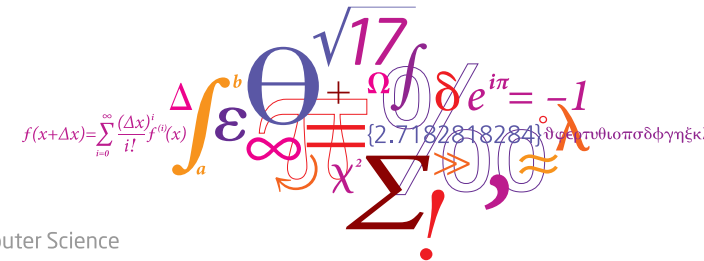
Automated and Early Detection of Disease Outbreaks

Kasper Schou Telkamp

Section for Dynamical Systems

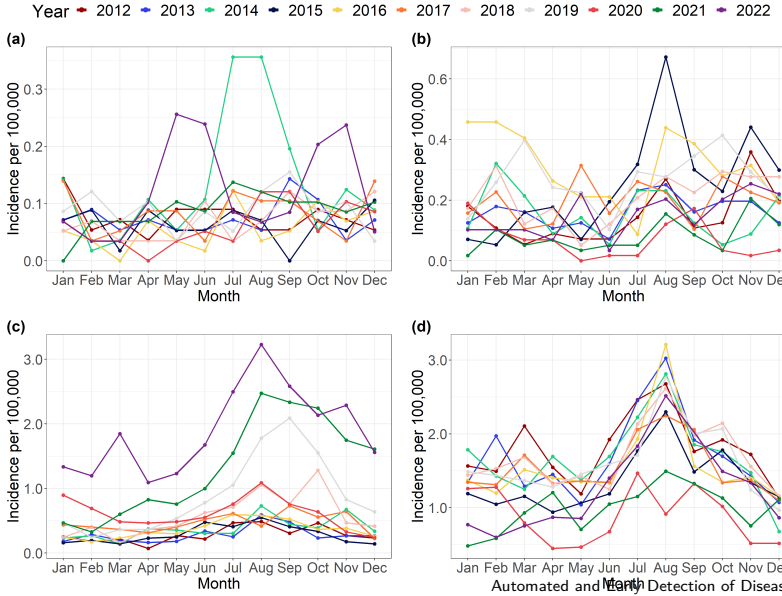


DTU Compute
Department of Applied Mathematics and Computer Science

Outline

- Data
- Novel Algorithm for Prospective Disease Outbreak Detection
  - Procedure
  - Hierarchical Poisson Normal model
  - Hierarchical Poisson Gamma model
- Proposed Models for the Novel Algorithm
  - Constant model
  - Trend model
  - Seasonality model
  - Combined trend and seasonality model
- State-of-the-art Algorithm for Prospective Disease Outbreak Detection
- Case studies
  - *Listeriosis*
  - *Shigellosis*
  - Shiga toxin (verotoxin)-producing *Escherichia coli*
  - *Salmonellosis*
- References

Data


Year 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
(a)
(b)
(c)
(d)
Incidence per 100,000
Month
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
0.0 0.1 0.2 0.3
0.0 0.2 0.4 0.6
0.0 1.0 2.0 3.0
1.0 2.0 3.0

Step 1: Modeling Framework

- Assume a hierarchical Poisson Normal or Poisson Gamma model to reference data using a log link.
- Assume count is proportional to population size n
- Incorporate covariates by supplying a model formula on the form.

$$\log(\lambda_{it}) = \boldsymbol{x}_{it}\boldsymbol{\beta} + \log(n_{it}), \quad i = 1, \ldots, m, \quad t = 1, \ldots, T \tag{1}$$

- Account for structural changes in the time series using a rolling window of width k.

