Tools for real-time nowcasting and R_t estimation

Katie Gostic, PhD

Modeling and data science lead

CDC Center for Forecasting and Outbreak Analytics

CSTE Infectious Disease Forecasting Workshop

2023-06-25

Thank you to our collaborators on the EpiNow2 and epinowcast teams

Sam Abbott, PhD

London School of Hygiene and Tropical Medicine

CDC Center for Forecasting and Outbreak Analytics

epinowcast

Sam Abbott
Adrian Lison
Sebastian Funk
Carl Pearson
Hugo Grunson
Hannah Choi
Pietro Monticone
Felix Guenther
Pratik Gupte
Luis Rivas
… and many community contributors

EpiNow2 1.3.5

Sam Abbott
Joel Hellewell
Katharine Sherratt
Katelyn Gostic
Joe Hickson
Hamada S. Badr
Michael DeWitt
Sebastian Funk
… and many community contributors

Take-home messages

1. Epidemic data inevitably contain lags, reporting effects, and other imperfections. This makes real-time analysis challenging.

2. Methods and tools designed to deal with predictable problems in the data are developing rapidly. But expert judgement is needed to choose the right tools and interpret the outputs.

3. CDC is working to support real-time epidemic analysis through
   - Funding for collaborations and growth of the talent pipeline
   - Open-source tool development
   - Testing tools on historical outbreak data → continuous improvement
   - Full-time team to produce and publish analyses at the state and territory level

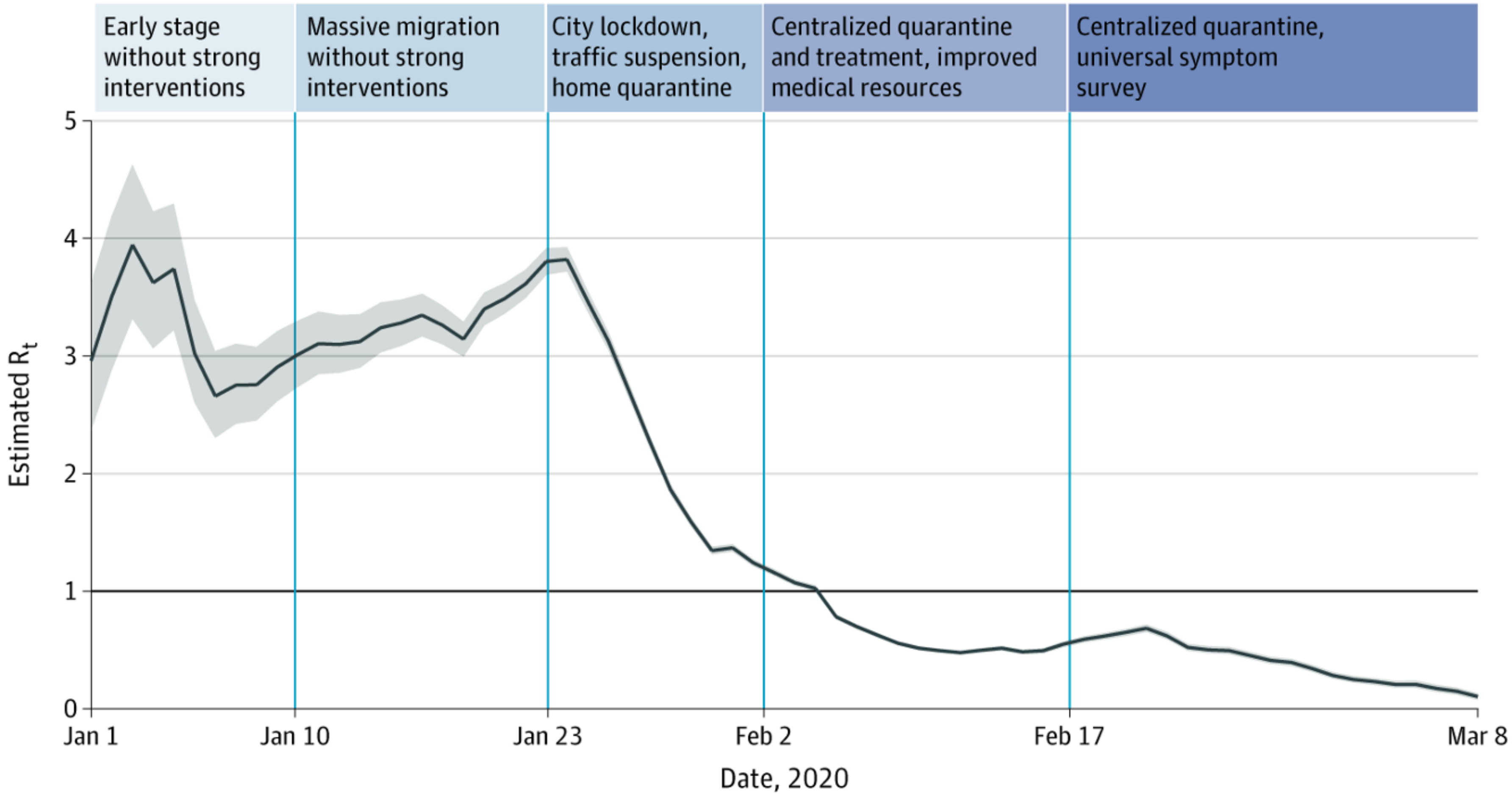
Agenda

R_t estimation with nowcasting

1. Case study: R_t estimation early in the COVID-19 pandemic
2. R_t definition and estimation
3. Pitfalls when estimating R_t
    - Observation delays
    - Day-of-week effects and reporting artifacts
    - Partial observation at the end of the time series
4. Tools designed to avoid these pitfalls: EpiNow2, and epinowcast

