A guide to accounting for reporting delays in state, local, and territorial public health surveillance data

Laura Jones[1]   Sam Abbott[2,*]

2026-02-02

State, local, and territorial surveillance systems are essential for public health decision making, but inherent delays between disease occurrence and reporting create challenges for real-time analysis. Other issues such as data revisions, site drop-in and drop-out, and data quality problems can also manifest as apparent delays in aggregate data. This guide provides practical guidance for epidemiologists, public health practitioners, and modellers working with reporting delays and related challenges in surveillance data. We describe challenges in real-time use of surveillance data, opportunities from modelling approaches, guidance on choosing appropriate methods, considerations for communicating results, and practical case studies with implementation resources. We also highlight gaps where new modelling methods could address unmet needs in public health practice.

[1] Massachusetts Department of Public Health, Boston, MA, USA
[2] Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, United Kingdom

[*] Correspondence: Sam Abbott <sam.abbott@lshtm.ac.uk>

Introduction

Section lead: Sam Abbott **Section support:** TBD

State, local, and territorial surveillance systems play a central role in public health decision making in the United States. Similarly, sub-national jurisdictions in other countries also play a signficant role in public health decision making. However, several challenges complicate the use of surveillance data in real time. Reporting delays arise from lab confirmation requirements, differences between electronic and manual reporting, and weekend and holiday effects. Data

revisions occur as duplicates are removed, cases are reclassified, and dates are corrected. Site-specific variations in reporting capabilities and intermittent reporting from some facilities create additional uncertainty. Other data quality issues such as missing or incorrect dates and incompatible formats can further limit the utility of recent data. Many of these issues can manifest as apparent delays when viewed in aggregate data, even when the underlying cause is not a true reporting delay.

Common approaches to handling these challenges include pruning recent data or using incomplete data with caveats. However, statistical methods exist that can adjust for these issues by learning from historical patterns. Addressing these challenges can improve situational awareness of true disease trends, supports decision making (such as timing of public health interventions), and can improve forecasting of disease trends and resource needs (i.e. hospitalisations). Examples of publicly available use of modelling to account for challenges data reporting include the Massachusetts Department of Public Health's respiratory illness dashboard, New York City's nowcasting during mpox and COVID-19 emergencies, California's nowcast of COVID-19 effective reproduction number, and the CDC's COVID-19 variant nowcast.

In this guide, we provide practical guidance for public health practitioners and modellers planning to account for reporting delays and other data quality issues in their analyses. We describe the challenges that arise when using surveillance data in real time, review modelling approaches that can address these challenges, and provide guidance on choosing and implementing appropriate methods. We also highlight gaps where new methods could address unmet needs in practice. This guide is accompanied by an interactive website with a decision tree for method selection and a community code repository with implementation examples.

Challenges in real-time use of surveillance data

Section overview lead: TBD **Section overview support:** TBD

Table 1 summarises the challenges described in this section.

Table 1: Summary of challenges in real-time surveillance data

Challenge	Description	Examples
Reporting delays	Time between event and report	Lab confirmation, manual entry
Data revisions	Changes to previously reported data	Duplicate removal, reclassification
Site drop-in/out	Facilities joining or leaving	New sites, closures, intermittent reporting
Other data quality	Additional data issues	Missing dates, format incompatibilities

Reporting delays

Section lead: TBD **Section support:** TBD

What it is

- Right-truncation

Why it happens

- Lab confirmation requirements
- Electronic vs manual reporting differences
- Weekend and holiday effects

How it affects analysis

How to identify it

Data revisions beyond reporting delays

Section lead: TBD **Section support:** TBD

What it is

Why it happens

- Downward corrections from duplicate removal
- Case reclassifications
- Date corrections
- De-duplication

How it affects analysis

How to identify it

Surveillance site drop-in and drop-out

Section lead: TBD **Section support:** TBD

What it is

Why it happens

- System-to-system delays
- Hospital/facility intermittent reporting
- Urban vs rural reporting capabilities
- EHR integration disparities

How it affects analysis

How to identify it

Other data quality issues

Section lead: TBD **Section support:** TBD

What it is

Why it happens

- Duplicate entries across systems
- Missing or incorrect dates
- Missing strata variables of interest (e.g., race/ethnicity)
- Incompatible formats

How it affects analysis

How to identify it

Opportunities from modelling

Section overview lead: TBD **Section overview support:** TBD

Table 2 provides an overview of modelling approaches that can address the challenges described in the previous section.

Table 2: Modelling approaches for addressing surveillance data challenges

Challenge	Approach category	Examples	Data requirements
Reporting delays	Chain ladder	baselinenowcast, ChainLadder	(**placeholder?**)
	GAMs	nowcaster, UKHSA GAMs	(**placeholder?**)
	Bayesian hierarchical	NobBS	(**placeholder?**)
	Ad-hoc	EpiNow2	(**placeholder?**)
	Semi-mechanistic	EpiNow2, epinowcast	(**placeholder?**)
Data revisions	Chain ladder	baselinenowcast	(**placeholder?**)
Site drop-in/out	(**placeholder?**)	(**placeholder?**)	(**placeholder?**)
Other data quality	(**placeholder?**)	(**placeholder?**)	(**placeholder?**)

