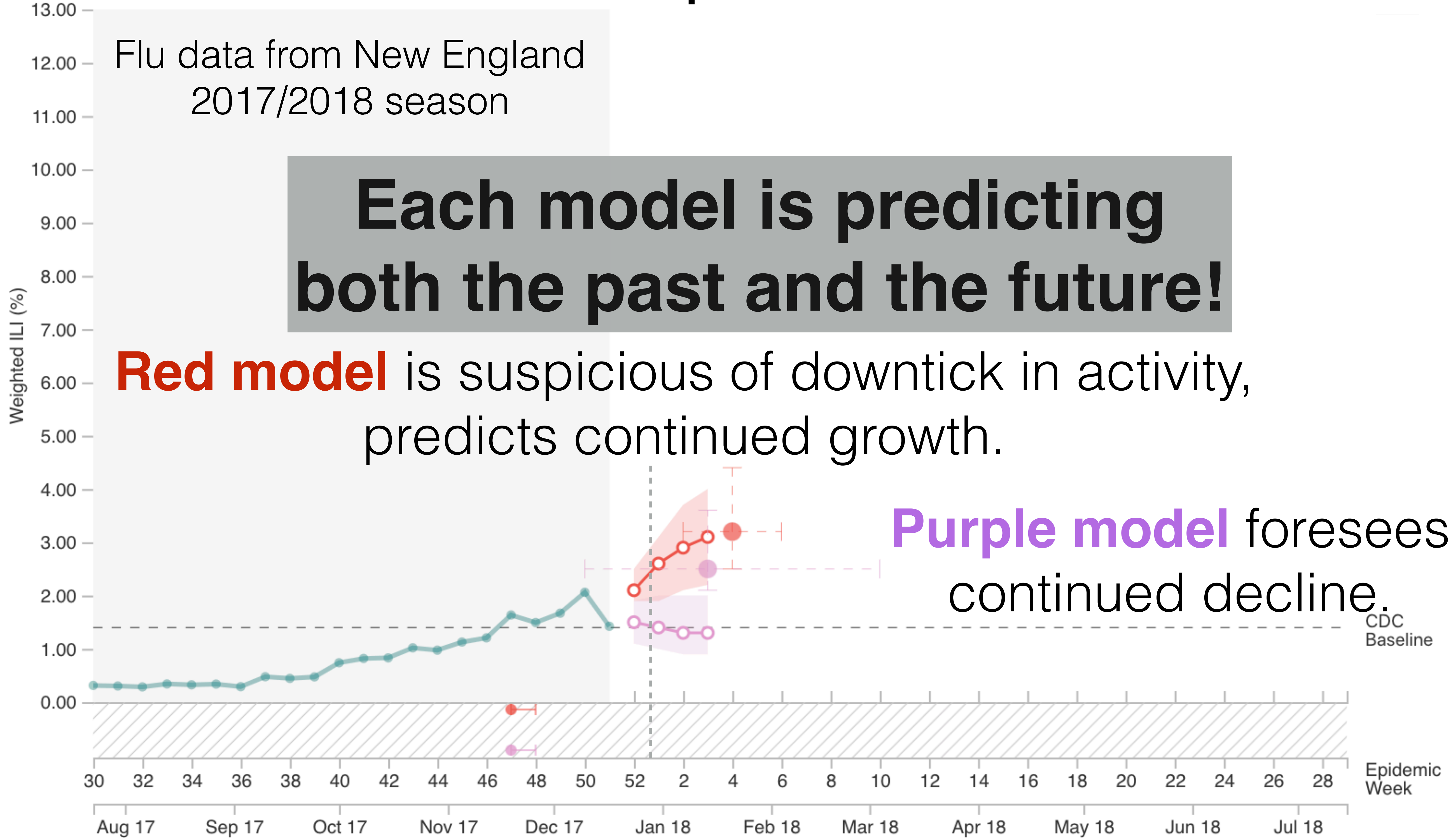


Purple model foresees continued decline.

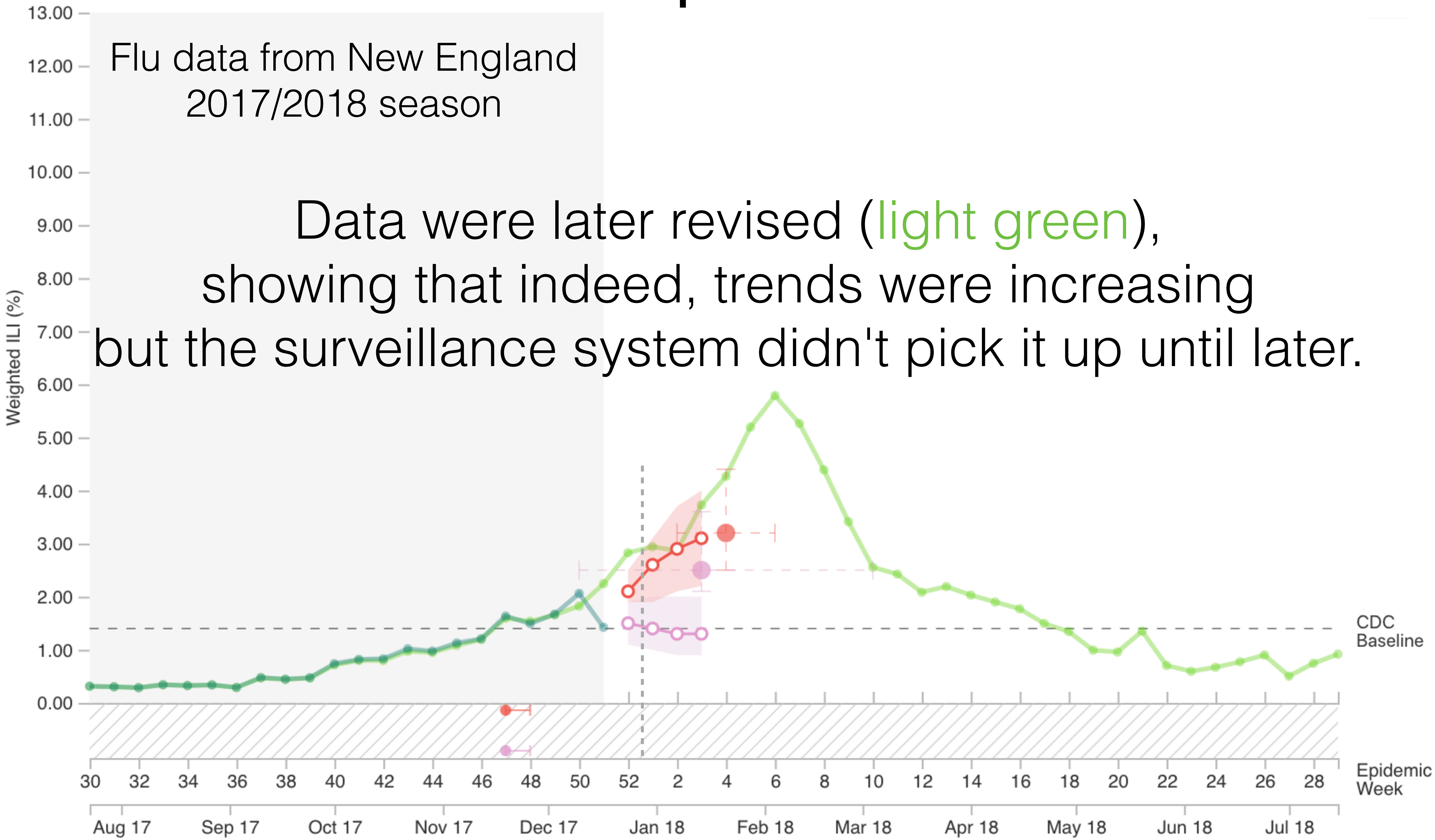
Real-time public health data is imperfect



Real-time public health data is imperfect



Real-time public health data is imperfect



Data were later revised (light green),
showing that indeed, trends were increasing
but the surveillance system didn't pick it up until later.

Good models might…

- Anticipate and adjust for data quality issues.
- Infer what is happening right now.
- Forecast what will be observed in the near future.
- Project hypothetical outcomes in the distant future.

Don't expect a single model to do all of these things well!

COVID-19 example

California COVID Assessment Tool

https://calcat.covid19.ca.gov/cacovidmodels/

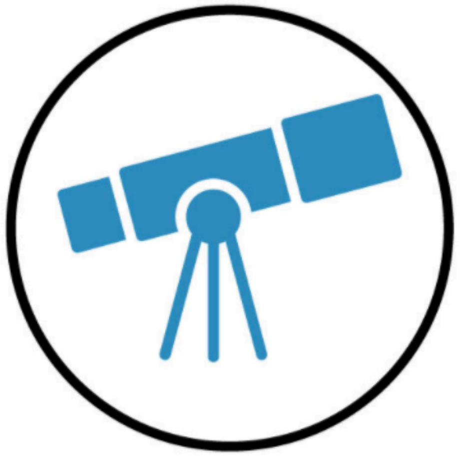
Nowcasts

How fast is COVID-19 spreading right now?

Forecasts

What can we expect in the next 2-4 weeks?

Scenarios

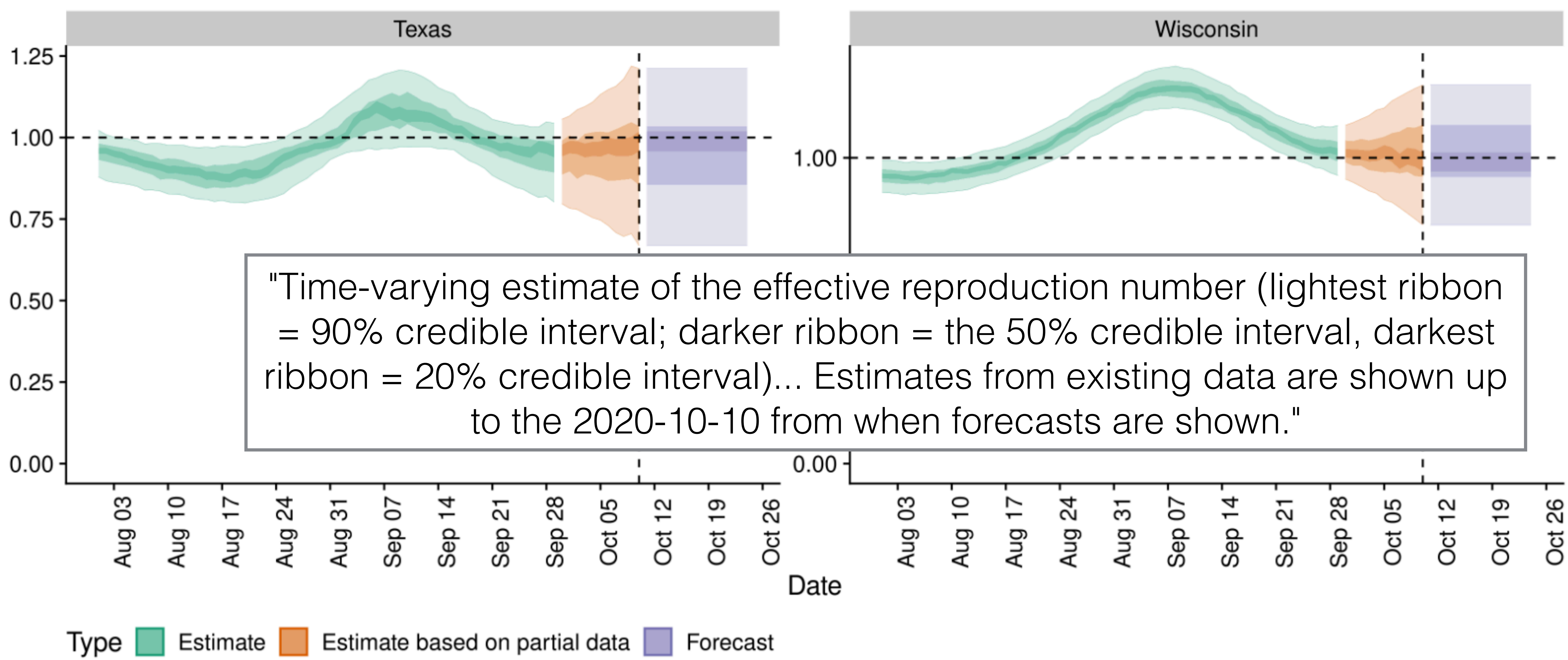


What are the long-term impacts under different scenarios?

