# Evaluation of Rolling Surveillance Methods in Context of Prior Aberrations

## A Simulation Study with Routine Data from Low- and Middle-Income Countries

Anuraag Gopaluni, Nicholas B. Link, Emma Boley, Isabel Fulcher, Muhammed Semakula, Bethany Hedt-Gauthier

November 6, 2024

# Background

- Infectious disease surveillance enables early detection of outbreaks, aiding timely public health responses [Buehler et al., 2004, Wagner et al., 2001].
- Syndromic surveillance leverages health symptoms as outbreak proxies, critical in settings with limited testing [Pearce et al., 2020, Fulcher et al., 2021].
- Focus on low- and middle-income countries (LMICs), where health management information systems (HMIS) offer routine, aggregated data [Lal et al., 2020, AbouZahr and Boerma, 2005].

# Syndromic Surveillance in LMICs

- HMIS data provides a basis for syndromic surveillance in LMICs, supporting infectious disease tracking in limited-resource settings [World Health Organization, 2021, Hussain et al., 2016].
- Example: Acute respiratory infections (ARI) monitoring for COVID-19 in Liberia [Fulcher et al., 2021].

# Rolling Surveillance vs Event Surveillance

- Rolling surveillance accumulates baseline data over time, allowing continuous comparison of observed data to historical expectations [Noufaily et al., 2013].
- Event surveillance, by contrast, assesses spikes relative to a fixed baseline period.

# Defining **Aberrations** and **Outbreaks**

- **Aberrations**: Retrospective anomalies, such as historical disease events, that distort baseline data.
- **Outbreaks**: Significant new deviations from the expected baseline, identified prospectively.
- Challenge: Unaccounted aberrations may bias baseline models, compromising outbreak detection [Farrington et al., 1996].

# Outbreak Detection Algorithms Overview

- Evaluated five outbreak detection algorithms:
    1. Early Aberration Reporting System (EARS)
    2. Farrington Detection Model
    3. Holt-Winters Predictive Models
    4. Weinberger-Fulcher Negative Binomial (WF NB)
    5. Weinberger-Fulcher Quasi-Poisson (WF QP)
- Each algorithm offers unique sensitivity and specificity characteristics [Farrington et al., 1996, Weinberger et al., 2020].

# Method 1: Early Aberration Reporting System (EARS)

- Utilizes a seven-month baseline period preceding the evaluation point.
- Calculates expected count and prediction intervals for detection:

$$E[Y_t] = \frac{1}{7} \sum_{i=t-7}^{t-1} Y_i,$$

$$\mathrm{Var}(Y_t) = \frac{1}{6} \sum_{i=t-7}^{t-1} (Y_i - E[Y_t])^2.$$

- Observations outside the 95% prediction interval are flagged as potential outbreaks [Hutwagner et al., 2005].

# Method 2: Farrington Detection Model

- A log-linear Poisson regression model with quasi-likelihood adjustment:

$$\log(E[Y_t]) = \alpha + \beta t.$$

- Uses residual-based weights to account for aberrations in baseline data.
- Flags deviations from the upper prediction interval as outbreaks [Farrington et al., 1996].

# Method 3: Holt-Winters Predictive Models

- Applies exponential smoothing, accounting for level, trend, and seasonality:

$$\hat{Y}_{t+h|t} = l_t + hb_t + s_{t+h-m}.$$

- Components include level ($l_t$), trend ($b_t$), and seasonal adjustments ($s_t$) [Chatfield, 1978].
- Flags values outside the 95% interval for outbreak detection.

# Methods 4 and 5: Weinberger-Fulcher Models

- Adapted for COVID-19 surveillance using a baseline model with a negative binomial (NB) or quasi-Poisson (QP) approach:

$$\log(E[Y_t]) = \beta_0 + \beta_1 \text{year}_t + \sum_{k=1}^{K} [\beta_{3k} \cos(2\pi kt/12) + \beta_{4k} \sin(2\pi kt/1$$

- Long-term trends and seasonal harmonics capture variation; flags facilities outside prediction intervals [Weinberger et al., 2020].