Step 2: Inference of Random Effects

- Infer one-step ahead random effects u_{it_0} for each group using the fitted model.
- Define outbreak detection threshold U_{t_0} as a quantile of the second stage model's random effects distribution
- Use either a Gaussian or Gamma distribution with respective plug-in estimates

Step 3: Outbreak Detection and Parameter Estimation

- Compare inferred random effects u_{it_0} to an upper bound U_{t_0}
- Raise and alarm if the inferred random effect exceeds the upper bound, i.e. $u_{it_0} > U_{t_0}$
- Omit outbreak related observations from future parameter estimation, reducing the rolling window's sample size until the observation is discarded.

Hierarchical Poisson Normal model

$$\boldsymbol{Y}|\boldsymbol{u} \sim \text{Pois}\left(\boldsymbol{\lambda}\exp(\boldsymbol{u})\right) \tag{2a}$$

$$\boldsymbol{u} \sim \text{N}(\boldsymbol{0}, I\sigma^2) \tag{2b}$$

Hierarchical Poisson Gamma model

$$\boldsymbol{Y}|\boldsymbol{u} \sim \text{Pois}(\boldsymbol{\lambda u}) \tag{3a}$$

$$\boldsymbol{u} \sim \text{G}(\mathbf{1}/\phi, \phi) \tag{3b}$$

Constant model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \log(n_{it}) \tag{4}$$

- λ_{it} is the outbreak intensity at time t for age group i.
- $\beta(ageGroup_i)$ is the fixed effect specific to age group i.
- $\log(n_{it})$ acts as an offset, accounting for the population size at time t for age group i.

Trend model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \beta_{trend}t + \log(n_{it}) \tag{5}$$

- In addition to Model 1, includes a trend component.
- β_{trend} quantifies the rate of change in the outbreak intensity over time.

Seasonality model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \sin\left(\frac{2\pi \cdot \tau_t}{12}\right)\beta_{\sin} + \cos\left(2\frac{\pi \cdot \tau_t}{12}\right)\beta_{\cos} + \log(n_{it}) \quad (6)$$

- In addition to Model 1, incorporates an annual seasonality pattern.
- τ_t represents the time period t within a year (1-12).
- $\beta_{\sin}$ and $\beta_{\cos}$ capture the effect of the seasonal pattern.

Combined trend and seasonality model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \beta_{trend}t + \sin\left(\frac{2\pi \cdot \tau_t}{12}\right)\beta_{\sin} + \cos\left(\frac{2\pi \cdot \tau_t}{12}\right)\beta_{\cos} + \log(n_{it}) \quad (7)$$

- Builds upon previous models, combining trend and seasonality components.
- Includes both β_{trend} and $\beta_{\sin}/\beta_{\cos}$ parameters.

State-of-the-art Algorithm for Prospective Disease Outbreak Detection

State-of-the-art methods for aberration detection is presented in Salmon, Schumacher, and Höhle 2016 and implemented in the R package **surveillance**. The R package includes methods such as the Farrington method introduced by Farrington et al. 1996 together with the improvements proposed by Noufaily et al. 2013.

Listeriosis

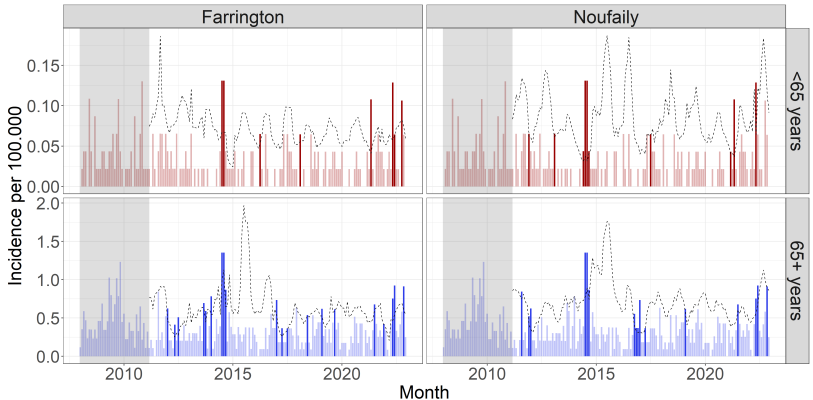


Figure: One-step ahead random effects u_{t_1} (circles) given the model described in (4) for *Listeriosis* in Denmark, 2011-2022. Poisson Normal model (right) and Poisson Gamma model (left) monitor the disease. Alarm raised (solid circle) if u_{t_1} exceeds the threshold U_{t_0} (dashed line).