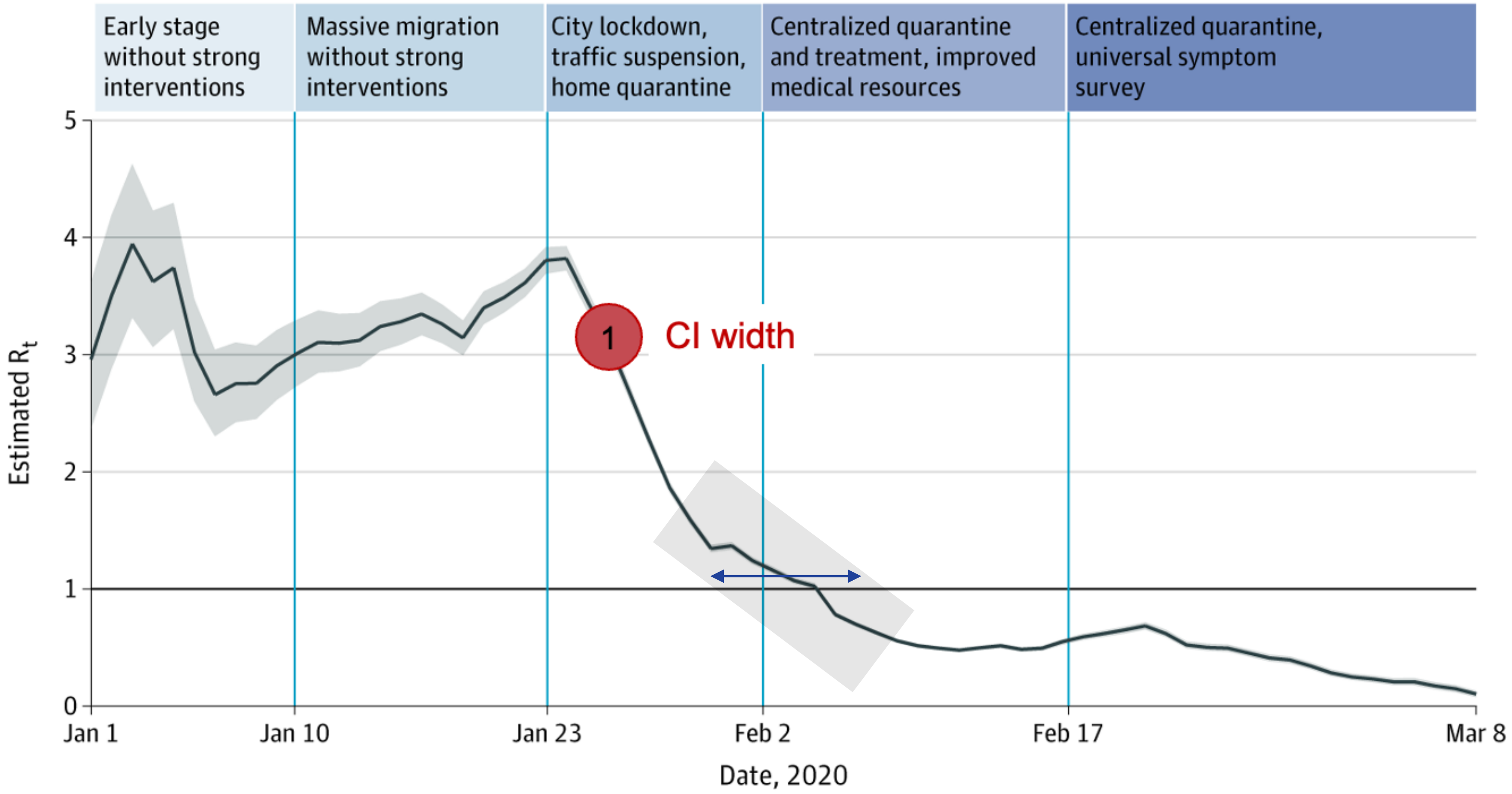
1. Case study: R_t estimation early in the COVID-19 pandemic

Case study: R_t estimation in practice



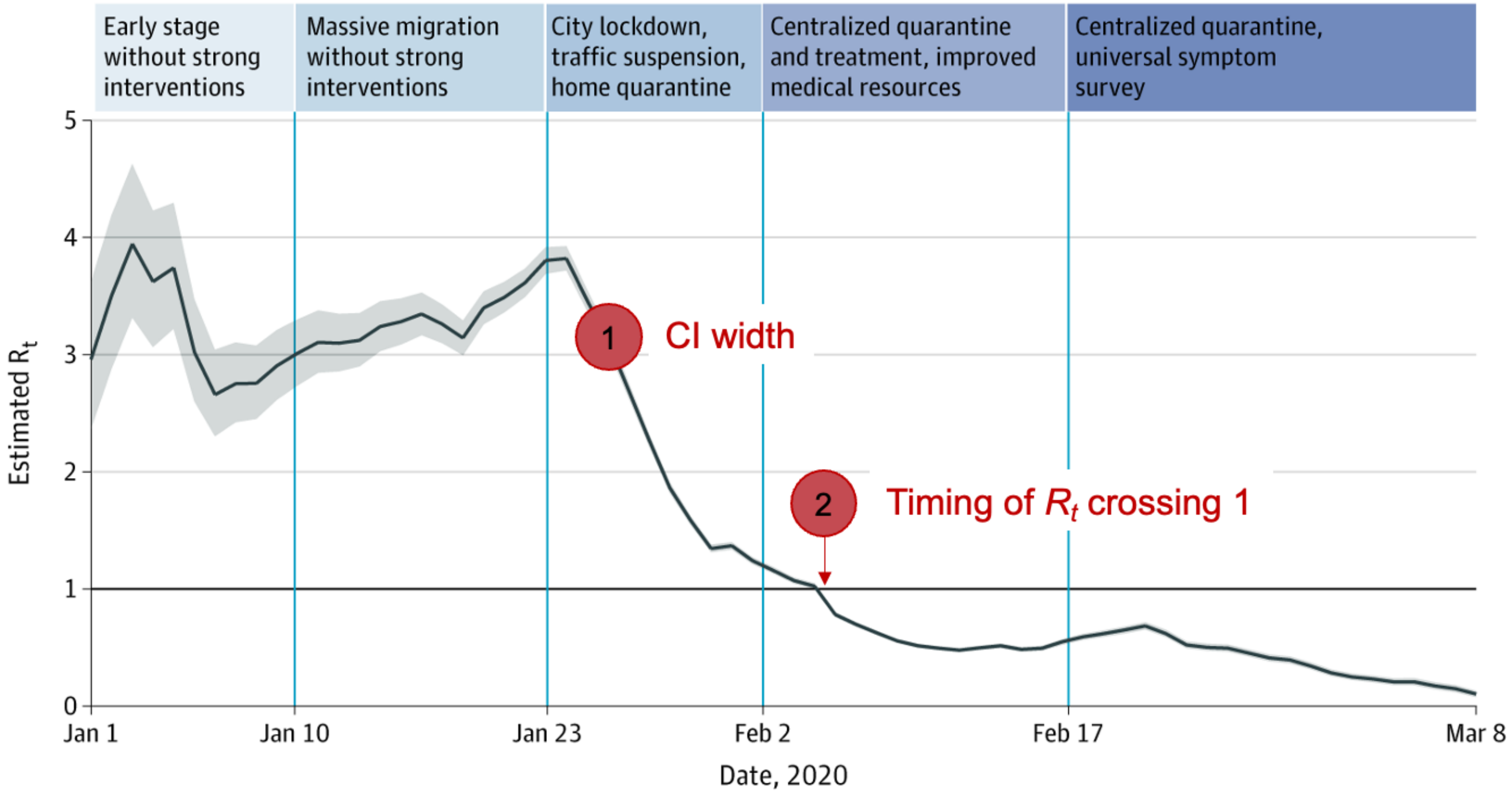
JAMA. 2020;323(19):1915-1923. doi:10.1001/jama.2020.6130