How fast is COVID-19 spreading right now?

Nowcasting

Building a model that draws inference about about trends the recent past.



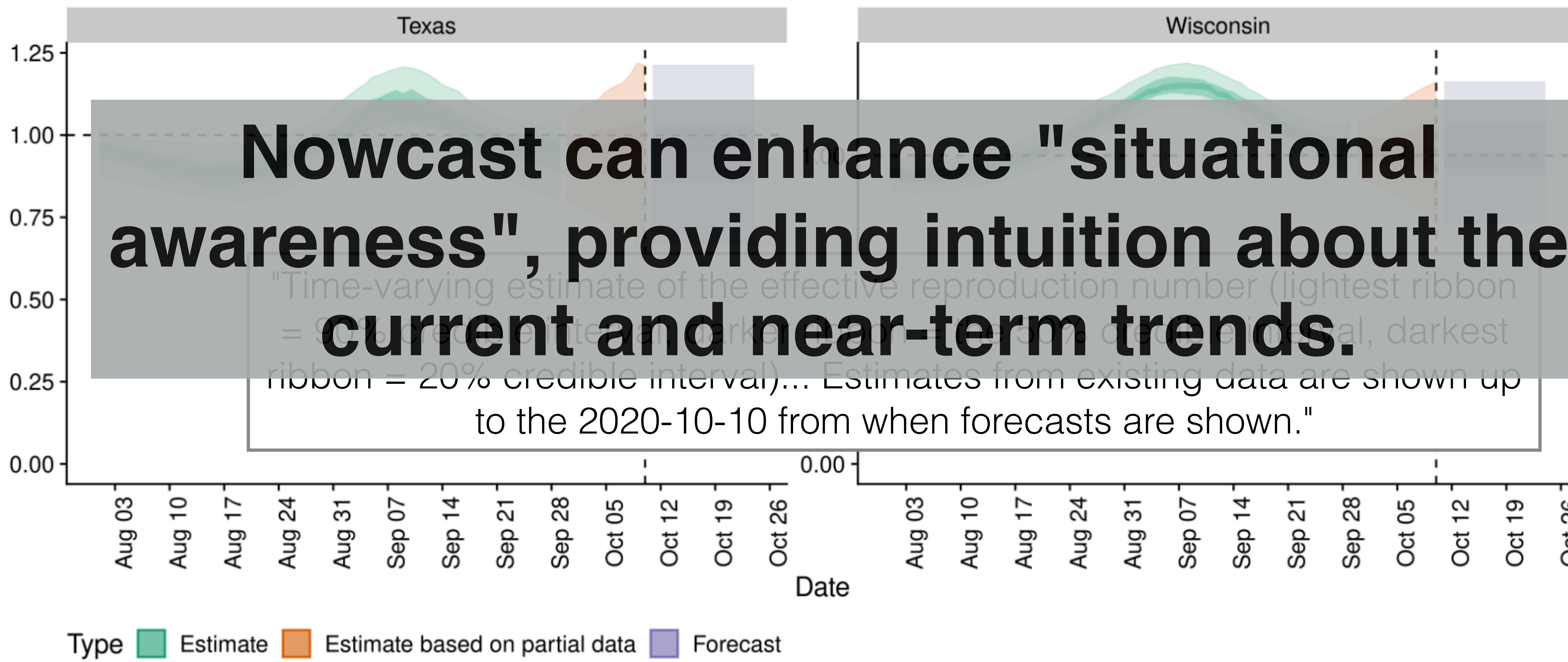
"Time-varying estimate of the effective reproduction number (lightest ribbon = 90% credible interval; darker ribbon = the 50% credible interval, darkest ribbon = 20% credible interval)... Estimates from existing data are shown up to the 2020-10-10 from when forecasts are shown."

https://epiforecasts.io/covid/posts/national/united-states/

How fast is COVID-19 spreading right now?

Nowcasting

Not as agreed upon definition, but I'd vote for "building a model that draws inference about about trends the recent past."



Nowcast can enhance "situational awareness", providing intuition about the current and near-term trends.

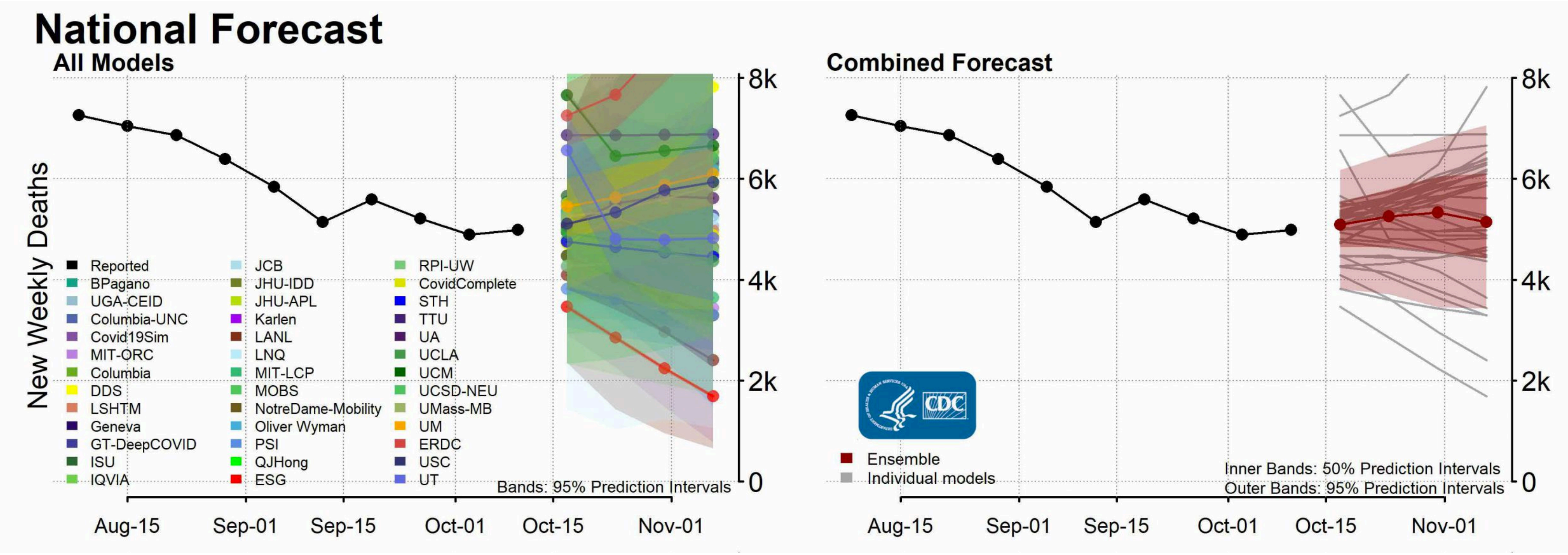
"Time-varying estimate of the effective reproduction number (lightest ribbon = 90% credible interval; darker ribbon = the 50% credible interval, darkest ribbon = 20% credible interval)... Estimates from existing data are shown up to the 2020-10-10 from when forecasts are shown."

https://epiforecasts.io/covid/posts/national/united-states/

Short-term Forecasting

What can we expect in the next 2-4 weeks?

Making **falsifiable, evaluable** predictions of observable future quantities.

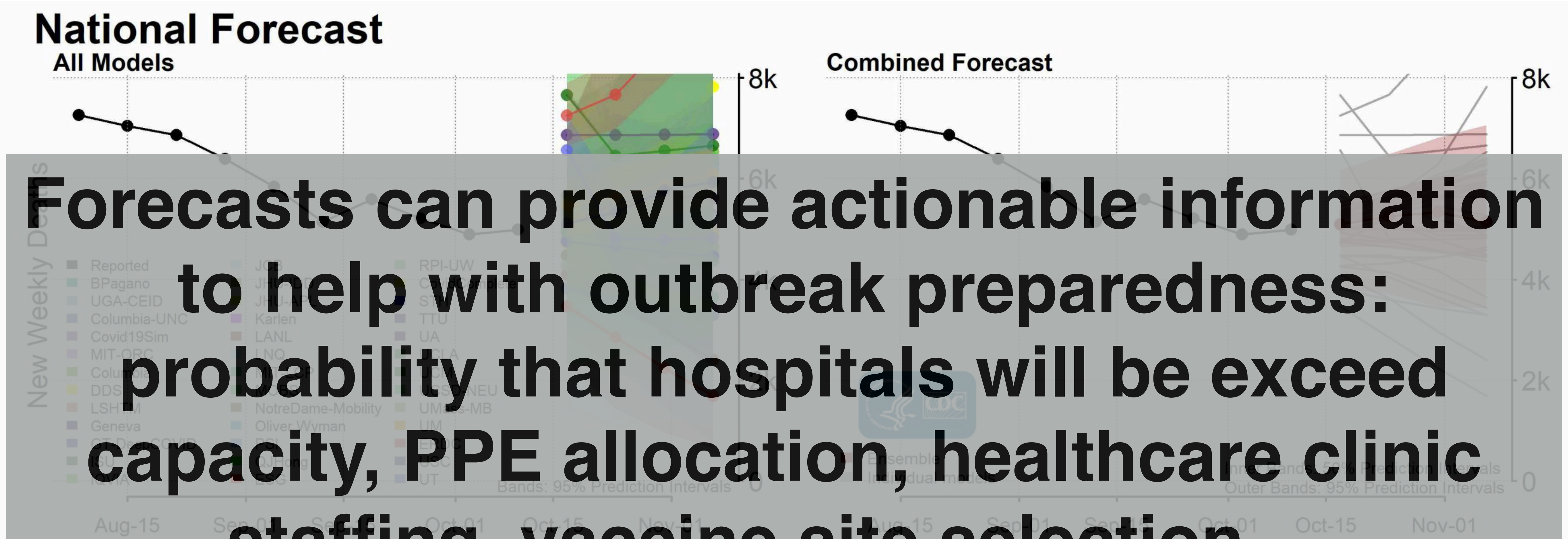


https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html

What can we expect in the next 2-4 weeks?