Listeriosis

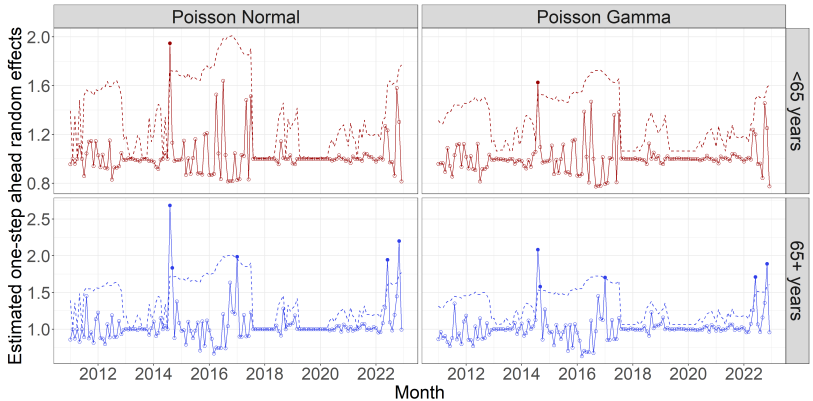


Figure: One-step ahead random effects u_{t_1} of *Listeriosis* in Denmark, 2011-2022, as circles. Poisson Normal model (top) and Poisson Gamma model (bottom) monitor the disease. Alarm raised (solid circle) if u_{t_1} exceeds the threshold U_{t_0} (dashed line).

Listeriosis

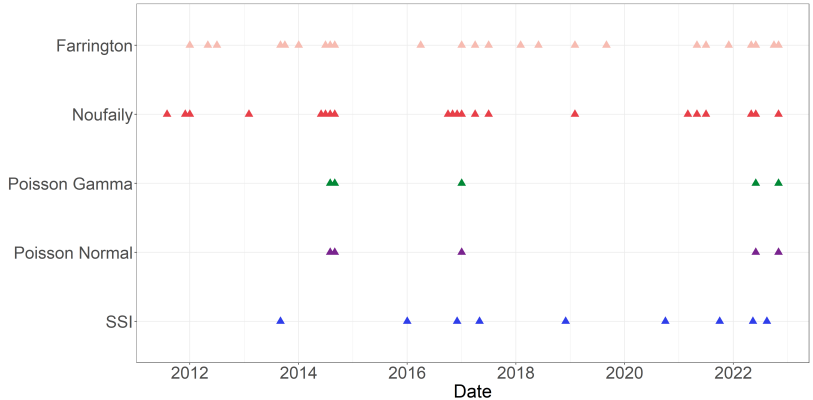


Figure: Alarm plot displaying alarms for *Listeriosis* time series using four different algorithms, along with outbreaks investigated by SSI.