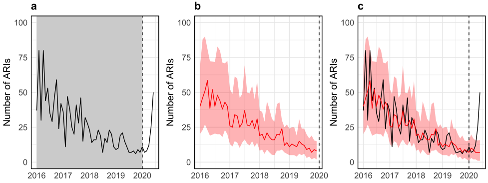



Figure: Steps in the Weinberger-Fulcher approach for outbreak detection.

# Simulation Study Framework

- General steps in the simulation framework (detailed in Figure 4):
    1. Generate baseline data and outbreaks under multiple scenarios to reflect diverse data-generating models (DGMs).
    2. Evaluate each rolling surveillance approach's sensitivity and specificity in a new month without baseline aberrations.
    3. Induce aberrations in the baseline data, refit models, and re-evaluate sensitivity and specificity.

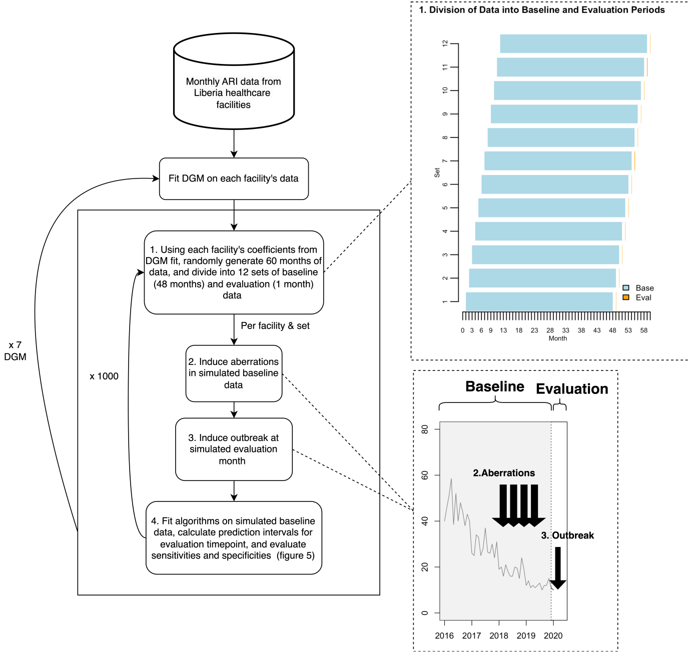



Figure: Simulation study framework steps.

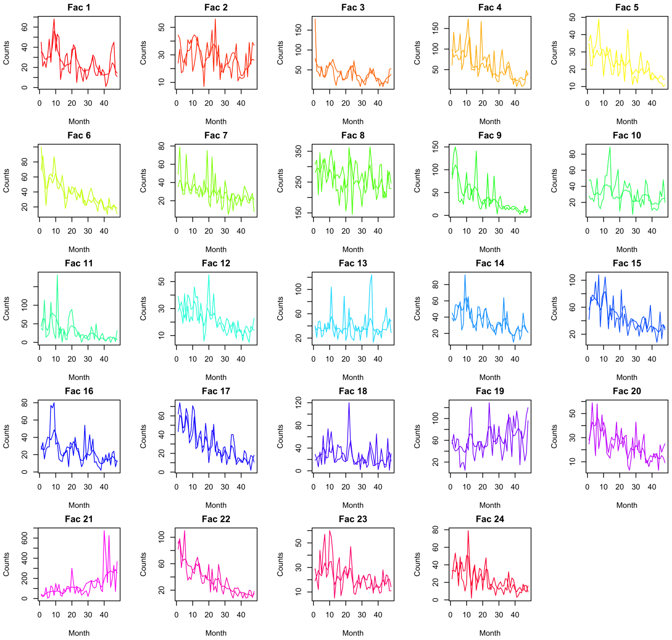
# Generating Baseline Data

- Baseline data generated using a negative binomial model with trend and seasonality components, inspired by prior studies [Zareie et al., 2023, Unkel et al., 2012, Noufaily et al., 2013].
- Model form:

$$\log(E[Y_t]) = \beta_0 + \beta_1 \text{year}_t + \sum_{k=1}^{K} \left[ \beta_{3k} \cos\left(\frac{2\pi kt}{12}\right) + \beta_{4k} \sin\left(\frac{2\pi kt}{12}\right) \right],$$

  where $K$ represents seasonality (triannual, biannual, annual), and $\beta_1$ captures trend.
- Seven distinct Data Generating Models (DGMs) considered, combining presence/absence of trend and different seasonality levels. DGM 1 corresponds to the WF model.
- Baseline data generated for 24 facilities, each with unique parameter sets, replicated 1,000 times per DGM for robust simulation.

# DGM 1 Example Data



# Inducing Outbreaks

- Outbreaks introduced in the 49th month with varying sizes:

$$E(Y_{49}) + k_O \cdot \mathrm{SD}(E(Y_{49})),$$

  where $k_O$ controls outbreak magnitude.
- Given the negative binomial distribution, the outbreak form becomes:

$$E(Y_{49}) + k_O \cdot \sqrt{E(Y_{49}) + \frac{E(Y_{49})^2}{\phi}}.$$

- Explored factors: outbreak size ($k_O$: 1-20) and timing (months 49 to 60).

# Inducing Aberrations

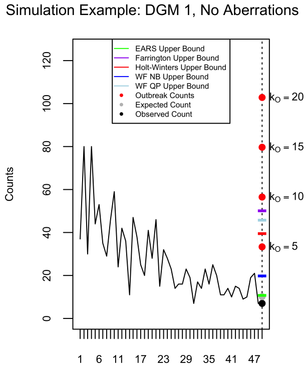
- Retrospective aberrations induced in the baseline data.
- Factors:
  1. Aberration size ($k_A$), modeled as:

     $$E(Y_t) + k_A \cdot \sqrt{E(Y_t) + \frac{E(Y_t)^2}{\phi}},$$

     with $k_A$ values of 5, 10, 15, and 20.
  2. Length ($l$): single month ($l = 1$) or four consecutive months ($l = 4$).
  3. Timing: unrestricted or restricted to past year.

# Evaluation Metrics

- **Sensitivity**: Probability an outbreak in the 49th count is flagged. An algorithm is sensitive if the induced outbreak is above the upper bound of its 95% prediction interval.
- **Specificity**: Probability a non-outbreak count in the 49th month is not flagged, independent of outbreak size.
- Figure 5 illustrates an example simulation run and evaluation metrics.



# Pseudo-Receiver Operating Characteristics Curves (ROCs)

- To compare algorithms, we varied prediction intervals by adjusting $\alpha$ levels (0.000000001 to 0.75), achieving specificities for each method.
- For each $\alpha$, calculated method sensitivity, constructing pseudo-ROC curves to compare sensitivity and specificity jointly.

# Results Overview

- **Data Generating Models (DGMs)**: Results shown for DGMs 1 and 2; additional DGMs in appendix. DGM 1 includes time trend and seasonality, DGM 2 includes seasonality only.
- **Outbreak Sizes**: Results for $k_O = 5, 10, 20$ presented.
- **Aberrations**: Focused on length $l = 4$ and size $k_A = 10$; other configurations are in the appendix.
- **Outbreak Month**: Examines effect of outbreak timing, then results are averaged across timepoints.
- **Facility**: Results shown as boxplots across facilities and per-facility pseudo-ROC curves.