Short-term Forecasting

Making **falsifiable, evaluable** predictions of observable future quantities.



Forecasts can provide actionable information to help with outbreak preparedness: probability that hospitals will be exceed capacity, PPE allocation, healthcare clinic staffing, vaccine site selection, ...

https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html

Long-term Scenarios

What are the long-term impacts under different scenarios?

Projections based on specific assumptions.

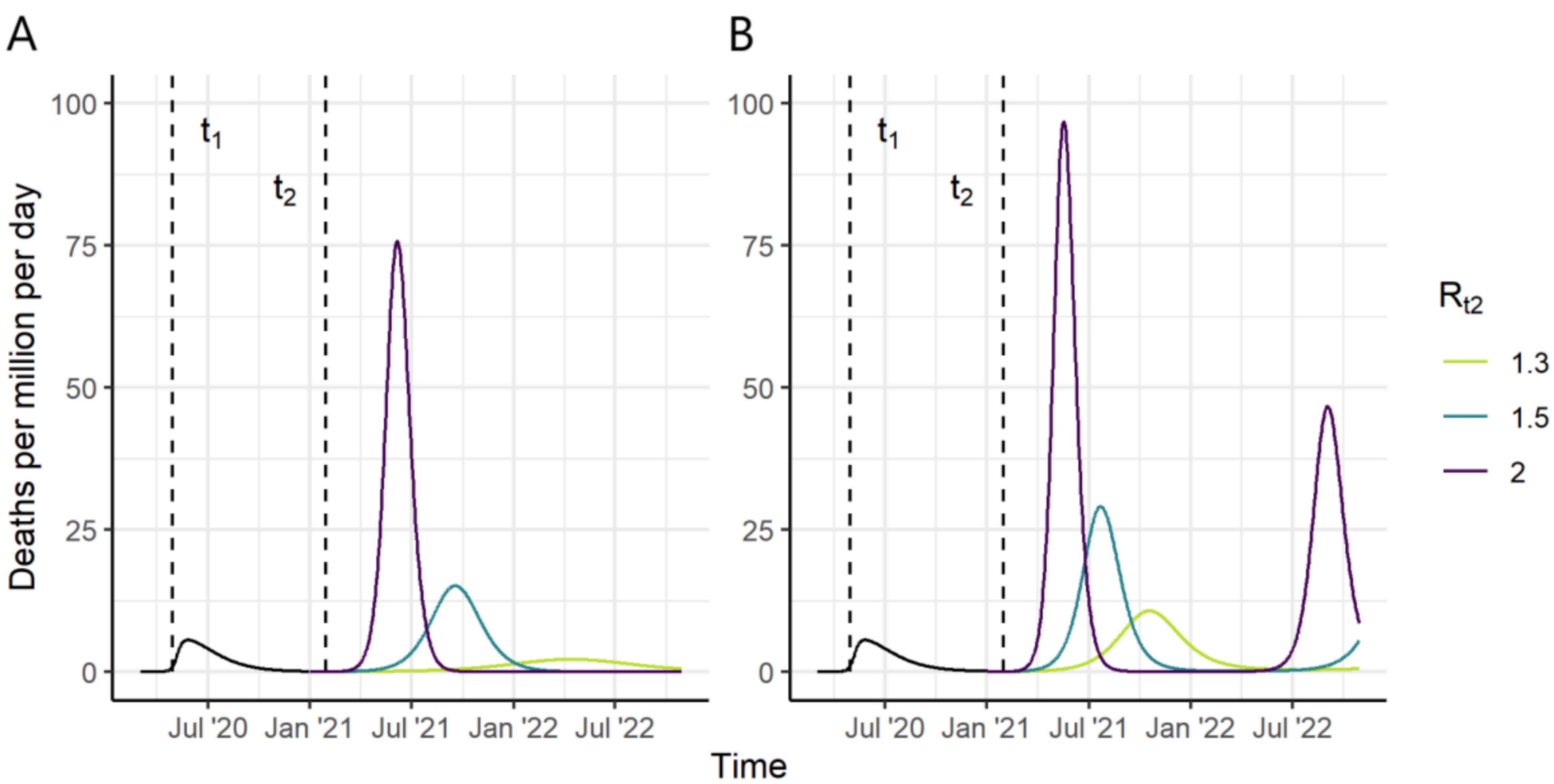


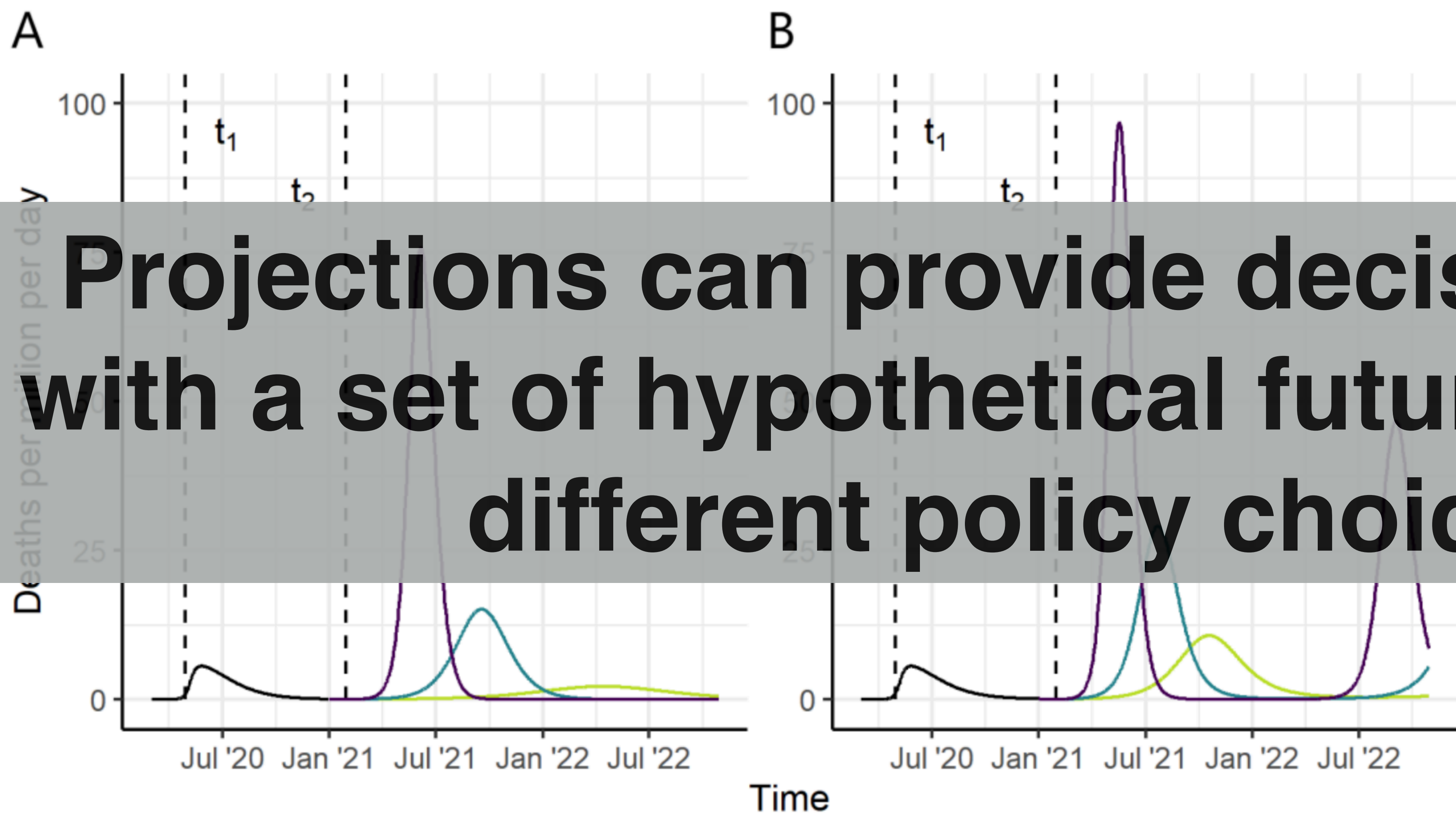
Figure 1: Scenarios for the Course of the Epidemic from 2020–2022, for a High-Income Country Setting, in the Absence of a Vaccine (counterfactual scenarios). (A) Assuming "long immunity" and (B) assuming an average duration of naturally acquired immunity of 1 year. We assume that R_0=2.5 up to time t_1 (May 2020) and that R_{t1}

https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-09-25-COVID19-Report-33.pdf

What are the long-term impacts under different scenarios?

Long-term Scenarios

Projections based on specific assumptions.



Projections can provide decision-makers with a set of hypothetical futures based on different policy choices.

Figure 1: Scenarios for the Course of the Epidemic from 2020–2022, for a High-Income Country Setting, in the Absence of a Vaccine (counterfactual scenarios). (A) Assuming "long immunity" and (B) assuming an average duration of naturally acquired immunity of 1 year. We assume that R_0=2.5 up to time t_1 (May 2020) and that R_{t1}

https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-09-25-COVID19-Report-33.pdf

Nowcasting:
what is under the hood?

slides created with help of Graham Casey Gibson

"Classic" Compartmental Models

- Compartmental models have long been used to model epidemics.

- Most common is the SIR/SEIR model

$$\text{Susceptible} \xrightarrow{\beta I} \text{Exposed} \xrightarrow{\sigma} \text{Infected} \xrightarrow{\gamma} \text{Recovered}$$

https://www.youtube.com/watch?v=CmhL4rVLwn0

Differential Equation Form

- Differential equation representation describes instantaneous rates of flow.