Shigellosis

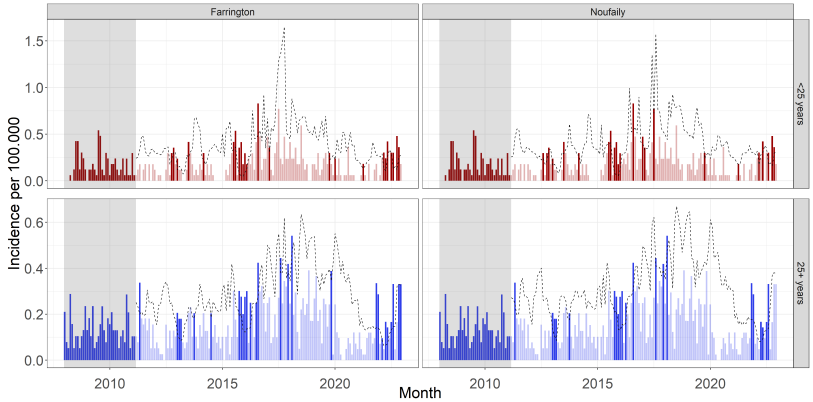


Figure: Monthly *Shigellosis* incidence per 100.000 in Denmark, 2008-2022. Monitored by (left) Farrington and (right) Noufaily method. Reference data for the estimation of model parameters from January 2008 to March 2011 (grey area). Threshold (dashed line) is computed for observations timepoints outside reference data. Alarm triggered (full opacity) if observations exceeds threshold.

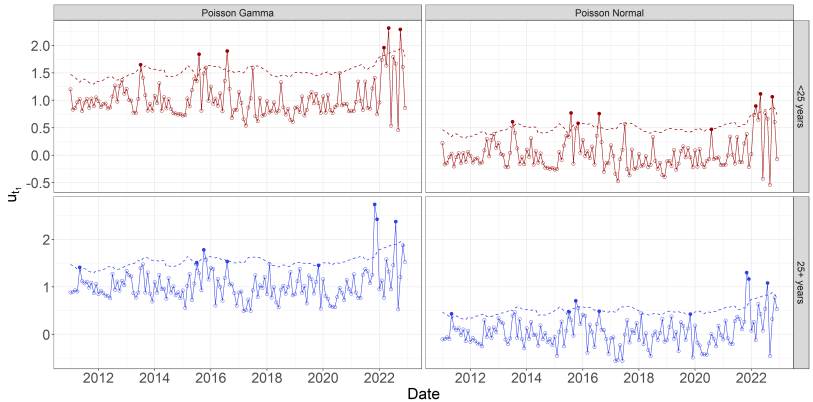
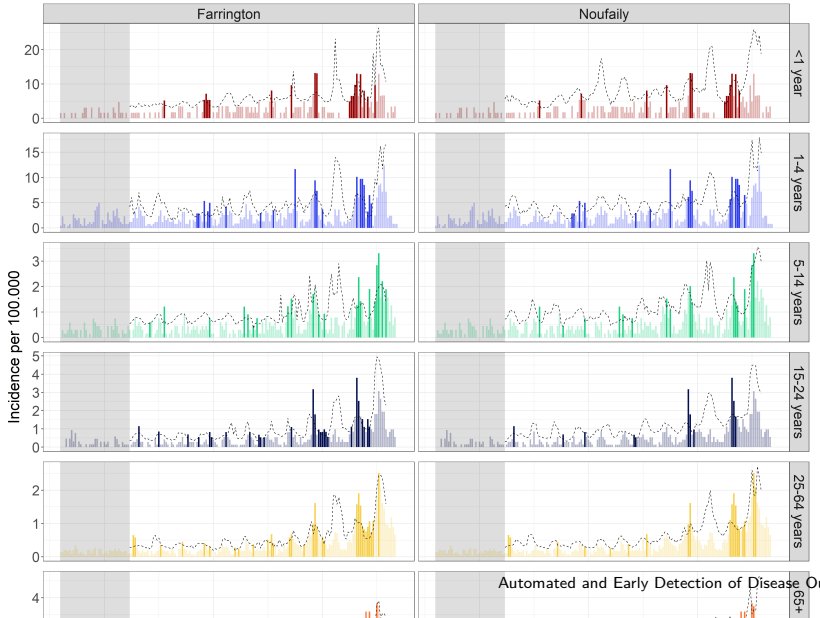


Figure: One-step ahead random effects u_{t_1} of *Shigellosis* in Denmark, 2011-2022, as circles. Poisson Normal model (top) and Poisson Gamma model (bottom) monitor the disease. Alarm raised (solid circle) if u_{t_1} exceeds the threshold U_{t_0} (dashed line).