Estimates must be accurate, timely, and account for uncertainty



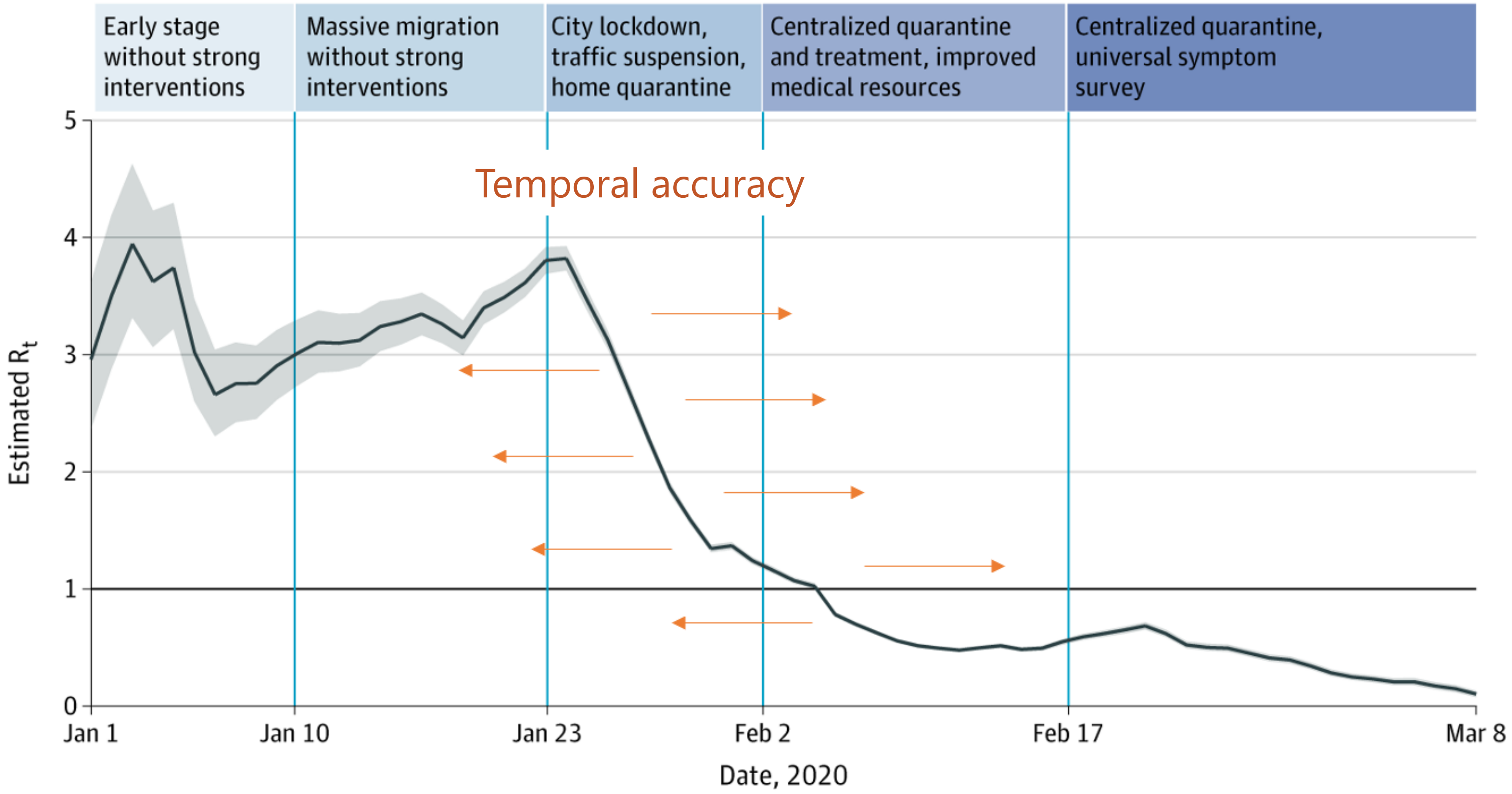
JAMA. 2020;323(19):1915-1923. doi:10.1001/jama.2020.6130

Estimates must be correctly oriented to the calendar



JAMA. 2020;323(19):1915-1923. doi:10.1001/jama.2020.6130

Temporal inaccuracy can change our interpretation

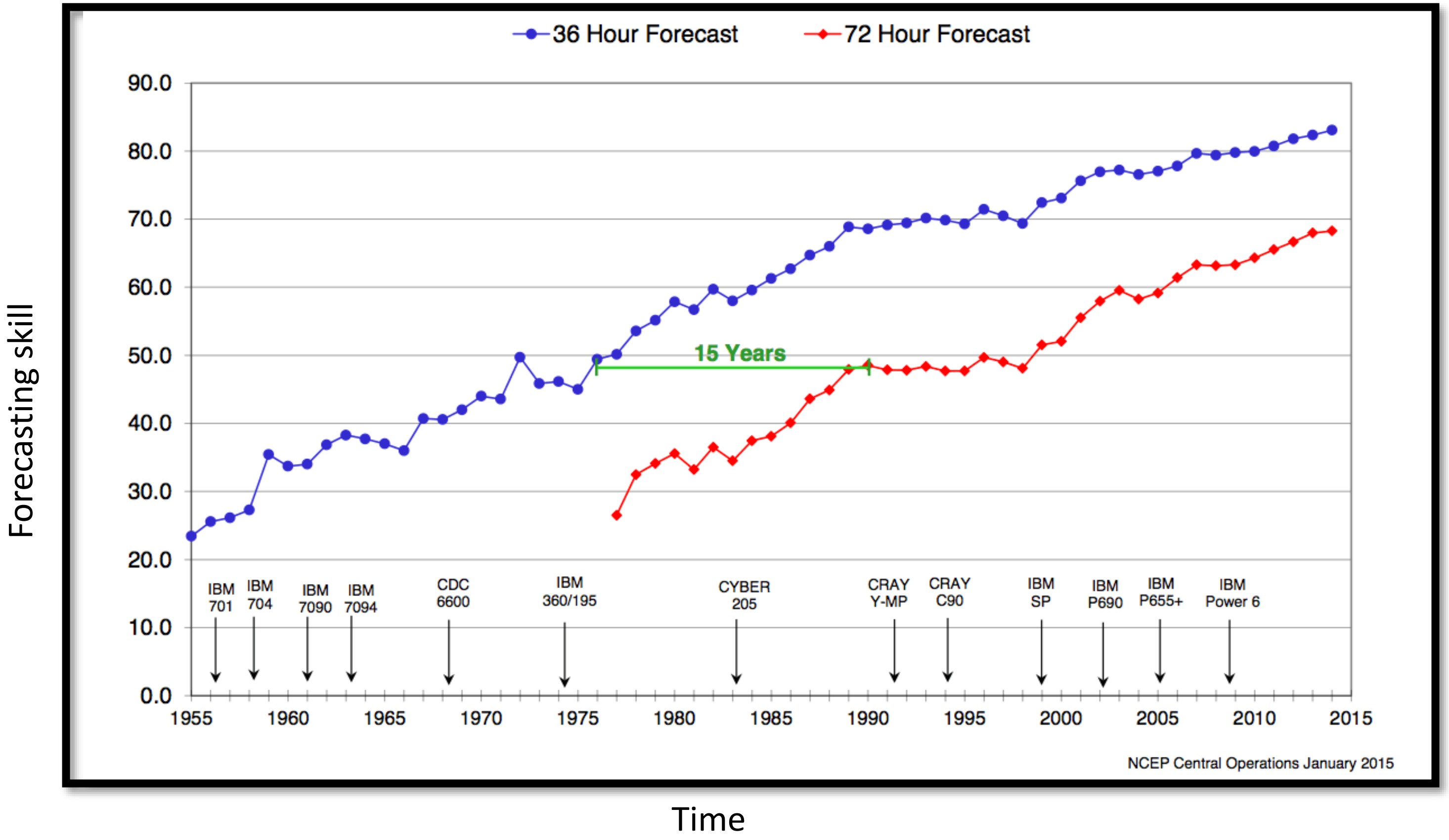


JAMA. 2020;323(19):1915-1923. doi:10.1001/jama.2020.6130

Developing Outbreak Analytics & Forecasting Takes Time and Resources

Advancing Weather Forecasting:
took time, data, models, and resources to develop

Increases in Numerical Weather Forecasting Skill Through Time



- Advancing weather forecasting capabilities took decades
- Needed ingredients:
    - Data
    - Models
    - People
    - Computational Power
    - Specific Use Cases
    - Sustained Funding
- Disease forecasting and analytics are still in early stages

R_t estimation methods have improved rapidly since 2019

But we still have a ways to go…

Pre COVID-19

Methods not designed to deal with predictable features of outbreak data, including lags and backfill.