- SIR model is equivalent to assuming that infected individuals become infectious immediately.

$$\frac{dS}{dt} = -\beta \cdot \frac{I}{N} \cdot S$$

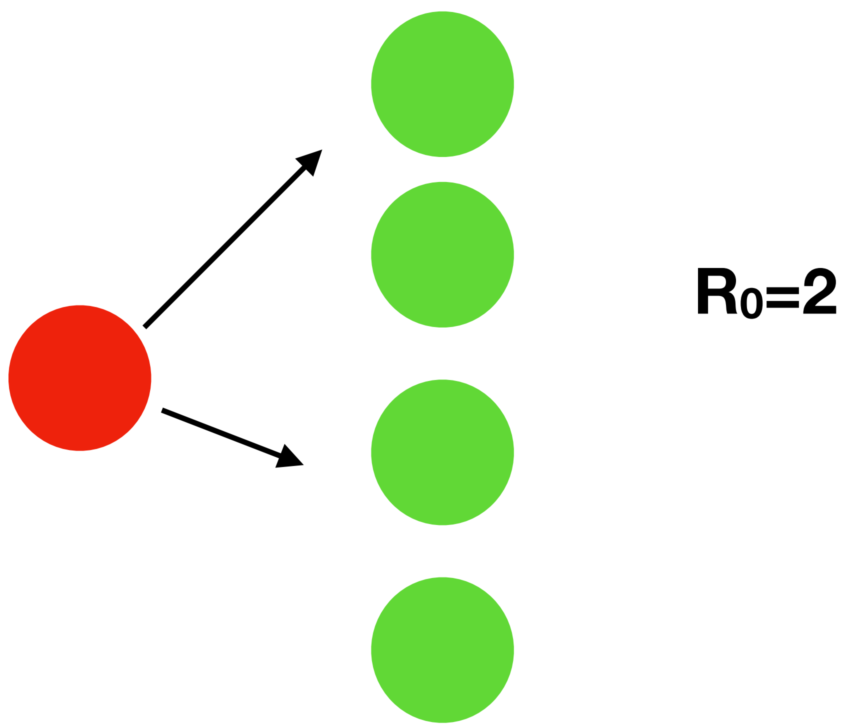
$$\frac{dE}{dt} = \beta \cdot \frac{I}{N} \cdot S - \sigma \cdot E$$

$$\frac{dI}{dt} = \sigma \cdot E - \gamma \cdot I$$

$$\frac{dR}{dt} = \gamma \cdot I$$

The reproduction number

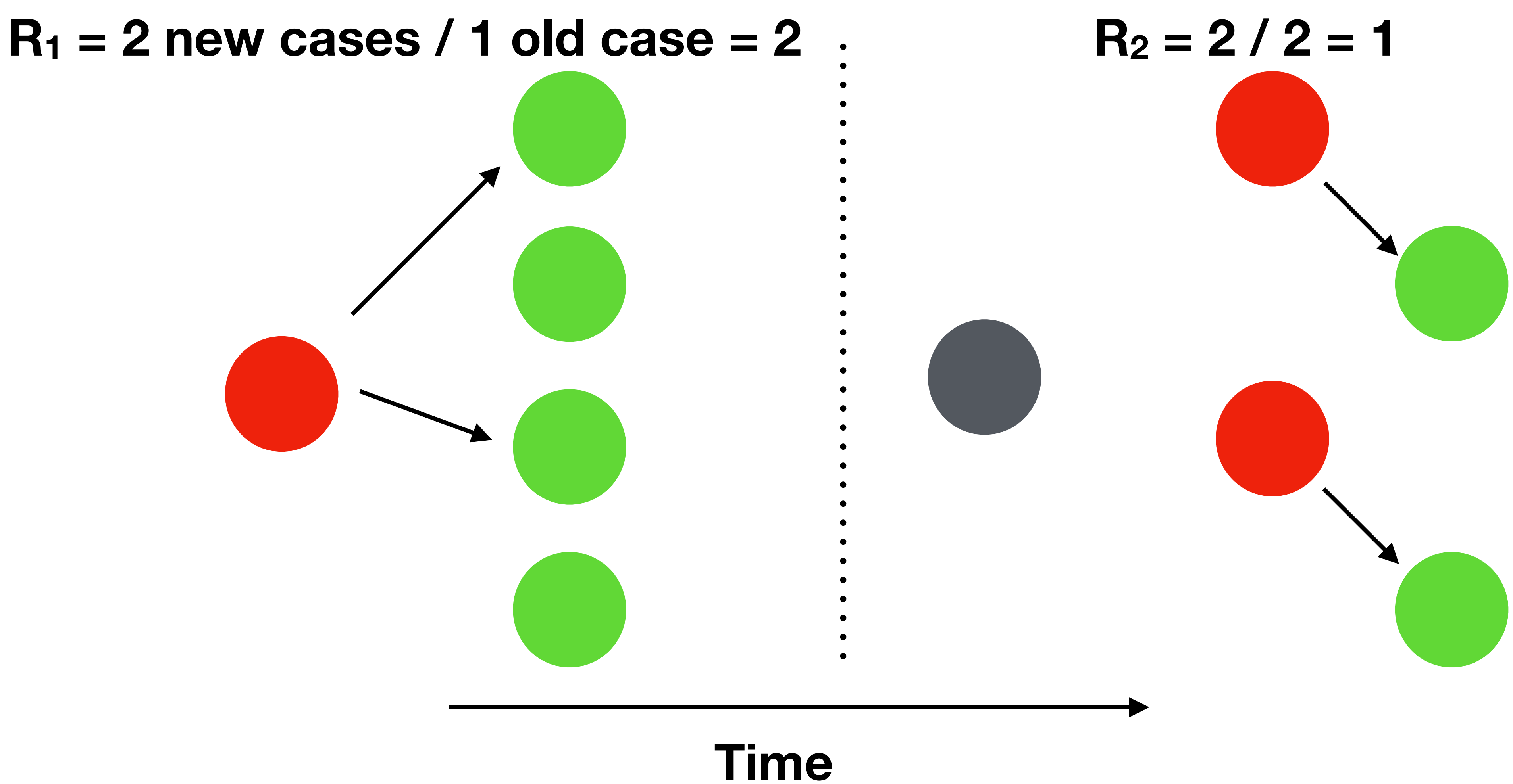
- R_0 (pronounced "R-naught") is the “expected the number of secondary cases one case would produce in a completely susceptible population”.



1. Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res. 1993;2:23–41
2. Fine PEM. Herd immunity: history, theory, practice. Epidemiol Rev. 1993;15:265–302

R_t

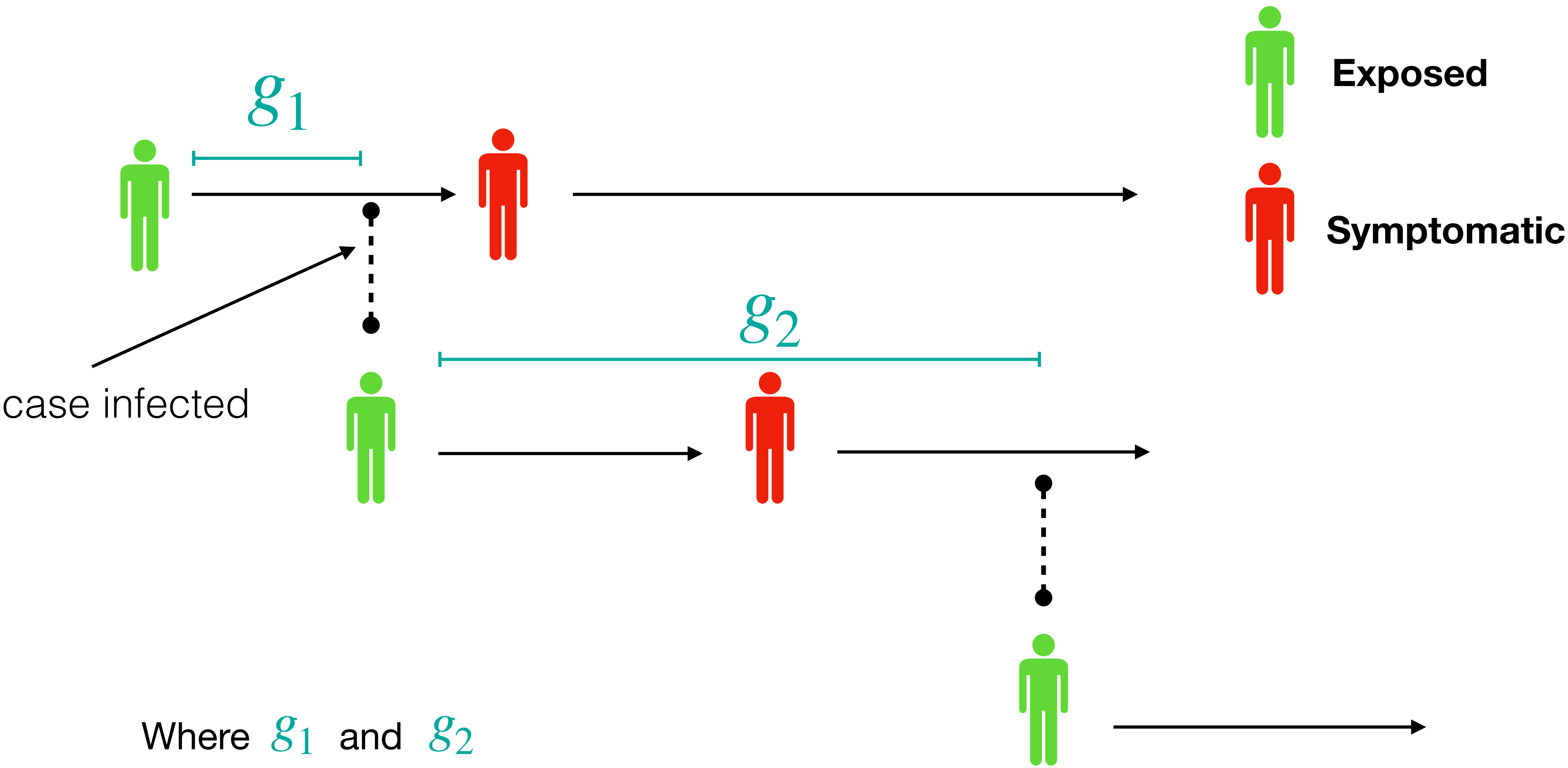
- The average number of secondary cases at time t [1]
- Naturally decreases over time due to decrease of S
- Is a product of biology and behavior.



https://staff.math.su.se/hoehle/blog/2020/04/15/effectiveR0.html

Generation Interval

- R_t relies on the **generation interval: time from exposure of infector to exposure of infectee.**



Where g_1 and g_2

are samples from the generation interval distribution

Renewal Equations

- We can also write the number of new infections at a given time point t using the "renewal style" equation (borrowed from population demography "renewal" of a population from new births).

Time-varying reproduction number

generation interval distribution

$$I_t = R_t \sum_{k=1}^{t} \underline{I_{t-k} g_k}$$

Number of new infections at time t

the expected fraction of infections exposed k time units ago who will generate a new infection at time t.

Intuition: New infections can be written as the fraction of historical infections that produce infections at time t multiplied by how many they produce on average.

R_t Estimator

- We can simply re-arrange the renewal style equation to obtain an estimator for R_t

$$\hat{R}_t = \frac{I_t}{\sum_k I_{t-k} g_k}$$

Intuition: The average number of secondary infections can be written as the observed new infections divided by the historical infections that produced at least one secondary infection at time t.

Cori, Ferguson, Fraser and Cauchemez. AJE, 2013. https://doi.org/10.1093/aje/kwt133

In Practice

- The naive estimator shown in the previous slide suffers from a variety of issues
    - Instability in small sample sizes
    - Sensitive to reporting variation
    - Does not take into account imported cases ("local" cases only)
- "Cori estimator" solves these issues via Bayesian estimation, implemented in R.