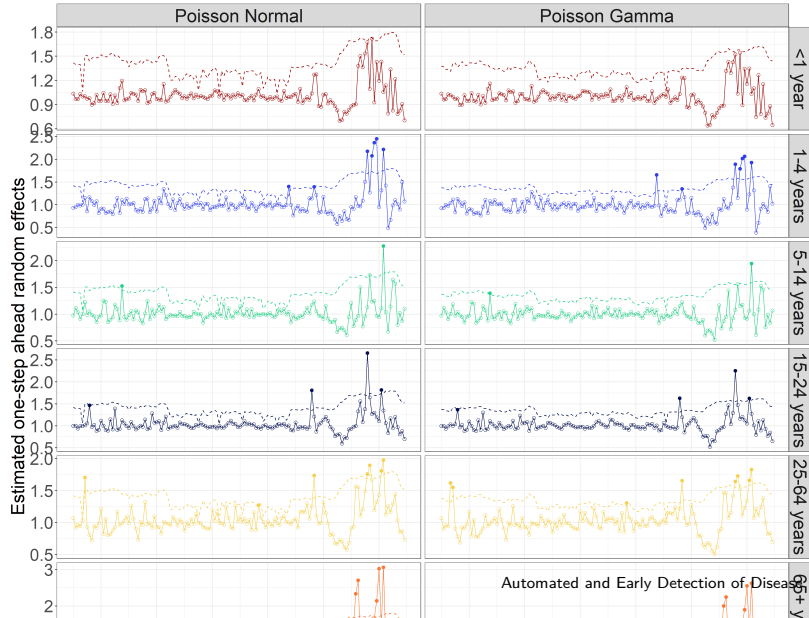
Shigellosis

Figure: Alarm plot displaying alarms for *Shigellosis* time series using four different algorithms, along with outbreaks investigated by SSI.

Shiga toxin (verotoxin)-producing *Escherichia coli*



Shiga toxin (verotoxin)-producing *Escherichia coli*



Shiga toxin (verotoxin)-producing *Escherichia coli*

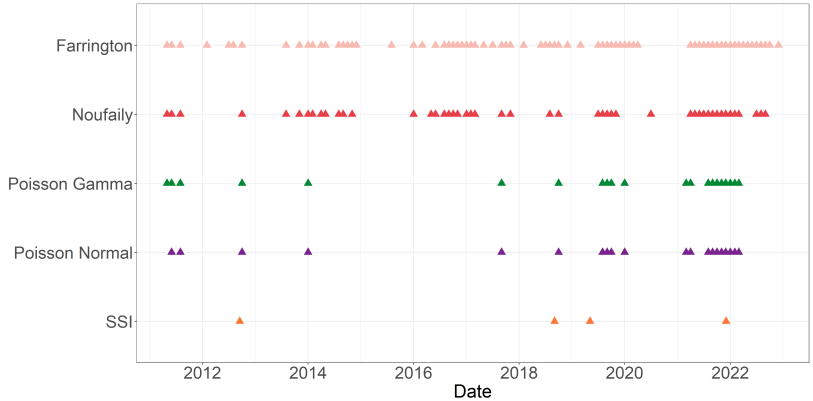
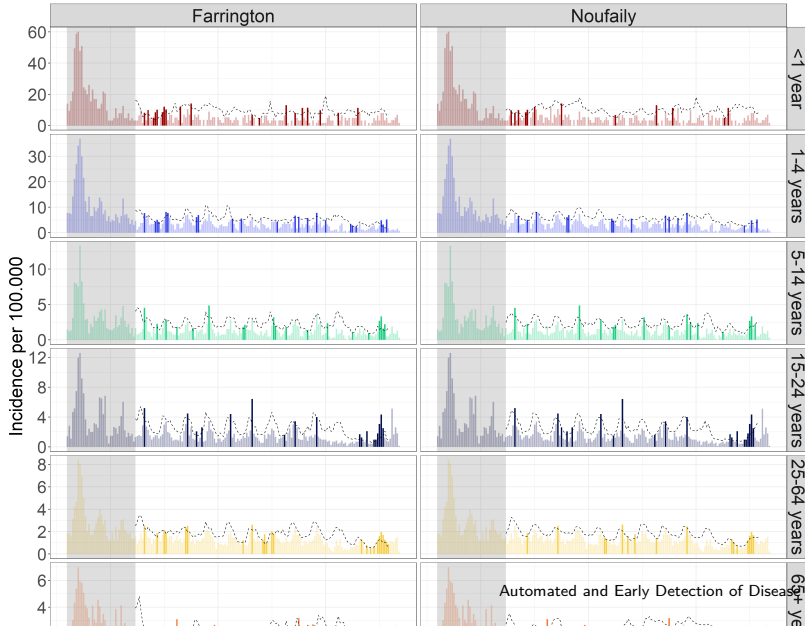
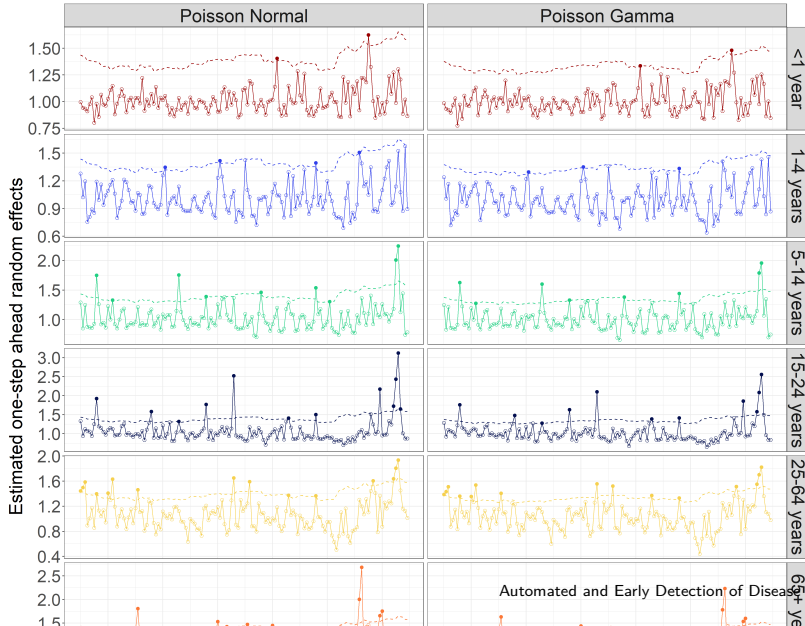


Figure: Alarm plot displaying alarms for *Shigellosis* time series using four different algorithms, along with outbreaks investigated by SSI.

Salmonellosis



Salmonellosis



Salmonellosis

Figure: Alarm plot displaying alarms for *Salmonellosis* time series using four different algorithms, along with outbreaks investigated by SSI.

References

Farrington, C. P. et al. (1996). "A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease". In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159.3, pp. 547–563. ISSN: 09641998, 1467985X. URL: http://www.jstor.org/stable/2983331 (visited on 01/27/2023).

Noufaily, Angela et al. (2013). "An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems". en. In: *Online Journal of Public Health Informatics* 32.7, pp. 1206–1222.

Salmon, Maëlle, Dirk Schumacher, and Michael Höhle (2016). "Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance". In: *Journal of Statistical Software* 70.10, pp. 1–35. DOI: 10.18637/jss.v070.i10. URL: https://www.jstatsoft.org/index.php/jss/article/view/v070i10.