→ Potential for temporal inaccuracy

→ Credible intervals too narrow

2020-2021

Rapid development of new tools, methods

e.g. EpiNow2, epidemia, packages

→ Large steps forward. Designed to deal with predictable problems in the data

→ But can be slow, inflexible

→ Some design features make it hard to incorporate rapidly evolving statistical methods

2022-present

Refinement of COVID-era tools

epinowcast package

→ Will replace EpiNow2

→ Provides speed and new flexibility, e.g. for geographic or demographic covariates

→ But still in development. Not yet user-friendly

The future

Continued improvement

→ Tools will become more accurate, flexible, and easier to use

→ Documentation and training will mature

- CDC is supporting this work through collaborations and open source tool development
- We are testing existing tools and identifying areas for improvement
- We are collaborating with the epinowcast team

2. R_t

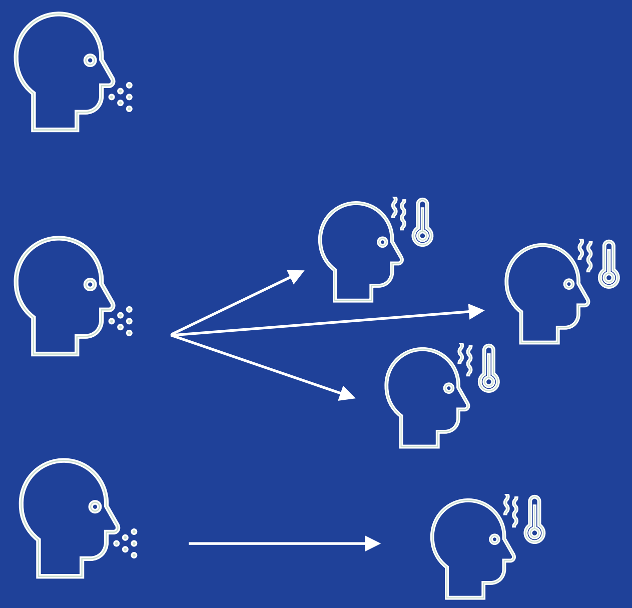
Interpretation and estimation

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

R_t: the time-varying effective reproductive number

- Defined as the average number of new infections caused by a single infectious individual at a specific time
- Accounts empirically for population susceptibility, interventions, and behavior
- Estimated from a time series of counts (e.g. cases, deaths, or hospitalizations per day)

If $R_t > 1$, the epidemic is growing



$$R_t = \frac{4 \text{ infectees}}{3 \text{ infectors}} = 1.33$$

R_t is a data-driven quantity. If we could directly observe every transmission chain, we could identify all the individuals who are infectious at a particular time and count the number of new infections they caused.

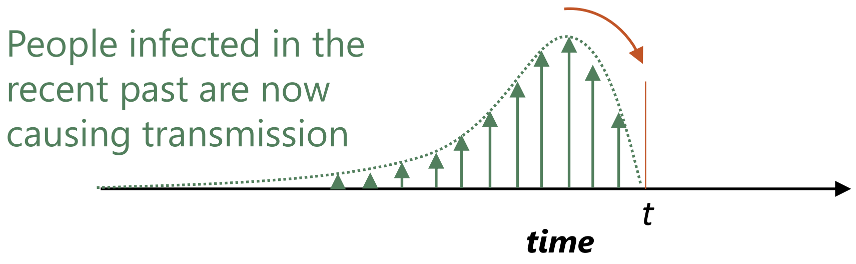
Estimating R_t from a time series of counts

Infectees
The number of infections observed on day *t*

$$R_t = \frac{I_t}{\sum_{s=1}^{t} I_{t-s}\,\omega(s)}$$

Infectors
Sum up the number of individuals infected *s* days prior to time *t*, scaled by their current infectiousness

Arrows show times at which **infectors** and their **infectees** appear in the data.



The generation interval, $\omega(s)$, scales the infectiousness of individuals infected s days ago