Generation interval

$$\mathbf{P}(R_t \mid I_0, I_1, I_2, \ldots, I_{t-\tau-1}, I_{t-\tau}^{\text{local}}, I_{t-\tau+1}^{\text{local}}, \ldots, I_t^{\text{local}}, w_s)$$

$$\propto \mathbf{P}(I_{t-\tau}^{\text{local}}, I_{t-\tau+1}^{\text{local}}, \ldots, I_t^{\text{local}} \mid I_0, \ldots, I_{t-\tau-1}, w_s, R_t)\mathbf{P}(R_t)$$

Only estimate Rt on local transmission

Smoothing window size

https://cran.r-project.org/web/packages/EpiEstim/index.html

Forecasting:
what is under the hood?

COVID-19 Forecast Hub: Background

- Each week the Hub receives forecasts of weekly incident and cumulative deaths and incident cases in the US due to COVID-19 from over 50 teams.

- The Hub builds an ensemble that combines predictions from these models for 1 through 4 week ahead forecasts.

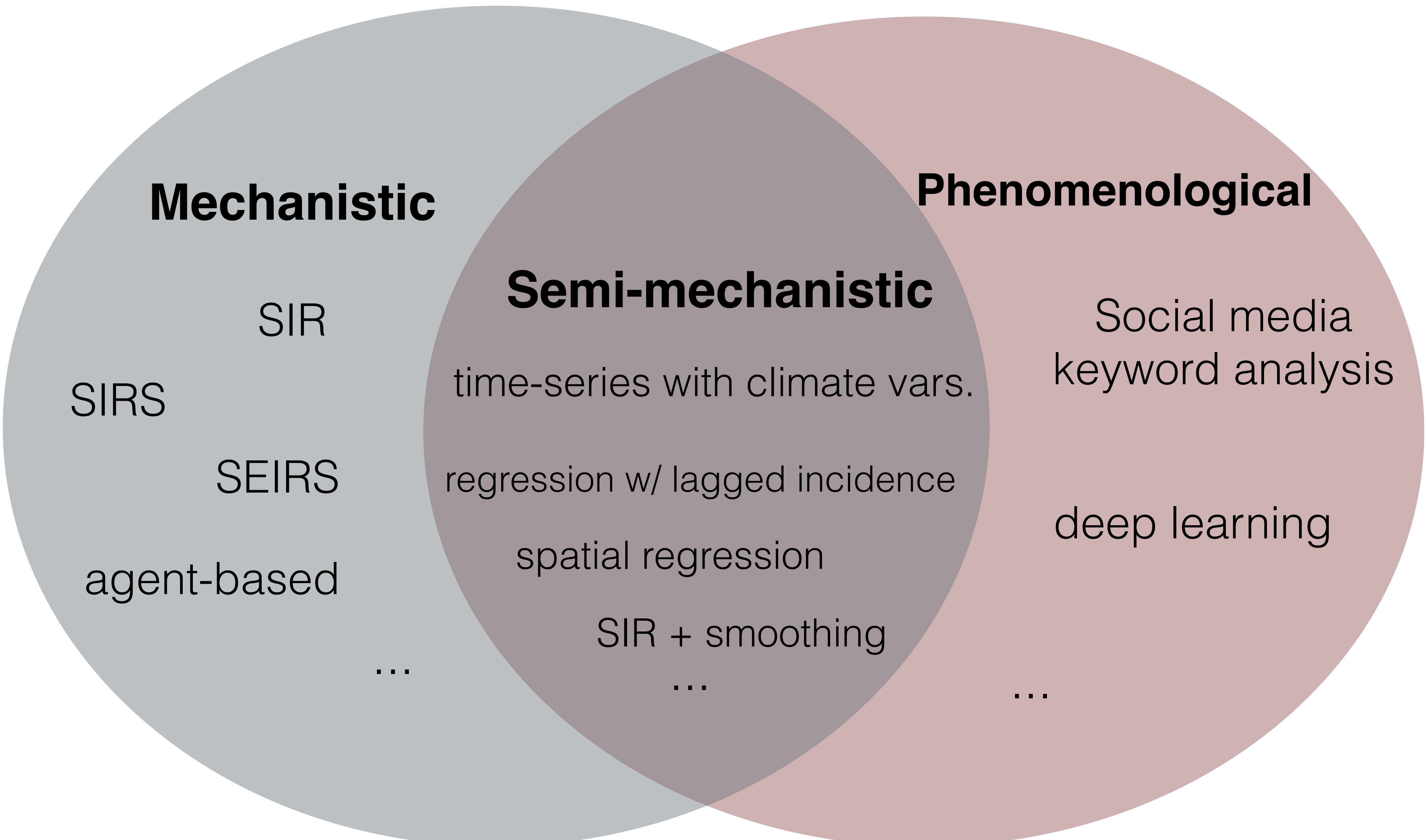
COVID-19 ForecastHub

https://covid19forecasthub.org/

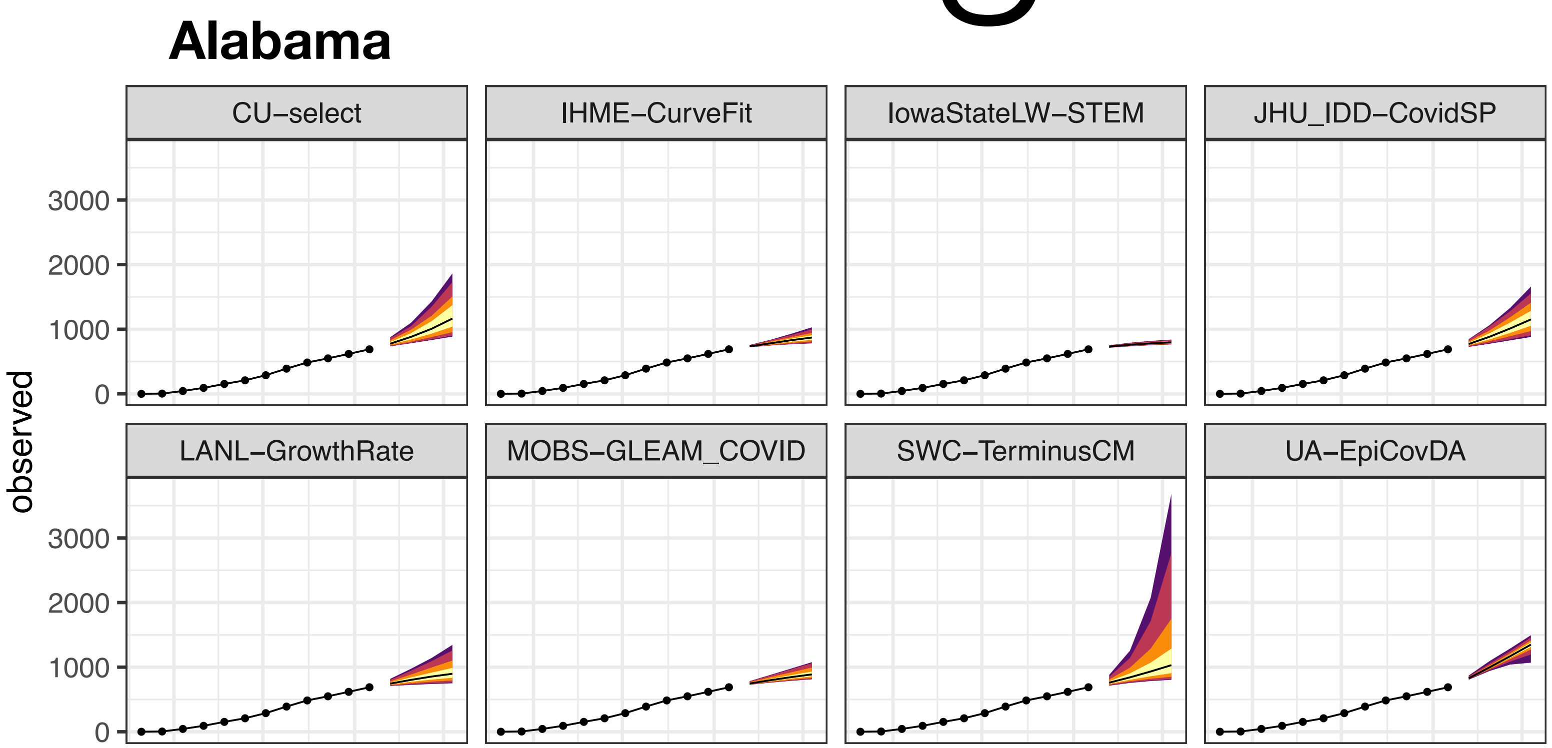
Modeling approaches vary

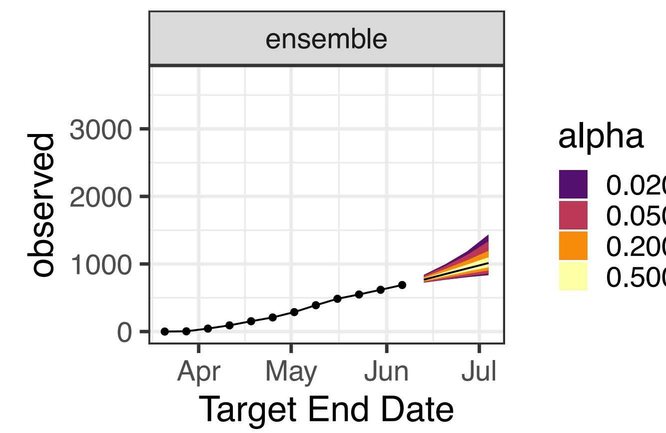
- YYG-ParamSearch: "**machine learning** techniques on top of a **classic infectious disease model** to make projections for infections and deaths."
- UMass-MechBayes: "**classical compartmental models from epidemiology**, prior distributions on parameters, models for time-varying dynamics, models for partial/noisy observations of confirmed cases and deaths."
- UCLA-SuEIR: "an improved **SEIR model** for predicting the dynamics among the cumulative confirmed cases and death of COVID-19"
- IHME-CurveFit: "**hybrid modeling approach** to generate our forecasts, which incorporates elements of statistical and disease transmission models."
- MOBS-GLEAM_COVID: "The GLEAM framework is based on **a metapopulation approach** in which the world is divided into geographical subpopulations. Human **mobility between subpopulations is represented on a network**."
- UT-Mobility: "For each US state, **we use local data from mobile-phone GPS traces** made available by [SafeGraph] to quantify the changing impact of social-distancing measures on 'flattening the curve.' "
- GT-DeepCOVID: "This **data-driven deep learning model** learns the dependence of hospitalization and mortality rate on various detailed syndromic, demographic, mobility and clinical data."