Correcting delayed data

Section lead: TBD **Section support:** TBD

Data requirements

- Defining report and reference dates
- Event date hierarchy (NYC mpox example)
- Ability to assess data revision history
- Volume of data available

Methods

- Chain ladder approaches (baselinenowcast, ChainLadder)
- GAMs (nowcaster, UKHSA GAMs)
- Bayesian hierarchical methods (NobBS)
- Ad-hoc methods (EpiNow2)
- Semi-mechanistic approaches (EpiNow2, epinowcast)

Implementation

- Technical adjustments to capture data revision history
- Aggregating data by report/reference date
- Considerations for stratified nowcasts

Managing data revisions beyond reporting delays

Section lead: TBD **Section support:** TBD

Data requirements

Methods

Implementation

Handling site drop-in and drop-out

Section lead: TBD **Section support:** TBD

Data requirements

Methods

Implementation

Managing other data quality issues

Section lead: TBD **Section support:** TBD

Data requirements

Methods

Implementation

Choosing a modelling method

Section overview lead: TBD **Section overview support:** TBD

?@fig-decision-tree provides a decision tree linking methods to data characteristics and use case considerations.

- Create decision tree figure linking methods to considerations
- Link methods to considerations with examples of common PH data sources
- Considerations for when NOT to nowcast

The following sections provide detailed guidance for specific aspects of method selection, referring back to the decision tree where relevant.

Adoption and sustainability

Section lead: TBD **Section support:** TBD

- Advocating to public health leadership
- Explaining nowcasts to decision makers
- Justifying system modifications

Data analysis

Section lead: TBD **Section support:** TBD

- Analysing delay distributions
- Choosing maximum delay and training window
- Representativeness
- What historic data is most informative?
- Determining variables needed, strata of interest

Implementation considerations

Section lead: TBD **Section support:** TBD

- Considerations for emergencies vs chronic delays
- Software/infrastructure requirements
- Failure modes
- Maintaining systems over time
- Tradeoff between flexibility and model complexity
- Model specific (link back to modelling opportunities implementation section)

Validating and evaluating a model

Section lead: TBD **Section support:** TBD

- Practical qualitative methods (flipbooking, alignment with trends)
- Alignment with overall trend vs inflection points
- Applied PH quantitative methods (coverage, correlation, residuals)
- Advanced methods (WIS, MAE/MSE) for method comparison
- Tension between domain expertise and theoretical scores
- Evaluation for public health utility
- Against a baseline and other common methods

Communicating and visualising results

Section lead: TBD **Section support:** TBD

Presenting uncertainty to decision makers

Section lead: TBD **Section support:** TBD

- Communicating uncertainty
- What prediction/confidence intervals to show or use for decisions

Public-facing output considerations

Section lead: TBD **Section support:** TBD

- Data presentation
- Placing nowcasts in context (e.g., seasonal intensity thresholds)

Common pitfalls

Section lead: TBD **Section support:** TBD

How-to case studies

Section overview lead: TBD **Section overview support:** TBD

Regardless of the specific challenge or method, implementing a model to account for reporting challenges in surveillance data involves common steps:

- Assessing the data source and predictability of the reporting challenge
- Analysing any delay and delay distributions
- Understanding the meaningful strata (i.e. ethnicity)
- Implementation
- Validation
- Communication and visualisation

The following case studies illustrate how these steps apply in practice.

Case study: Syndromic surveillance

Section lead: Laura Jones **Section support:** TBD

Case study: To be determined

Section lead: TBD **Section support:** TBD

Additional tools and resources

Section lead: TBD **Section support:** TBD

- Interactive website with decision tree for method selection
- Community code repository with implementation snippets
- Links to software packages and documentation
- Example datasets for testing and learning

Discussion

Section lead: Sam Abbott **Section support:** TBD

Note: Subtitles in this section are for organisation during writing and will be removed at submission.

Summary

We presented a practical guide for public health practitioners and modellers working with surveillance data affected by reporting delays and other data quality issues. We described the challenges that arise when using surveillance data in real time, reviewed modelling approaches that can address these challenges, and provided guidance on choosing and implementing appropriate methods. We also provided case studies demonstrating how these methods can be applied in practice.

Strengths and limitations

This guide brings together practical experience from public health practitioners and modellers working with delayed surveillance data. We focus on methods that have been applied in real-world settings rather than purely theoretical approaches. However, our coverage of methods is not exhaustive and the field continues to evolve.

Related work

Several reviews have examined nowcasting methods in specific contexts ((**placeholder?**)). Our guide complements these by focusing on practical implementation considerations for state, local, and territorial public health settings.

Future directions

Key areas for methodological development include methods for handling site drop-in and drop-out, approaches for managing data revisions beyond reporting delays, and tools that integrate multiple data quality adjustments. Improved software implementations that lower barriers to adoption would also benefit practitioners.

Conclusions

Reporting delays and data quality issues are inherent to surveillance systems but need not prevent timely public health decision making. Statistical methods exist that can adjust for these issues, improving situational awareness and supporting evidence-based responses. We hope this guide helps practitioners identify when such methods may be useful and provides a starting point for implementation. The companion website and code repository provide additional resources for those implementing these methods. We also hope this guide highlights gaps where new methods could address unmet needs, encouraging further methodological development in this area.

References