Inputs needed

1. A time series of incident events
2. An estimate of the generation interval distribution

$$R_t = \frac{\text{total infectees}}{\text{total infectors}}$$

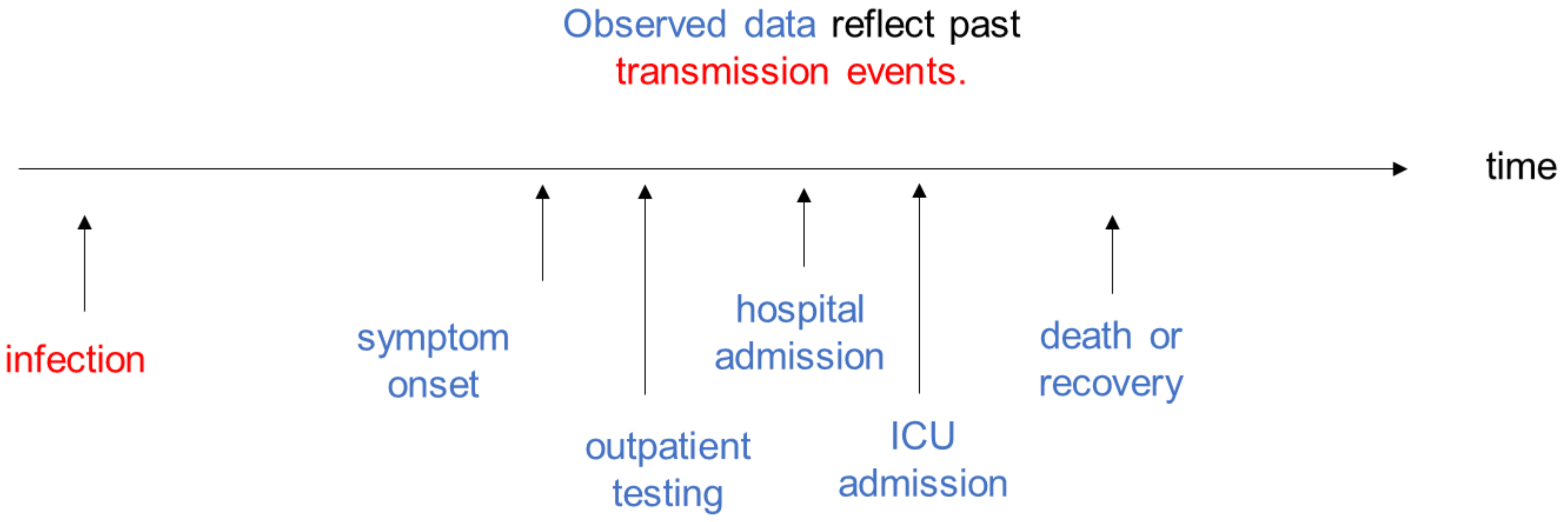
3. Pitfalls and how to avoid them

2.1 Observation delays

2.2 Day of week effects and other reporting anomalies

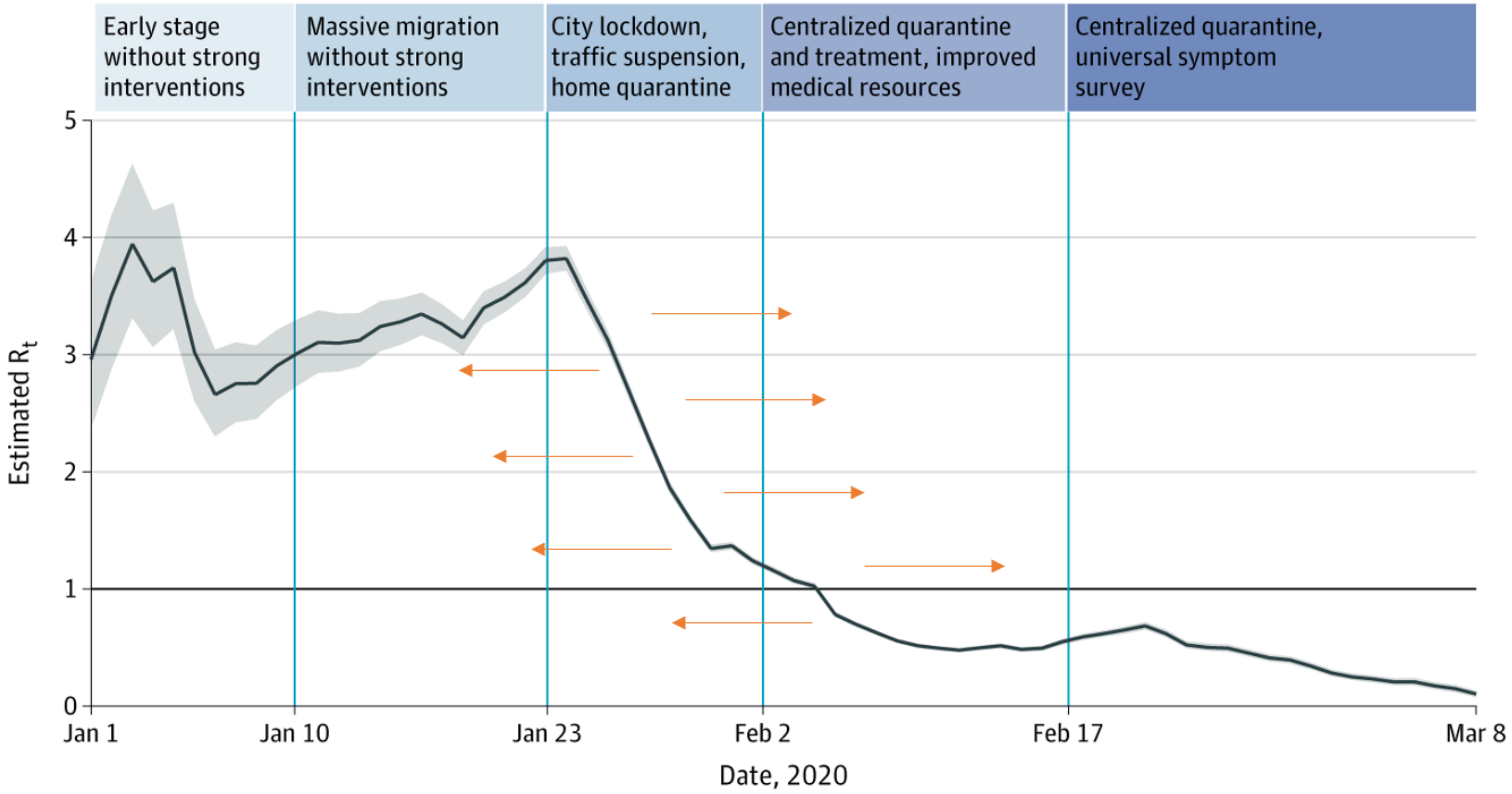
2.3 Partial observation at the end of the time series

Problem: Observation delays



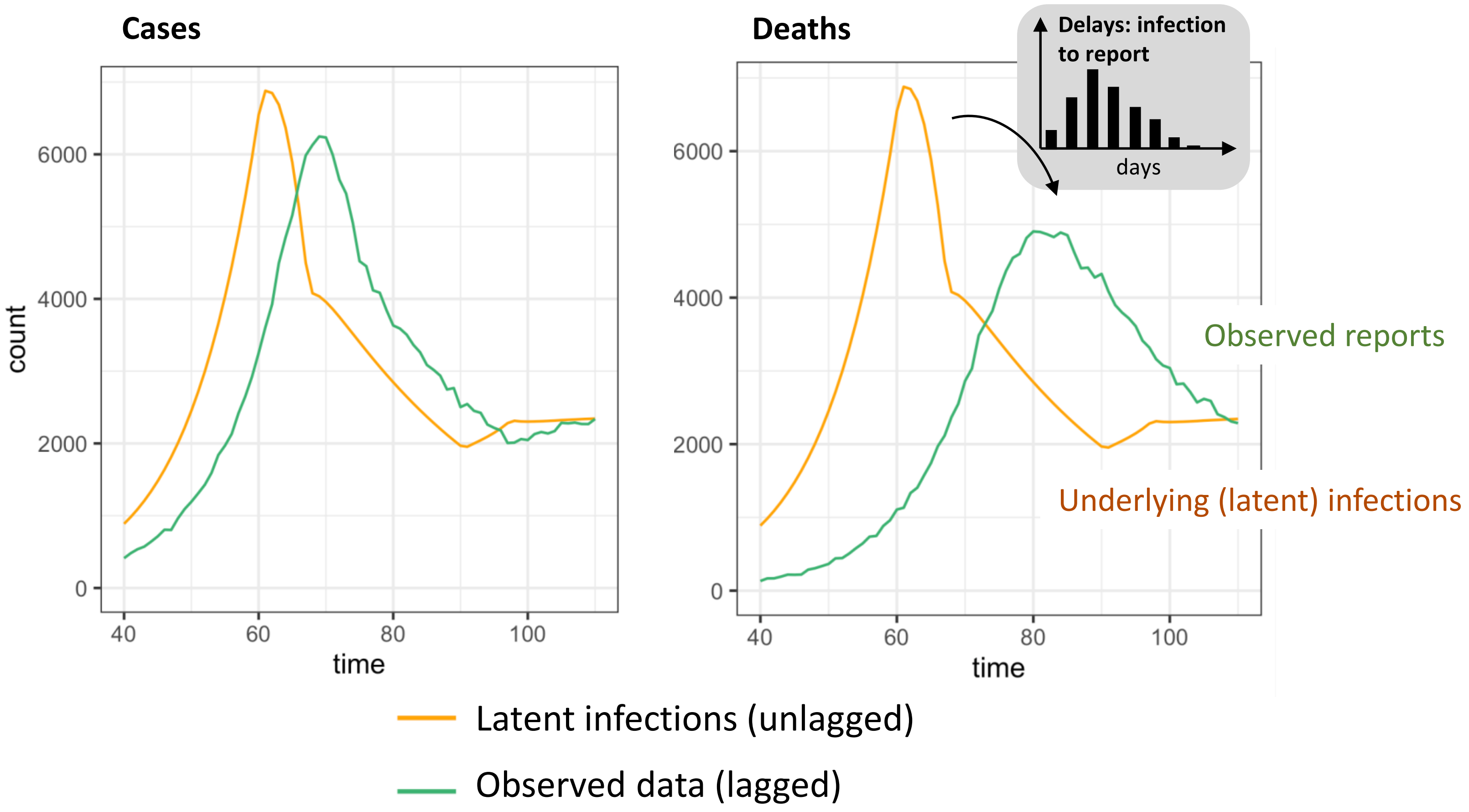
If we don’t account for the delay between infection and observation, the R_t estimates that we obtain today actually reflect transmission levels days or weeks ago