ID Epi Prediction Model Taxonomy



Building the Ensemble





- Teams are required to submit 23 quantiles of a predictive distribution:

$$\widehat{P}(Y \leq q_1) = 0.01,\ \widehat{P}(Y \leq q_2) = 0.025,\ \ldots,\ \widehat{P}(Y \leq q_{12}) = 0.5,\ \ldots,\ \widehat{P}(Y \leq q_{23}) = 0.99$$

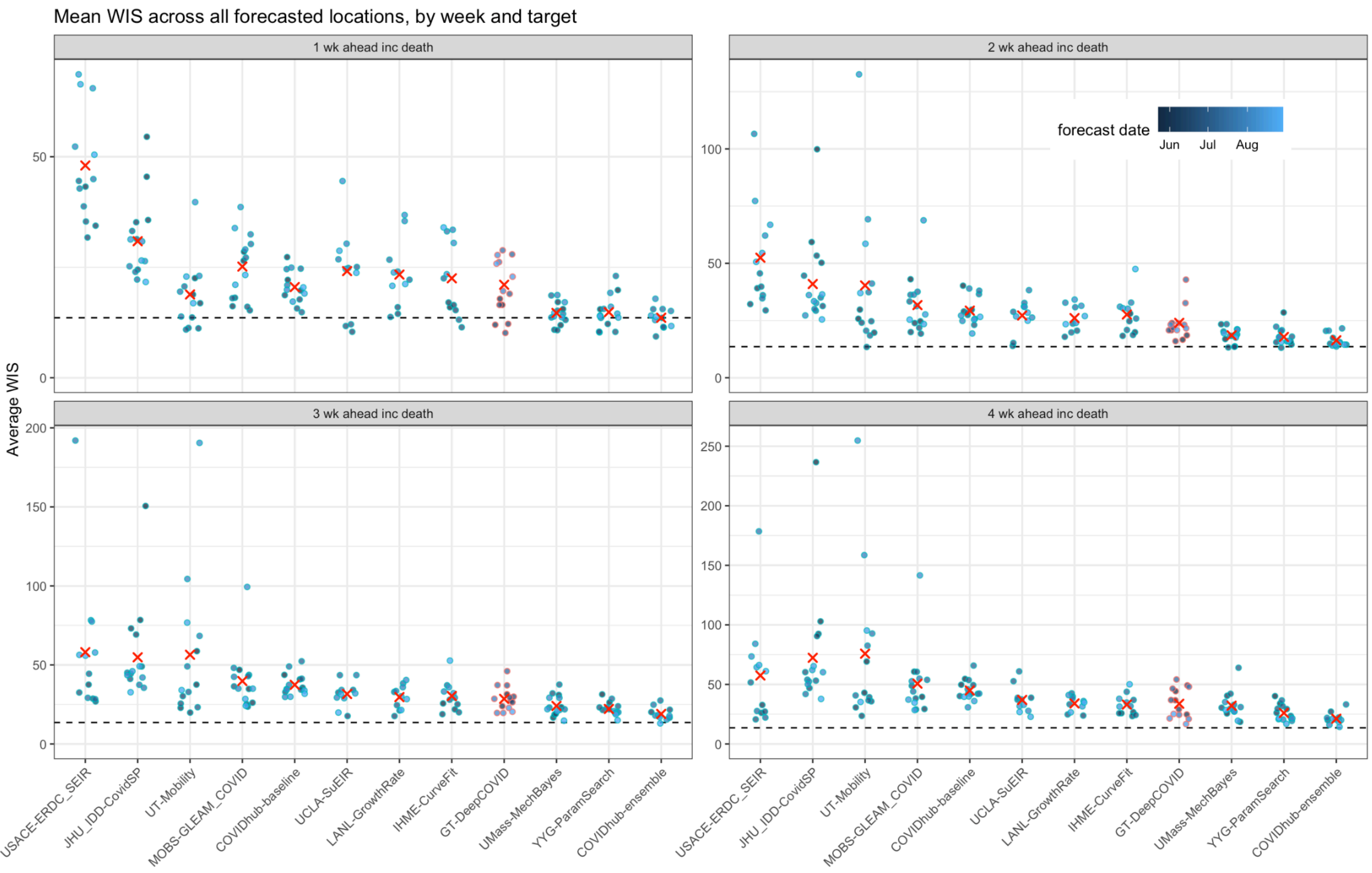
The predictive median

Limits of a 98% prediction interval

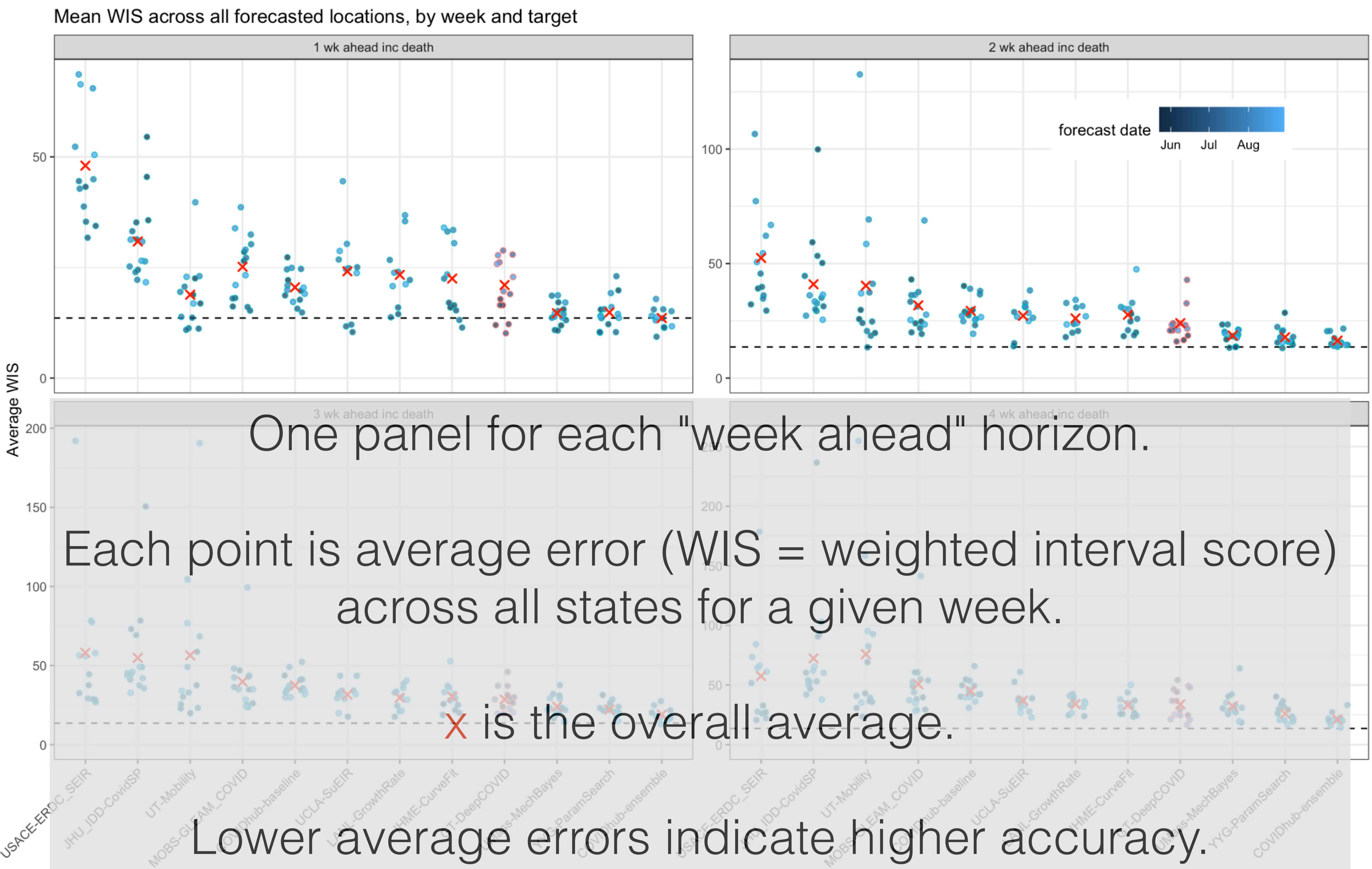
- At each quantile level, the ensemble combines the predictions from all models:
  - Before July 28, we used the simple average across all models, after manually screening unreasonable forecasts
  - From July 28 on, we have used the median without subjective screening

1. ensembles are robust

Ensemble is most accurate



Ensemble is most accurate



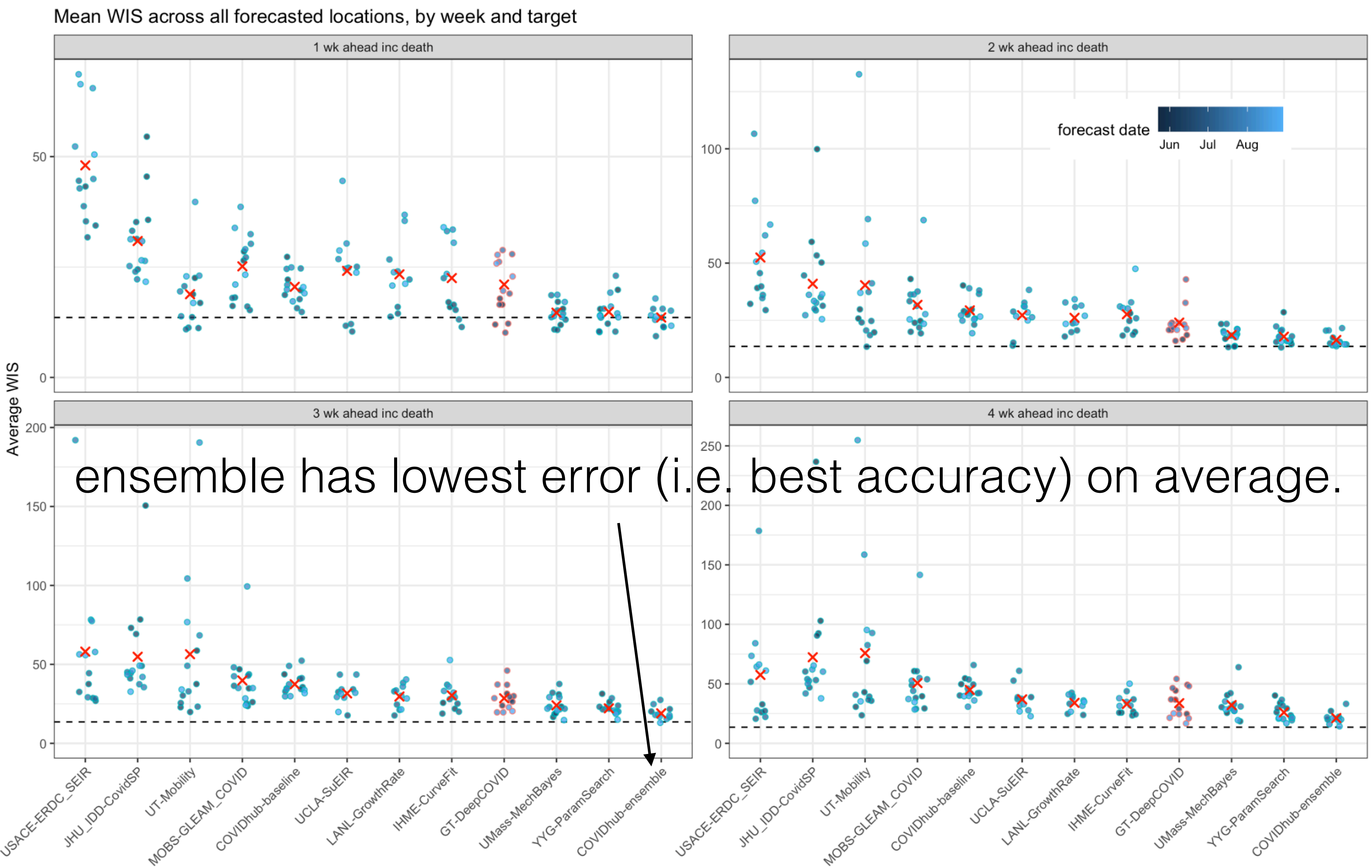
One panel for each "week ahead" horizon.

Each point is average error (WIS = weighted interval score) across all states for a given week.

X is the overall average.

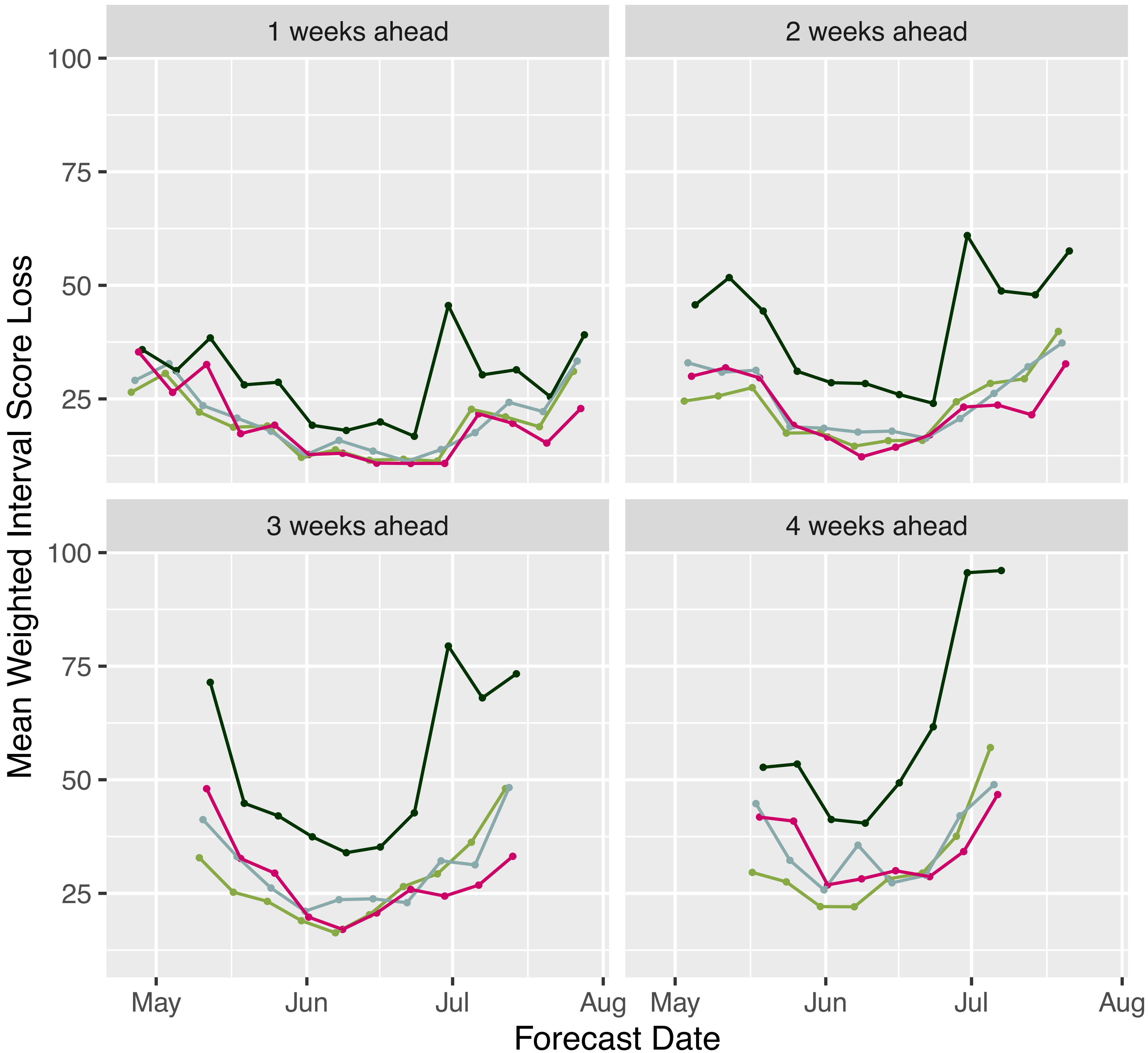
Lower average errors indicate higher accuracy.

Ensemble is most accurate



2. complexity not needed

Simple ensembles do well



- All ensembles improve over a baseline (dark green).
- Ensemble with weights that are based on past performance (pink line) are similar to ensembles that combine all models with equal weight (light green lines).

Ongoing challenges

Challenge 1: Data sparsity

(infectious disease dynamics cannot be observed like the weather)

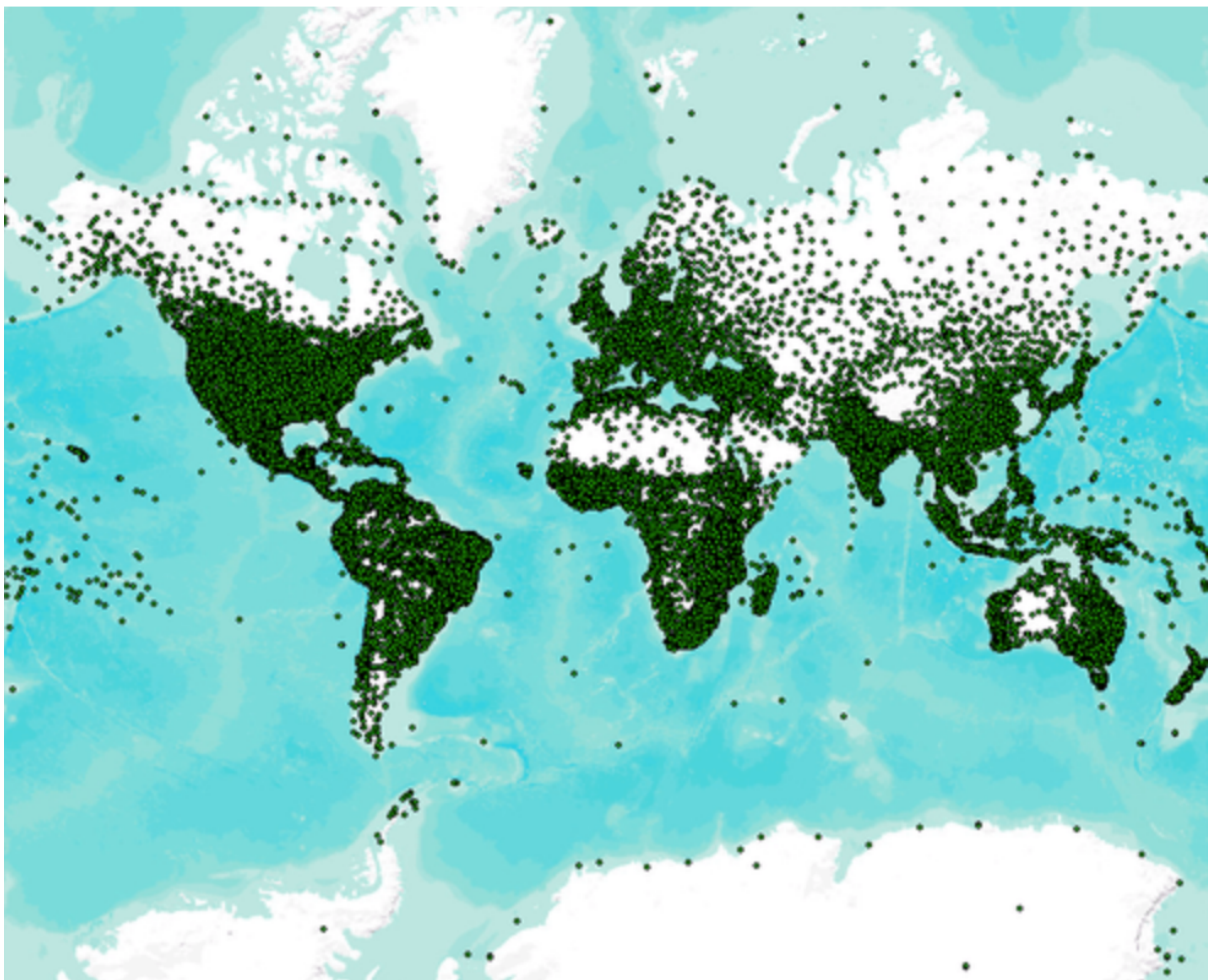


image credit: https://goo.gl/images/CSSQRv

Each dot represents
a weather station whose
data was used to create
the WorldClim dataset.

image credit: https://databasin.org/datasets/15a31dec689b4c958ee491ff30fcce75

Challenge 2: Feedback loop

- Weather forecasts can't change the weather.
- An outbreak forecast could change an outbreak.

US military troops heading to Liberia to assist with Ebola outbreak.
image: defense.gov

Images of vector-control activities to control dengue in Thailand
courtesy of Sopon Iamsirithaworn, Thailand Department of Disease Control