# Results for DGM 1: Time Trend and Triannual Seasonality

- **Effect of Outbreak Timing on Sensitivity and Specificity**:
  - Next figure, for outbreak size $k_O = 10$, aberration size $k_A = 10$, length $l = 4$, shows median specificity and sensitivity.
  - Consistent relative performance across timepoints; thus, results were aggregated across all timepoints.
- **Effect of Facility-Level Parameters and Outbreak Size**:
  - Following Figure: Sensitivity and specificity across no aberrations, aberrations anytime, and last year, as boxplots.
  - Larger outbreaks generally increase sensitivity; WF NB, EARS most sensitive for smaller outbreaks.

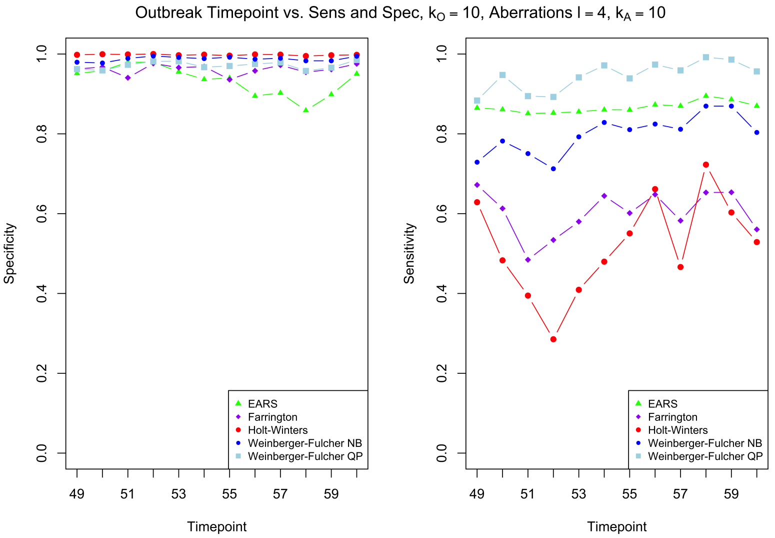



Figure: Effect of outbreak timing on sensitivity and specificity (DGM 1).

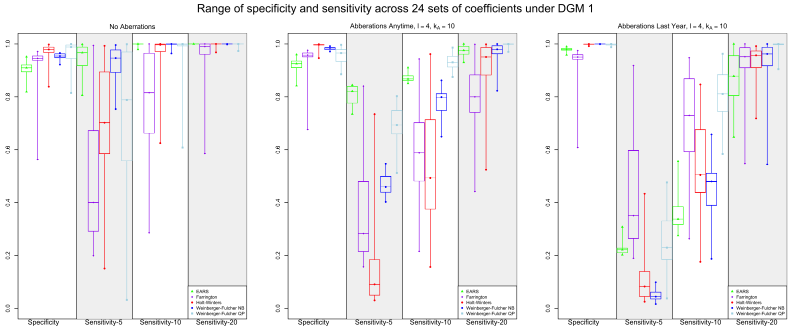



Figure: Sensitivity and specificity by outbreak size and aberration timing (DGM 1).

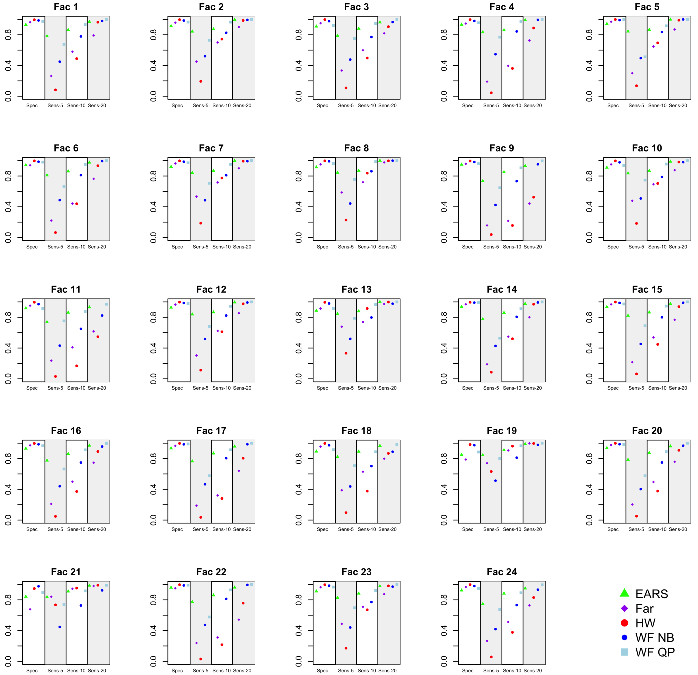
# Effect of Outbreak Sizes and Aberration Timing

- **Outbreak Sizes**:
  - Sensitivities increase with outbreak size.
  - For $k_O = 5$, EARS and WF NB most sensitive (median sensitivity = 0.97, 0.95); larger outbreaks show high sensitivity across methods.
- **Aberration Timing**:
  - WF QP and EARS most sensitive for $k_O = 10$ when aberrations can occur anytime; HW, WF NB most specific.
  - EARS sensitivity reduced when aberrations restricted to last year, WF QP and Farrington remain sensitive.

# Effect of Facility-Level Parameter Sizes

- Variability in specificity and sensitivity across facility parameters:
  - No aberrations: EARS and WF NB have low variability (e.g., IQR = (0.92, 0.99)).
  - With aberrations: EARS, WF NB, and WF QP maintain lower variability.
- Next figure: Facility-specific sensitivity and specificity results for aberrations anytime and restricted to last year.

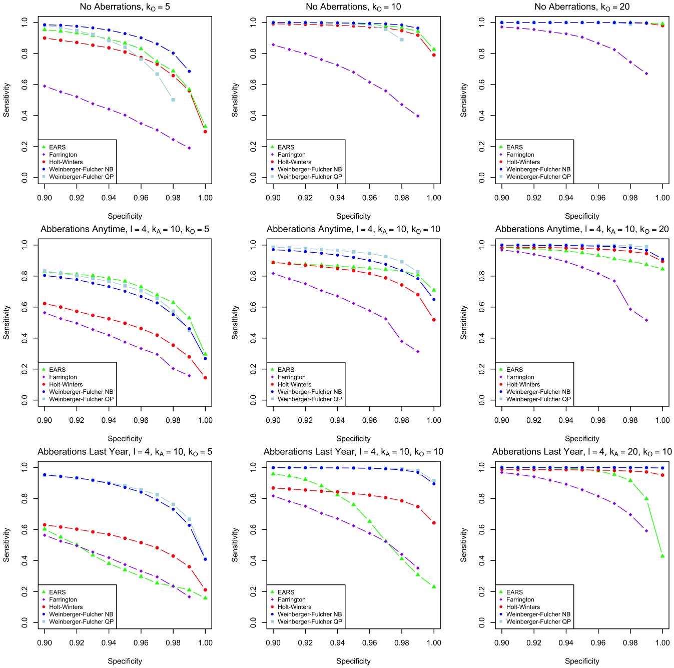
Sensitivity and Specificity by facility coefficients under DGM 1, Abberations Anytime, $l = 4$, $k_A = 10$



# Pseudo-Receiver Operating Characteristic (ROC) Curves for DGM 1

- **Pseudo-ROC Curves**:
  - For $k_O = 5$, with specificity at 0.95, WF NB and EARS achieve highest sensitivity without aberrations.
  - When aberrations are present anytime, EARS, WF QP, and WF NB balance sensitivity and specificity well.
  - Aberrations restricted to last year reduce EARS sensitivity.

# Pseudo-ROC Curves DGM 1



# Results for DGM 2: Triannual Seasonality without Time Trend

- **Outbreak Sizes**:
  - Sensitivities increase with outbreak size.
  - For $k_O = 5$, EARS and WF QP most sensitive (median sensitivity = 0.98); all methods highly sensitive for larger outbreaks.
- **Aberration Timing**:
  - WF QP and EARS remain sensitive for $k_O = 10$ when aberrations occur anytime.
  - Farrington sensitivity highest when aberrations are restricted to last year, though specificity is reduced.

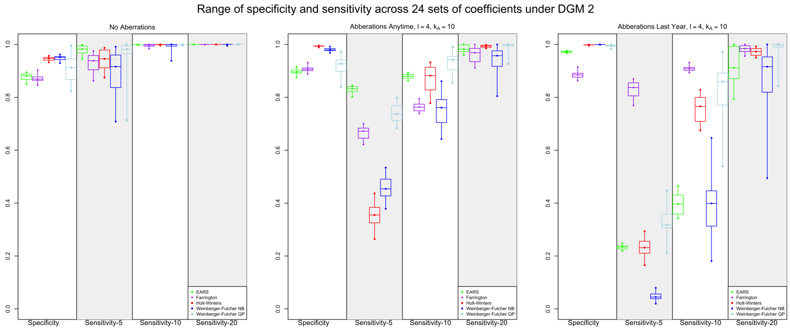
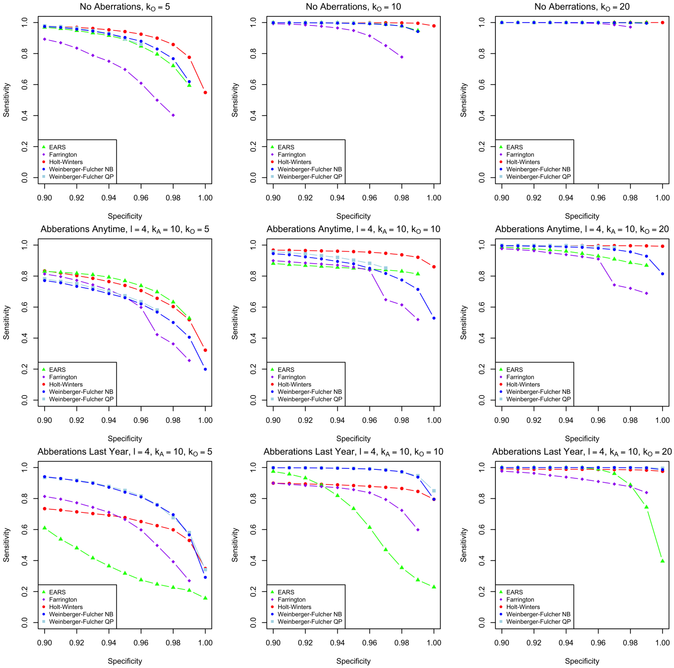



Figure: Sensitivity and specificity results for DGM 2 by outbreak size and aberration timing.

# Pseudo-ROC Curves for DGM 2

- **Pseudo-ROC Analysis**:
  - Without aberrations, HW and EARS offer best balance for smaller outbreaks.
  - With aberrations restricted to last year, WF QP and WF NB provide the best balance.

# Pseudo-ROC Curves DGM 2



# Discussion

- **Study Focus**: Examined rolling surveillance accuracy with baseline aberrations, using HMIS-derived simulation parameters.
- **Key Observations**: All algorithms' sensitivity decreased with longer or larger aberrations; specificity slightly increased due to wider prediction intervals from baseline variability.

# Discussion (Continued)

- **Algorithm Performance Summary**:
  - **EARS**: Strong performance, minimal baseline requirements, but sensitive to recent data disruptions.
  - **Farrington**: Useful for small outbreak detection, but lower specificity, especially in weak seasonality.
  - **WF QP**: Balanced sensitivity and specificity in recent aberrations; suitable for operational use.
  - **Holt-Winters**: Promising for large outbreaks with minimal time trend but sensitive to recent aberrations.

# Discussion (Limitations)

- Focused on five algorithms; future work could explore additional approaches like CUSUM and spatial Bayesian models.
- Specific aberration types evaluated; future work could consider varied aberration patterns (e.g., non-consecutive, spatial).
- Considered sensitivity and specificity; future assessments could include timeliness, predictive value, and practical trade-offs.

# Discussion (Broader Context)

- Real-world applications: All algorithms implemented in various settings [Weinberger et al., 2020, Zareie et al., 2023].
- Emerging challenge: Detecting new outbreaks amidst historical aberrations, especially relevant post-COVID-19.
- Suggested future focus: Evaluations across timepoints, balancing sensitivity and specificity rigorously.

# Conclusion

- Rolling surveillance of ARI in Liberia was used to assess five common strategies.
- **Algorithm Choice**: Depends on data availability, recent outbreaks, and accuracy needs.
- **Impact of Aberrations**: All algorithms struggle with smaller outbreaks when recent aberrations are present.
- Informing best approaches for HMIS-based rolling surveillance in LMICs.

# References I


C AbouZahr and T Boerma. Health information systems: the foundations of public health. *Bull World Health Organ*, 83(8): 578–583, 2005.

JW Buehler, RS Hopkins, JM Overhage, DM Sosin, V Tong, and CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the cdc working group. *MMWR Recomm Rep*, 53(RR-5):1–11, 2004.

C Chatfield. The holt-winters forecasting procedure. *J R Stat Soc C*, 27(3):264–279, 1978. doi: 10.2307/2347162.

CP Farrington, NJ Andrews, AD Beale, and MA Catchpole. A statistical algorithm for the early detection of outbreaks of infectious disease. *J R Stat Soc A*, 159(3):547–563, 1996.

# References II


IR Fulcher et al. Syndromic surveillance using monthly aggregate health systems information data: methods with application to covid-19 in liberia. *Int J Epidemiol*, 50(4): 1091–1102, 2021. doi: 10.1093/ije/dyab094.

Z Hussain, JA Ansari, M Salman, EA Khan, and J Asghar. An evaluation of acute respiratory infection surveillance systems in gilgit-baltistan pakistan. *JPMA J Pak Med Assoc*, 66(6): 682–687, 2016.

LC Hutwagner, WW Thompson, GM Seeman, and T Treadwell. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Stat Med*, 24(4):543–550, 2005. doi: 10.1002/sim.2034.

# References III


A Lal, HC Ashworth, S Dada, L Hoemeke, and E Tambo. Optimizing pandemic preparedness and response through health information systems: lessons learned from ebola to covid-19. *Disaster Med Public Health Prep*, pages 1–8, 2020.

A Noufaily, DG Enki, P Farrington, P Garthwaite, N Andrews, and A Charlett. An improved algorithm for outbreak detection in multiple surveillance systems. *Stat Med*, 32(7): 1206–1222, 2013.

N Pearce, JP Vandenbroucke, TJ VanderWeele, and S Greenland. Accurate statistics on covid-19 are essential for policy guidance and decisions. *Am J Public Health*, 110(7): 949–951, 2020.

S Unkel, CP Farrington, PH Garthwaite, C Robertson, and N Andrews. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Soc A*, 175 (1):49–82, 2012.

# References IV

MM Wagner, FC Tsui, JU Espino, VM Dato, DF Sittig, RA Caruana, et al. The emerging science of very early detection of disease outbreaks. *J Public Health Manag Pract*, 7(6):51–59, 2001.

DM Weinberger, J Chen, T Cohen, et al. Estimation of excess deaths associated with the covid-19 pandemic in the united states, march to may 2020. *JAMA Intern Med*, 180(10): 1336–1344, 2020. doi: 10.1001/jamainternmed.2020.3391.

World Health Organization. Core health facility indicators, 2021. URL `https://www.who.int/publications/m/item/core-health-facility-indicators`. Accessed May 28, 2021.

B Zareie, J Poorolajal, A Roshani, et al. Outbreak detection algorithms based on generalized linear model: a review with new practical examples. *BMC Med Res Methodol*, 23:235, 2023. doi: 10.1186/s12874-023-02050-z.