Temporal accuracy



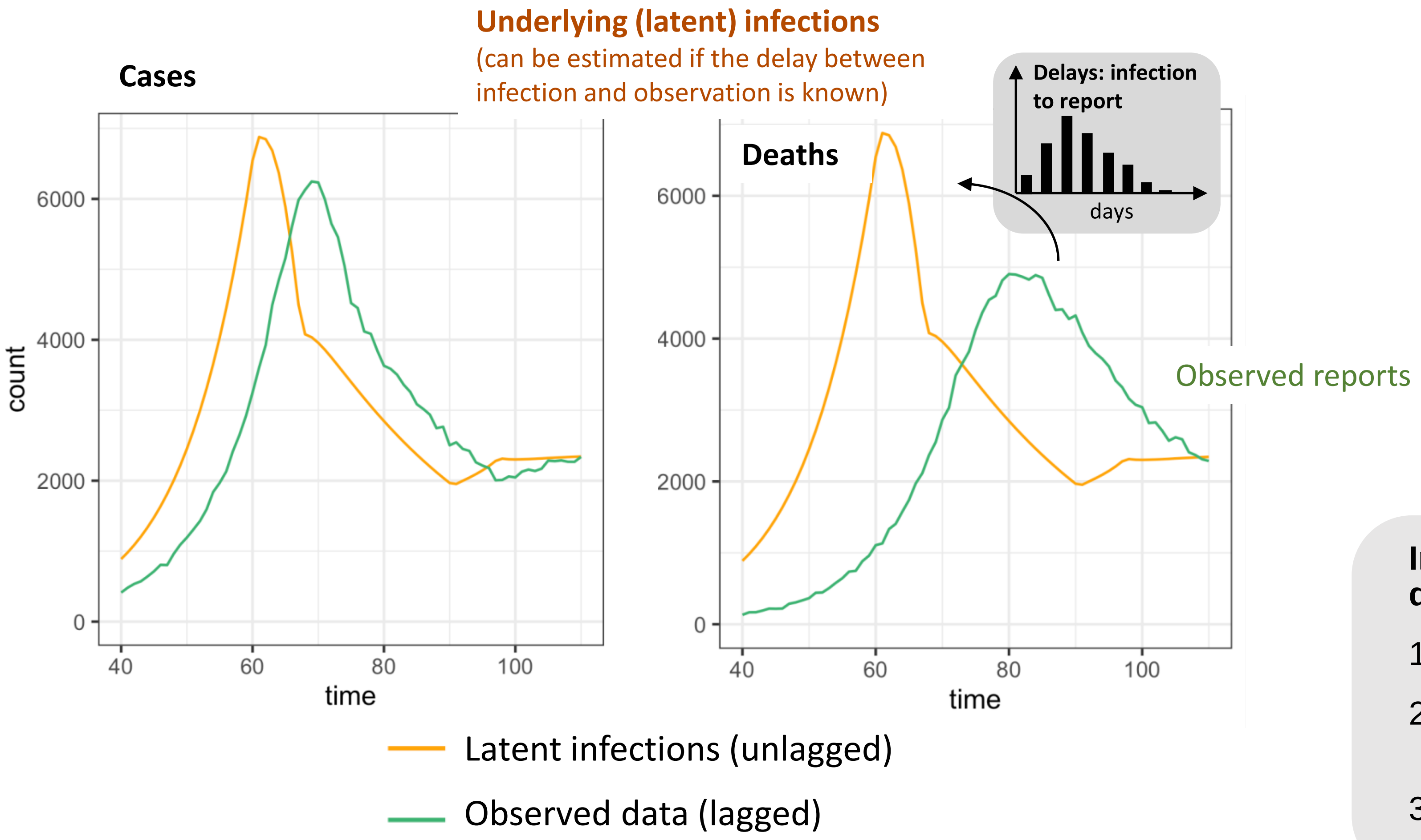
JAMA. 2020;323(19):1915-1923. doi:10.1001/jama.2020.6130

Problem: delays change the shape and timing of the observed curve, relative to underlying infections



Solution: Use a tool that estimates the unlagged infection curve along with R_t

e.g. EpiNow2, epinowcast, epidemia packages in R

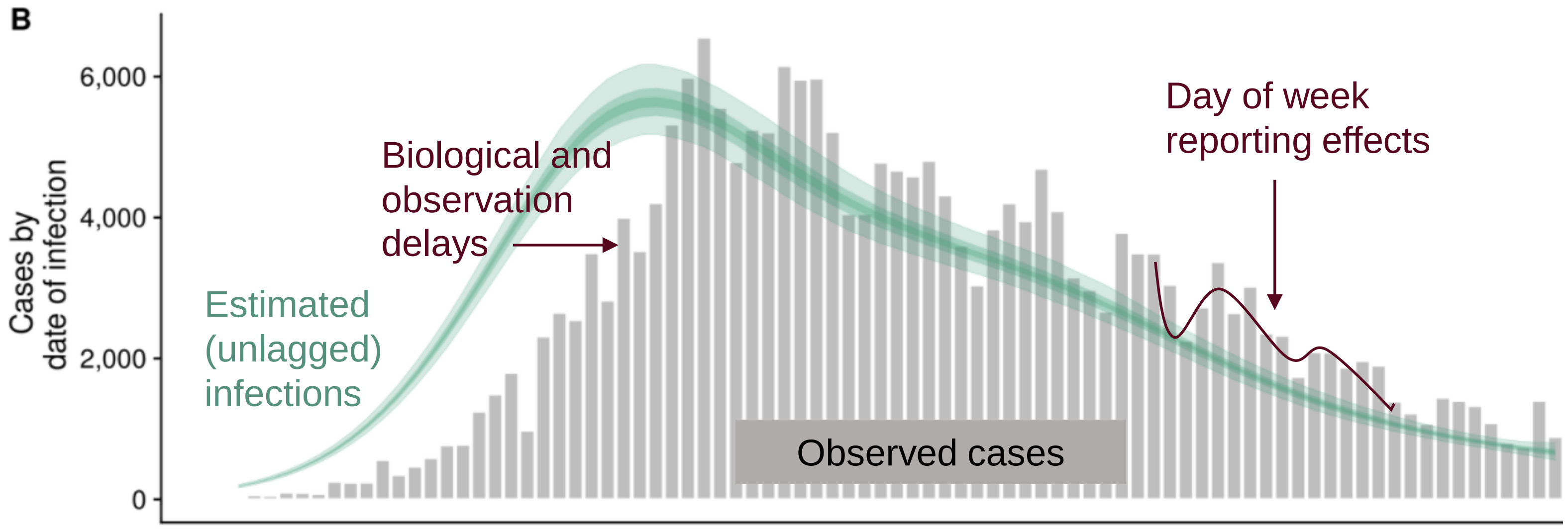


Inputs needed to estimate R_t, accounting for delays

1. A time series of observed reports
2. An estimate of the generation interval distribution
3. An estimate of infection-to-report delay distribution

Problem: Day of week effects

Solution: Use a model in which the reporting hazard depends on the day of the week. Estimate a fixed effect for each day of the week.