Challenge 3: Translation into action

Dan Jernigan, Director of Influenza Division, CDC
September 2018

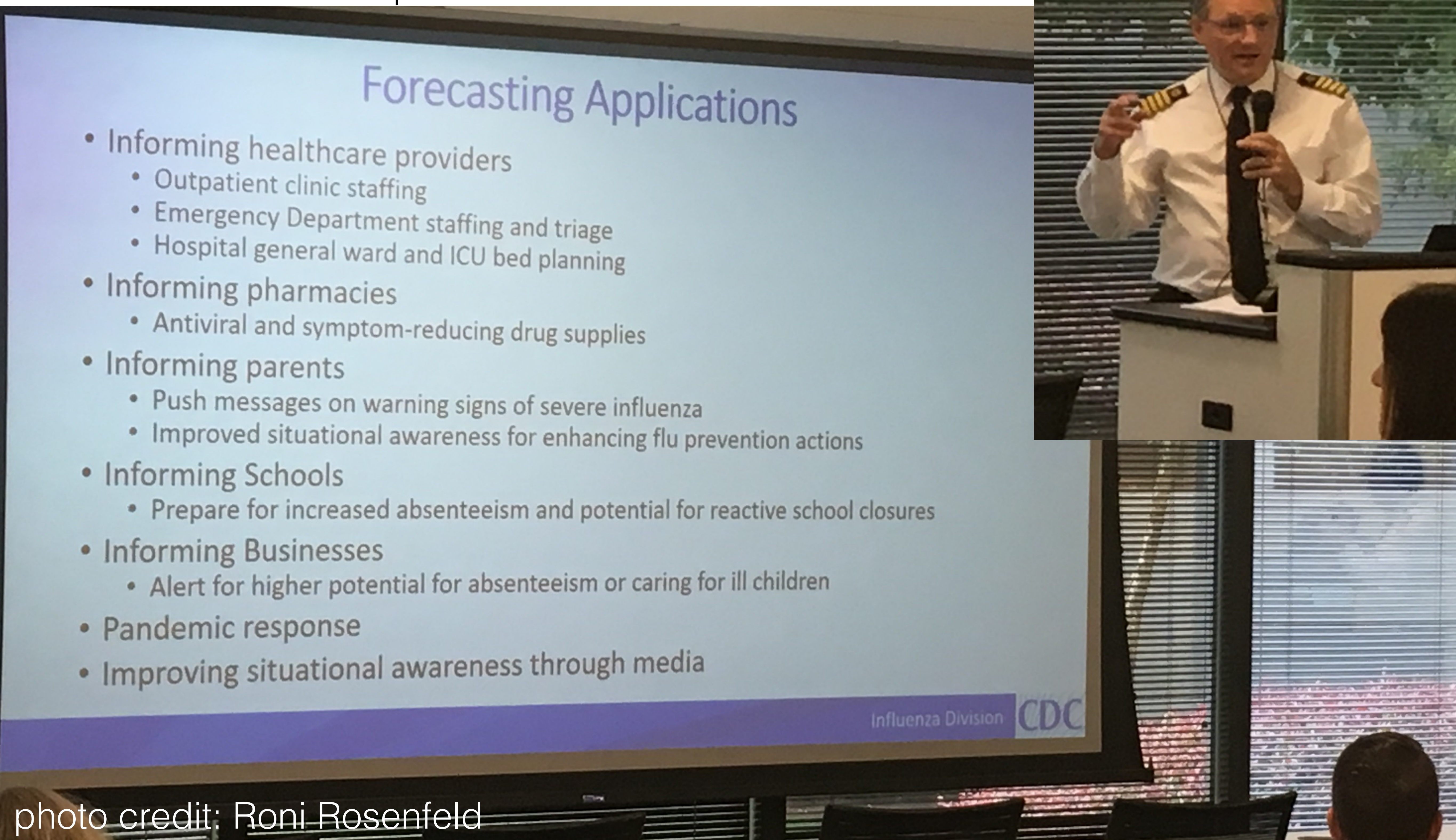


photo credit: Roni Rosenfeld

What is the appropriate role for these models in outbreaks?

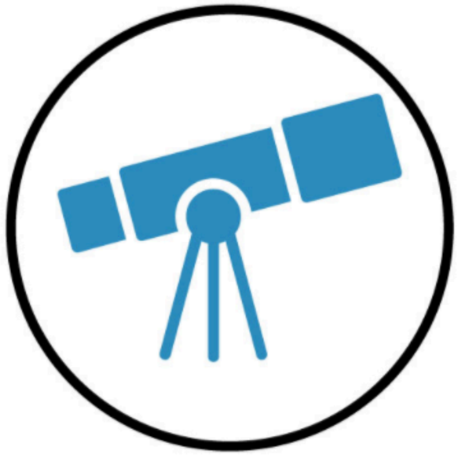
Nowcasts

How fast is COVID-19 spreading right now?

Forecasts

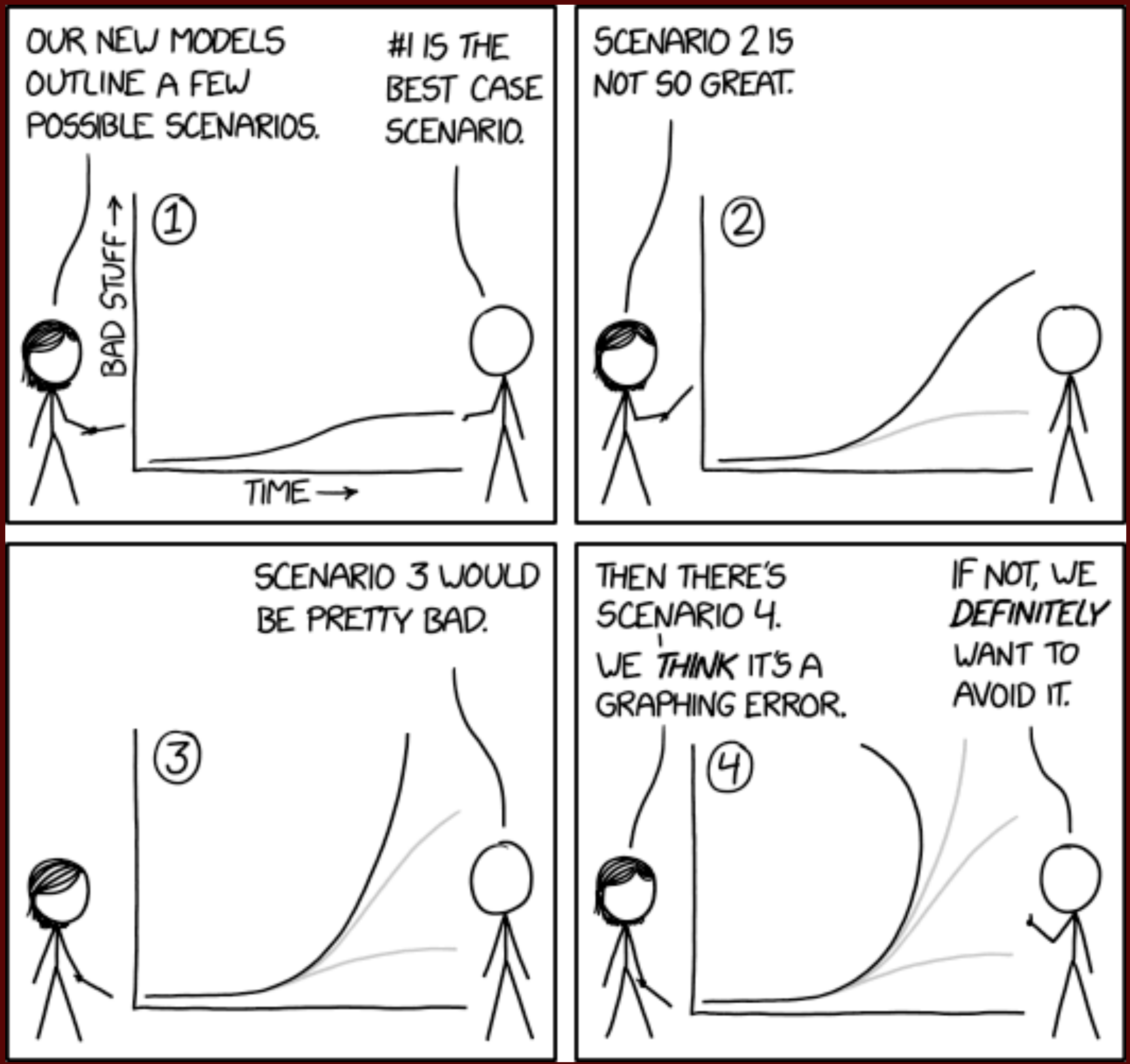
What can we expect in the next 2-4 weeks?

Scenarios



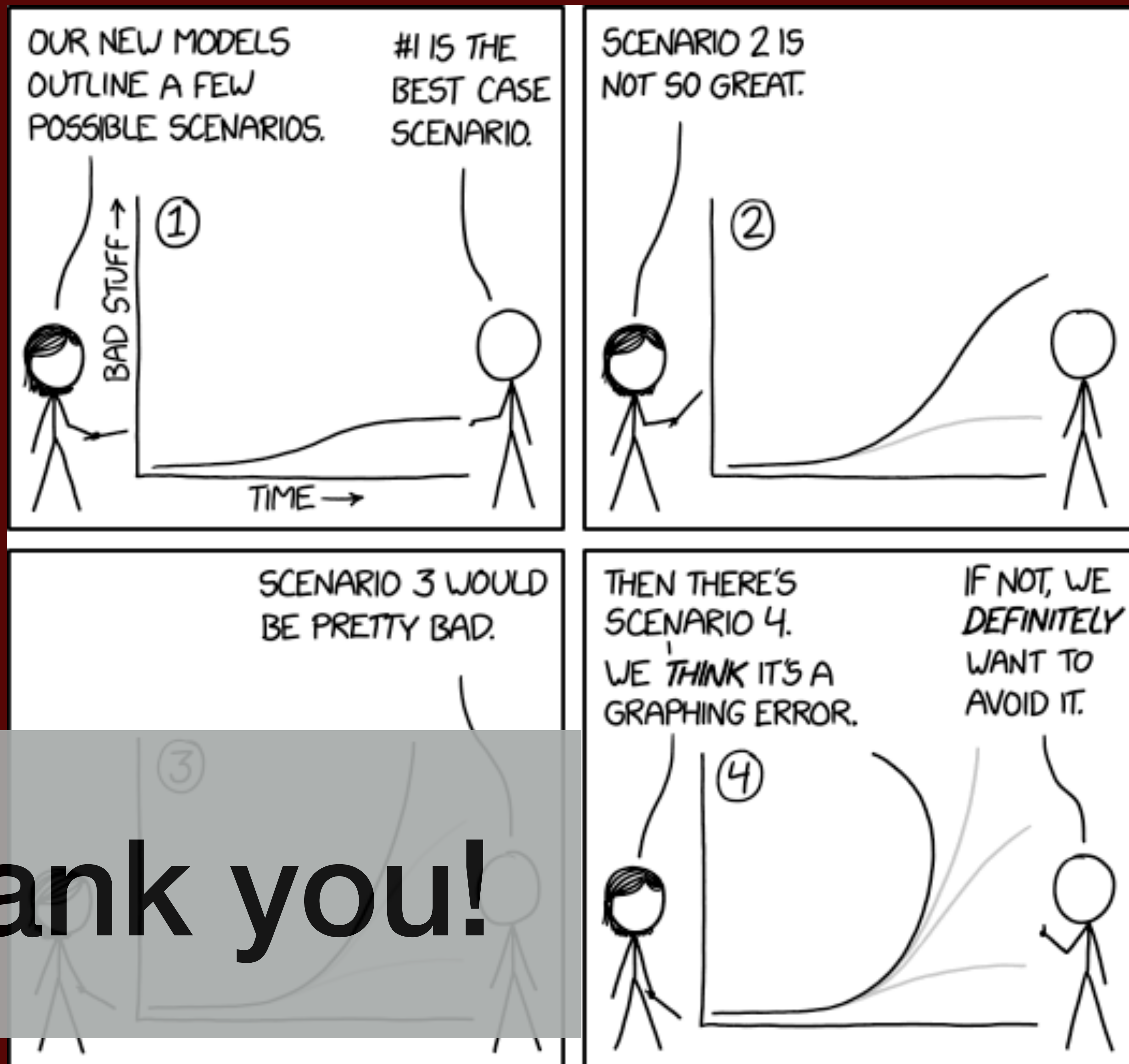
What are the long-term impacts under different scenarios?

We have lots of work to do!



https://xkcd.com/2289/

We have lots of work to do!



Thank you!

https://xkcd.com/2289/