Abbott, S. et al. (2020). EpiNow2. doi:10.5281/zenodo.3957489.

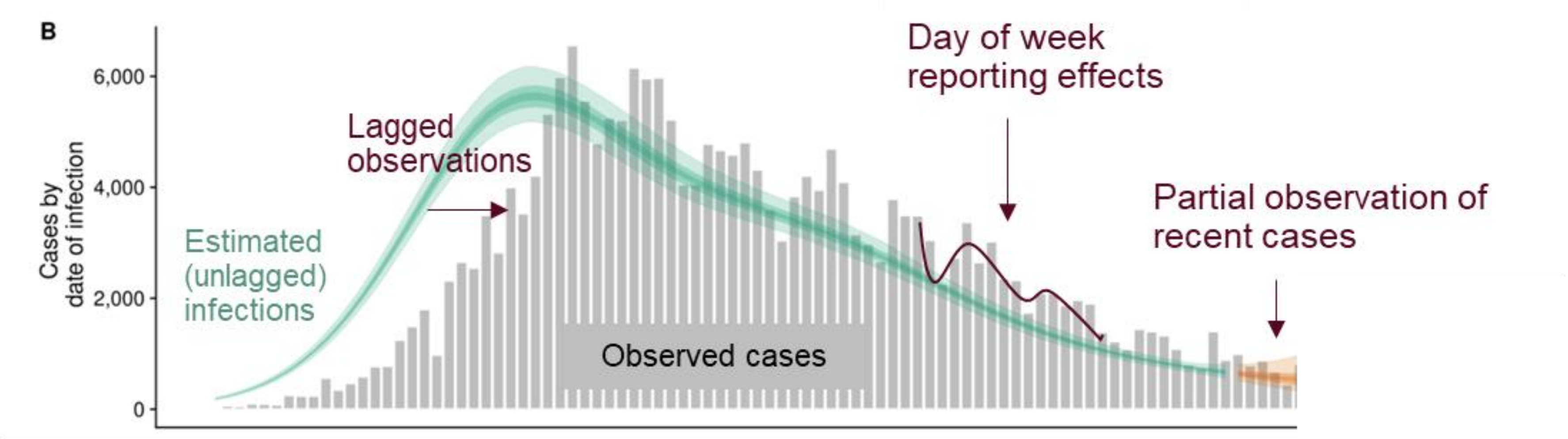
EpiNow2 and epinowcast are two packages that provide this option.

(Stay tuned for the coding demo!)

Problem: Partial observation of recent counts

Lags, backfill, revision, are inevitably present in epidemiological data

Solution: Use a method that can **nowcast** the end of the time series and estimate R_t simultaneously



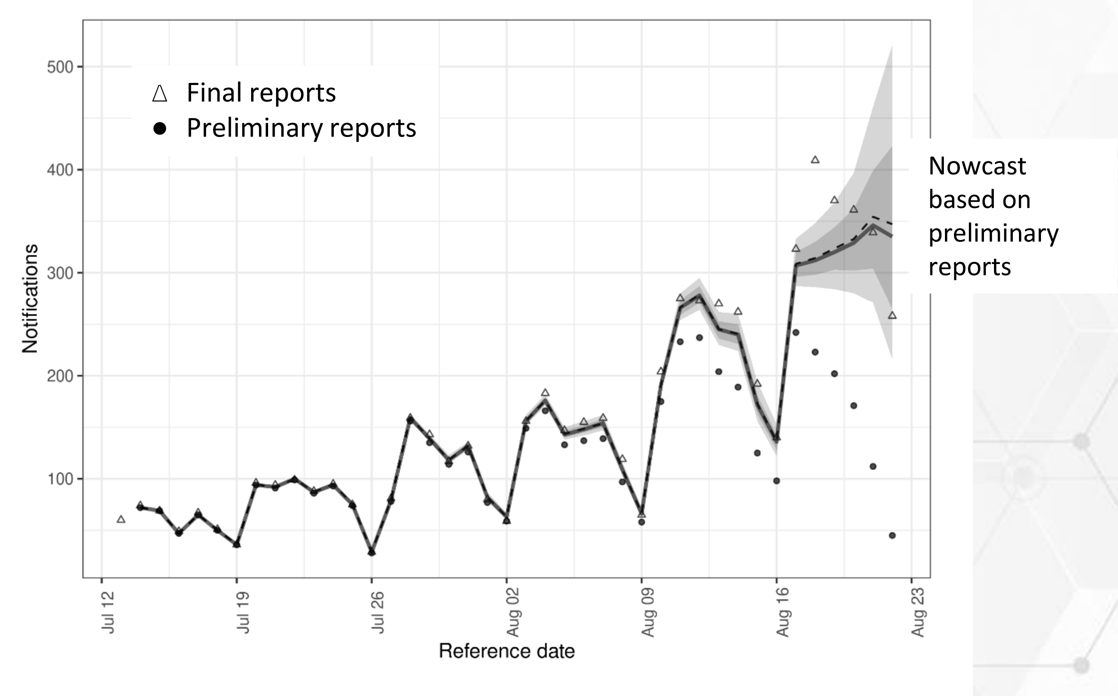
Abbott, S. et al. (2020). EpiNow2. doi:10.5281/zenodo.3957489.

EpiNow2 and epinowcast are two packages that provide this option

(Stay tuned for the coding demo!)

Nowcasting

Estimate final reports from those observed so far



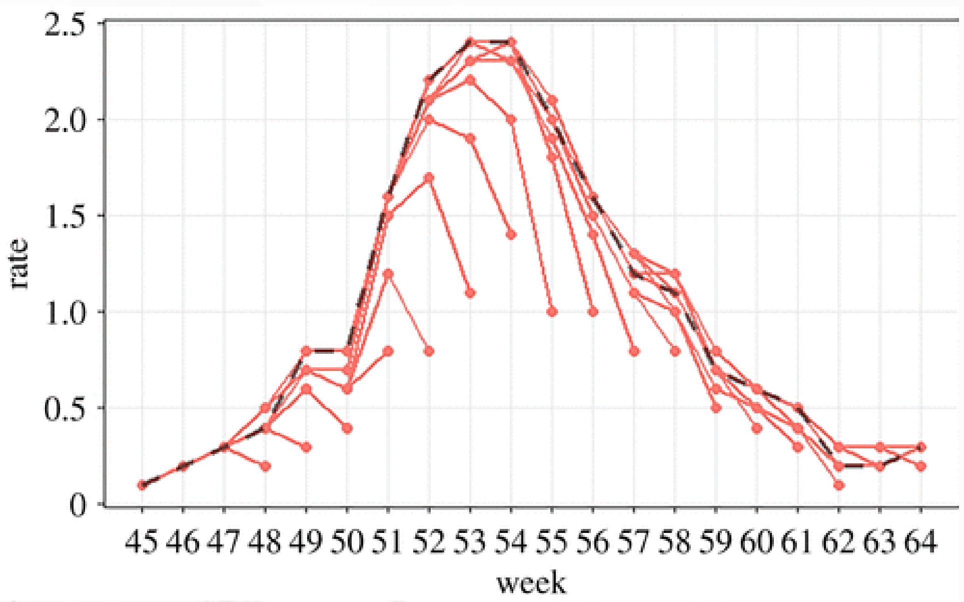
https://package.epinowcast.org/

- Basic nowcasting approach
  - Estimate the fraction of total reports we expect to have observed 1, 2, 3, … n days after the fact
  - Scale up recent incomplete reports accordingly
- Several nowcasting packages available
  - NobBS ("Nowcasting by Bayesian Smoothing")
  - epinowcast
  - nowcaster
  - surveillance
  - EpiNow2
- Sophisticated tools can stratify by geographic, demographic factors, or allow reporting delays to evolve over time
  - epinowcast
  - nowcaster

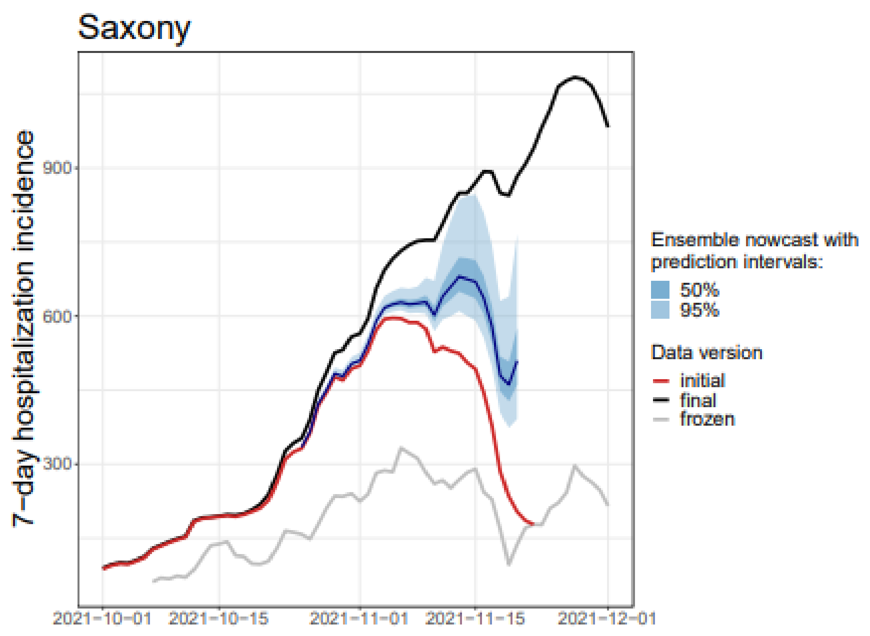
Inputs needed for nowcasting

Estimate of the delays between initial and final reports

- Need to characterize waiting times between event and report
- Data snapshots or line lists with event date and report date are sufficient
- Some public databases already store daily archives of key epidemic signals for this purpose
    - Covidcast
    - HealthData.gov
    - Stay tuned for the coding demo!
- Some existing R packages make it easy to
    - estimate delays (EpiNow2, epinowcast)
    - perform nowcasting (EpiNow2, epinowcast, NobBS)



Kandula et al. *Journ. Roy. Soc. Interface.* 2019.



(a) Saxony, nowcast made on 22 November 2022: overwhelmed hospitals lead to severe underreporting and thus too low nowcasts.

Wolfram et al. *medrxiv.* 2023.

Limitations

- Nowcast accuracy is greatest when the data observation process is stable and predictable
- Reporting anomalies and inaccurate nowcasts can occur at critical times for public health decision making
  - Early in an outbreak
  - When healthcare systems are over capacity
  - During periods of social upheaval
- Nowcasting is an active area of research
- CDC is using historical outbreak data to test and improve methods
- CDC's open funding opportunity will build on this research and support nowcast implementation at the state and local levels

https://www.cdc.gov/forecast-outbreak-analytics/nofo.html

4. Putting it all together

Two tools we recommend for R_t estimation

Designed to adjust for observation lags, day of week effects, and partial observation while estimating R_t

EpiNow2

- Developed during the COVID-19 pandemic
- Used for CDC's mpox technical reports

\+ Relatively easy to use and install

\- Limited flexibility for nowcasting
  - Two-step process
  - When nowcasting, reporting delays can't vary over time or across demographic strata

\- Model fitting is slower, less stable than it could be

Focus of coding demo

epinowcast

- Developed by the same team and intended to replace EpiNow2

\+ Faster, more stable, improved statistical methods

\+ Nowcasting is a one-step plug-and-play process

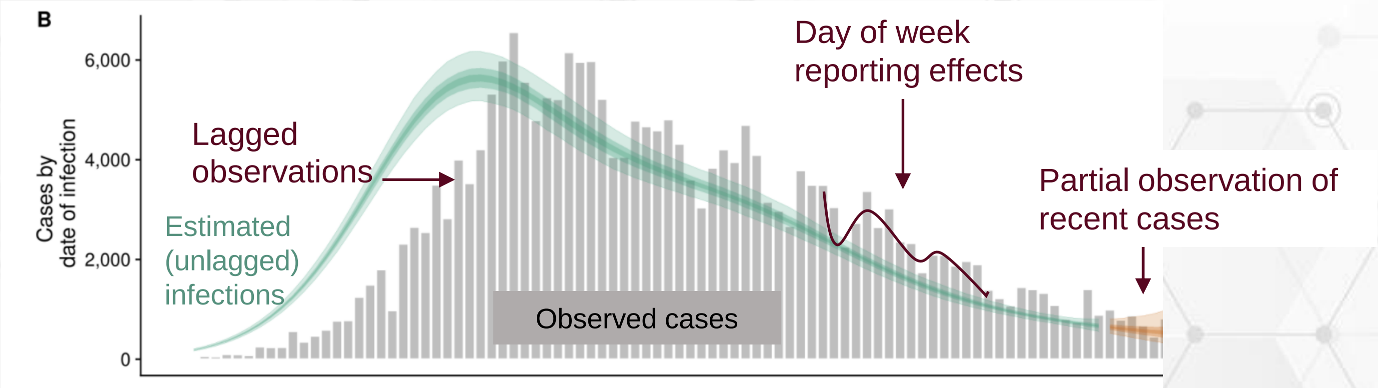
\+ More flexibility: model structure, covariates, changes over time in reporting

\- Still in development

\- Harder to install

\- For now, requires expert knowledge to take advantage of flexibility, generate plots, etc.

Real-time epidemic analysis requires models and tools designed for imperfect epidemic data



Abbott, S. et al. (2020). EpiNow2. doi:10.5281/zenodo.3957489.

Imperfections in outbreak data are predictable and need to be dealt with before the analysis begins

Pre-positioned, well-tested computational tools and data pipelines needed for accurate, timely decision support

Tools are rapidly improving, but there is still a long way to go

CDC is supporting these efforts

- Funding for collaborations, and growth of the talent pipeline
- Open-source tool development
- Testing tools on historical outbreak data → continuous improvement
- Full-time team to produce and publish analyses at the state and territory level

CDC24/7

PROTECTING AMERICA'S SAFETY, HEALTH, AND SECURITY

Thank you!

Questions?

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov