Detecting changes in dispersion in COVID-19 case counts using a negative binomial model

JSM 2024

Rachael Aber, Yanming Di, Ben Dalziel

August 5, 2024

Oregon State University

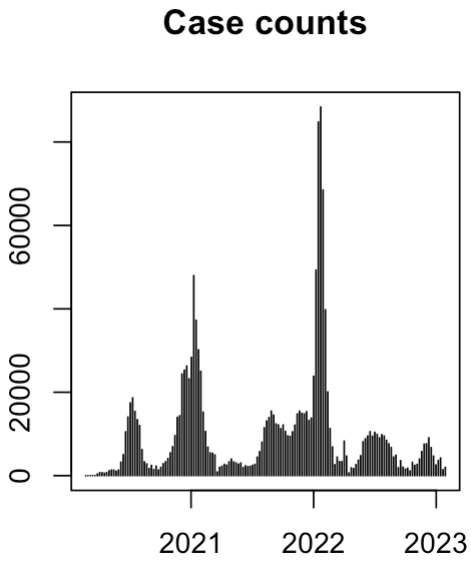
Table of Contents

Oregon State University

Table of Contents

Oregon State University

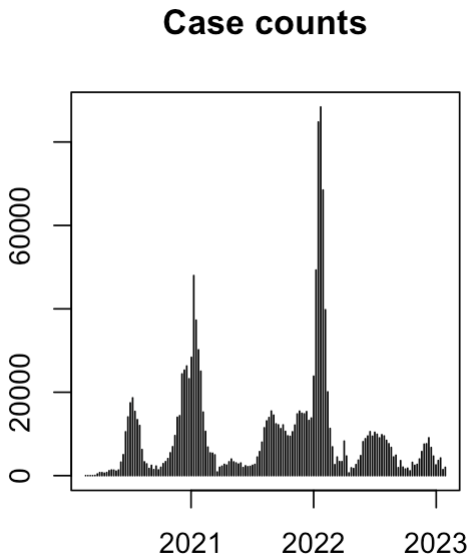
Why study variability?



- Highly variable case count time series suggest transmission heterogeneity or changes in R

Oregon State University

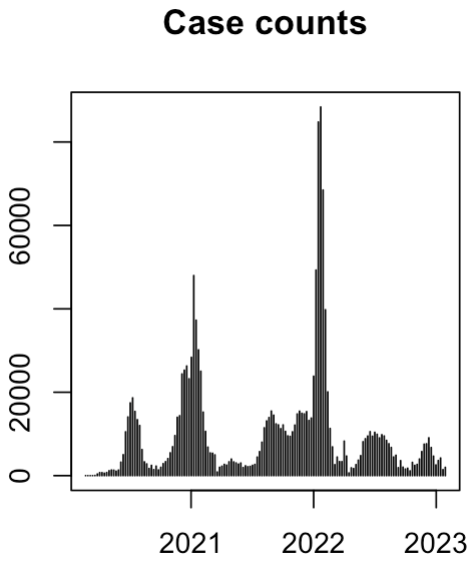
Why study variability?



- Highly variable case count time series suggest transmission heterogeneity or changes in R
- Metrics of variability are overlooked: "How is variability related to different phases of an epidemic?" [2]

Oregon State University

Why study variability?

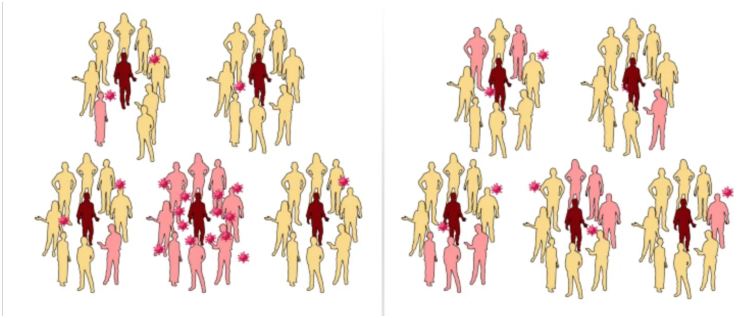


- Highly variable case count time series suggest transmission heterogeneity or changes in R
- Metrics of variability are overlooked: "How is variability related to different phases of an epidemic?" [2]
- Adam et al.[1] found that COVID-19 transmission heterogeneity decreased over time

Oregon State University

Why study variability?

- Dispersion of a case count time series may be a useful metric–part of a framework that models variance flexibly

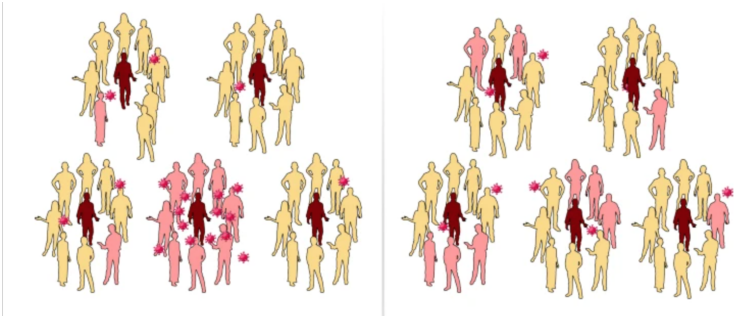


[5]

Oregon State University

Why study variability?

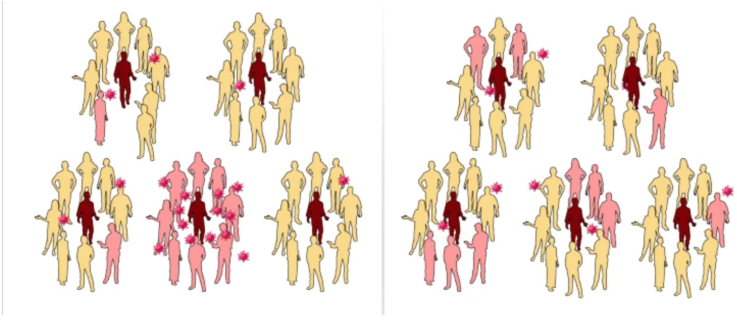
- Dispersion of a case count time series may be a useful metric–part of a framework that models variance flexibly
- A 'mean crowding' parameter [4] was proposed: mean number per individual of other individuals in the same quadrat



[5]

Why study variability?

- Dispersion of a case count time series may be a useful metric–part of a framework that models variance flexibly
- A 'mean crowding' parameter [4] was proposed: mean number per individual of other individuals in the same quadrat
- Useful way to think about dispersion in case count time series, degree of dispersion is degree of clustering/crowding of cases



[5]

Table of Contents

Oregon State University

Introduction to the method

- Let λ_t be epidemic intensity at time t, I_t be incidence at time t, and θ_t be dispersion at time t

Oregon State University

Introduction to the method

- Let λ_t be epidemic intensity at time t, I_t be incidence at time t, and θ_t be dispersion at time t
- $$\mathrm{I_t} = \mathrm{NB}(\mu = \lambda_\mathrm{t}, \theta_\mathrm{t} = \mathrm{I_{t-1}}) \quad (1)$$

  [3]

Oregon State University

Introduction to the method

- Let λ_t be epidemic intensity at time t, I_t be incidence at time t, and θ_t be dispersion at time t
- $$I_t = NB(\mu = \lambda_t, \theta_t = I_{t-1}) \quad (1)$$

  [3]
- Adjusted for population size using an offset in the model

Oregon State University

Introduction to the method

- Let λ_t be epidemic intensity at time t, I_t be incidence at time t, and θ_t be dispersion at time t
- $$\mathrm{I_t} = \mathrm{NB}(\mu = \lambda_\mathrm{t}, \theta_\mathrm{t} = \mathrm{I_{t-1}}) \quad (1)$$

  [3]
- Adjusted for population size using an offset in the model
- Linear predictor includes a natural spline in time to account for autocorrelation in case counts

$$log(E[Y_i]/n_i) = \beta_1 h_1(t_i) + \beta_2 h_2(t_i) + \beta_3 h_3(t_i) \quad (2)$$

$$log(E[Y_i]) - log(n_i) = \beta_1 h_1(t_i) + \beta_2 h_2(t_i) + \beta_3 h_3(t_i) \quad (3)$$

$$log(E[Y_i]) = \beta_1 h_1(t_i) + \beta_2 h_2(t_i) + \beta_3 h_3(t_i) + log(n_i) \quad (4)$$

Oregon State University

Introduction to the method

- Let λ_t be epidemic intensity at time t, I_t be incidence at time t, and θ_t be dispersion at time t
- $$\mathrm{I_t} = \mathrm{NB}(\mu = \lambda_\mathrm{t}, \theta_\mathrm{t} = \mathrm{I_{t-1}}) \quad (1)$$

  [3]
- Adjusted for population size using an offset in the model
- Linear predictor includes a natural spline in time to account for autocorrelation in case counts

$$log(E[Y_i]/n_i) = \beta_1 h_1(t_i) + \beta_2 h_2(t_i) + \beta_3 h_3(t_i) \quad (2)$$

$$log(E[Y_i]) - log(n_i) = \beta_1 h_1(t_i) + \beta_2 h_2(t_i) + \beta_3 h_3(t_i) \quad (3)$$

$$log(E[Y_i]) = \beta_1 h_1(t_i) + \beta_2 h_2(t_i) + \beta_3 h_3(t_i) + log(n_i) \quad (4)$$

- Model fit on a rolling basis to each time series

Oregon State University

Negative binomial model

- $$f_t(I) = \binom{I + \theta - 1}{I} \left(\frac{\mu}{\mu + \theta}\right)^I \left(\frac{\theta}{\mu + \theta}\right)^\theta \quad (5)$$

Oregon State University

Negative binomial model

- $$f_t(I) = \binom{I+\theta-1}{I}\left(\frac{\mu}{\mu+\theta}\right)^{I}\left(\frac{\theta}{\mu+\theta}\right)^{\theta} \tag{5}$$
- $$E(I) = \mu \tag{6}$$
$$Var(I) = \mu + \frac{\mu^2}{\theta} \tag{7}$$

Oregon State University

Negative binomial model

- $$f_t(I) = \binom{I + \theta - 1}{I} \left(\frac{\mu}{\mu + \theta}\right)^I \left(\frac{\theta}{\mu + \theta}\right)^\theta \tag{5}$$

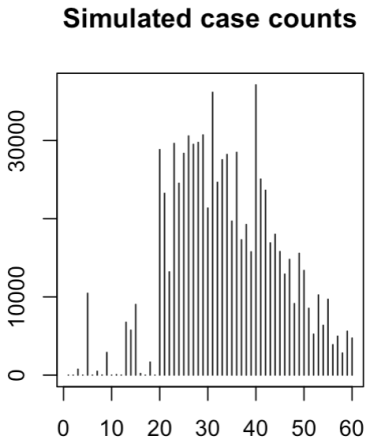
- $$E(I) = \mu \tag{6}$$

$$Var(I) = \mu + \frac{\mu^2}{\theta} \tag{7}$$

- LRT framework: at each time step along a time series, fit null model and θ change model, conduct LRT to produce p-value

Oregon State University

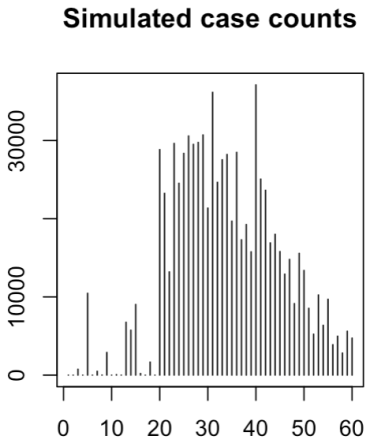
Simulated time series data set



- Validity/power simulations: Gaussian and uniform epidemic curves with an attack rate of 0.1

Oregon State University

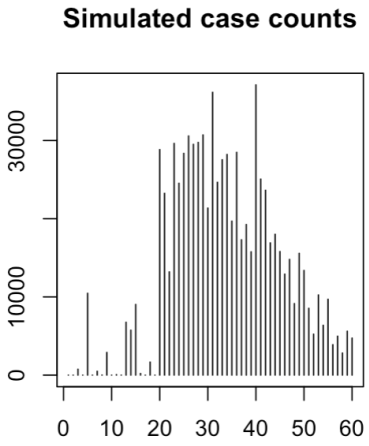
Simulated time series data set



- Validity/power simulations: Gaussian and uniform epidemic curves with an attack rate of 0.1
- Varied magnitude of θ change, location of the change, underlying population size, and epidemic curve shape

Oregon State University

Simulated time series data set



- Validity/power simulations: Gaussian and uniform epidemic curves with an attack rate of 0.1
- Varied magnitude of θ change, location of the change, underlying population size, and epidemic curve shape
- Epidemic curves over 60 time steps each were produced, and the LRT procedure was applied to each

Oregon State University

Performance: simulated data

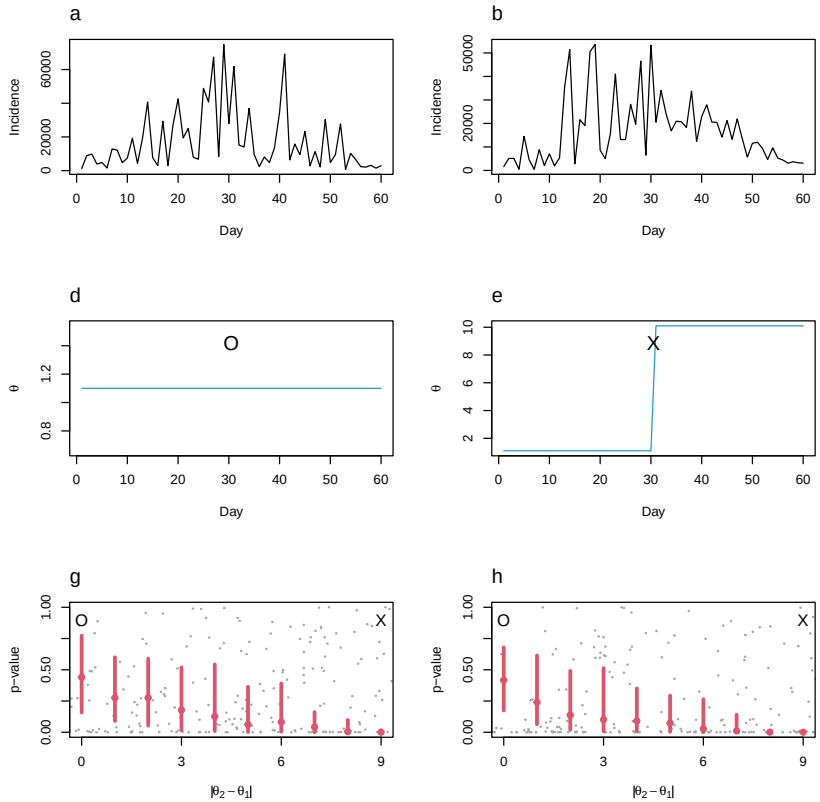


Figure 1: Detecting dispersion changes in incidence time series. A/B: Simulated incidence when dispersion is constant/changes. C/D: Constant/changing dispersion used above. E: Performance of LRT with simulated data that has different absolute differences in theta (x-axis of each pane) illustrates p-value distribution across different population sizes. O and X mark the null and alternative hypotheses indicated in panels C and D.

Oregon State University

Application to empirical data

- Weekly case counts from US counties between 2020-01-04 and 2023-03-18

Oregon State University

Application to empirical data

- Weekly case counts from US counties between 2020-01-04 and 2023-03-18
- Estimated θ_t for 154 time steps for 144 US counties (IRLS procedure implemented via the NBPSeq package[7] and from Di et al.[6])

Oregon State University

Application to empirical data

- Weekly case counts from US counties between 2020-01-04 and 2023-03-18
- Estimated θ_t for 154 time steps for 144 US counties (IRLS procedure implemented via the NBPSeq package[7] and from Di et al.[6])
- Investigated large counties (largest three counties in each state)

Oregon State University

Table of Contents

Oregon State University

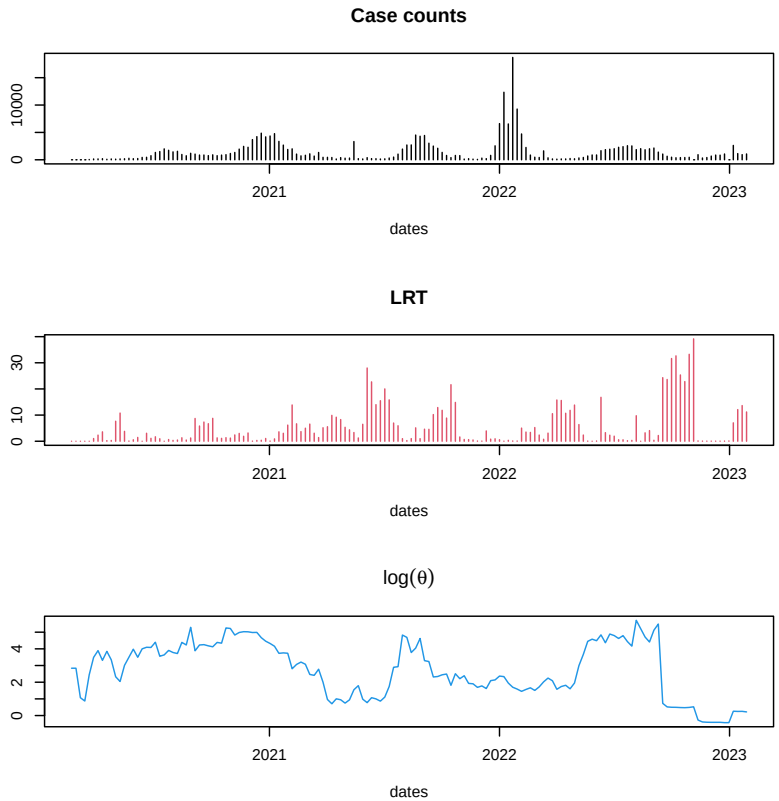


Figure 2: Method applied to case counts between 2020-01-04 and 2023-03-18 for Jefferson County, AL. A: Case counts . B: LRT statistic C: Log dispersion parameter.

Oregon State University

Results: empirical data

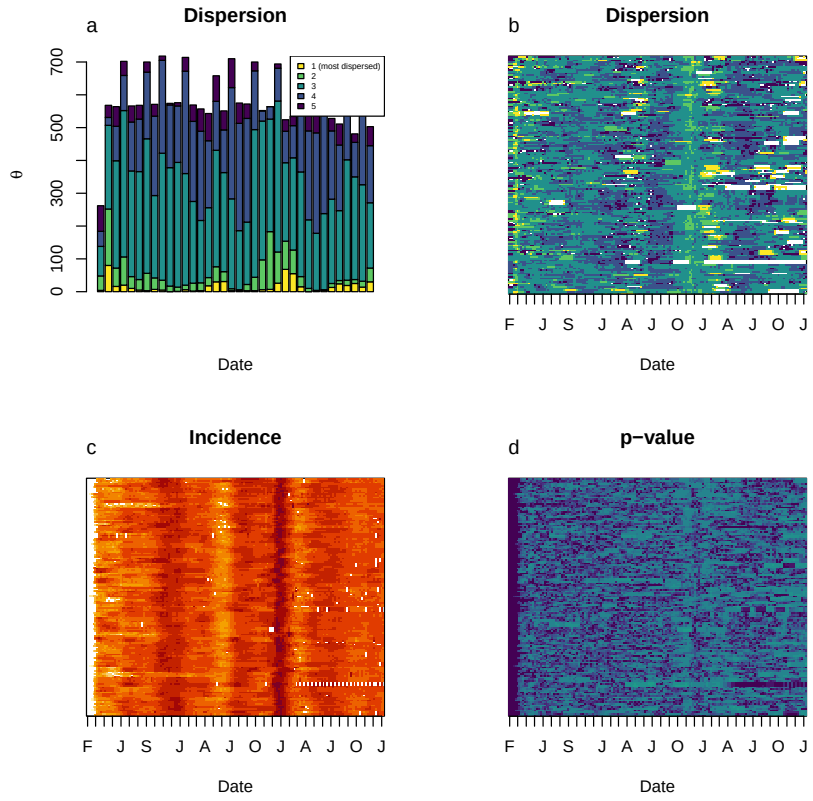


Figure 3: Incidence and dispersion in large counties in the US. A: Binned log of the dispersion parameter. B: Log of the dispersion parameter for each of the large counties (y-axis). C: Log incidence for each of the large counties (y-axis). D: LRT p-values for each of the large counties (y-axis).

Oregon State University

Table of Contents

Oregon State University

Concluding remarks

- LRT procedure performed well for relevant range of population sizes

Oregon State University

Concluding remarks

- LRT procedure performed well for relevant range of population sizes
- Dispersion is high at unexpected times (near peak incidence)

Oregon State University

Concluding remarks

- LRT procedure performed well for relevant range of population sizes
- Dispersion is high at unexpected times (near peak incidence)
- Concentration of changes in dispersion parameter near peak incidence

Oregon State University

Concluding remarks

- LRT procedure performed well for relevant range of population sizes
- Dispersion is high at unexpected times (near peak incidence)
- Concentration of changes in dispersion parameter near peak incidence
- Methods that use population-level data are important: timing/allocation of public health resources can be achieved with limited resources

Oregon State University

Table of Contents

Oregon State University

[1] Dillon Adam et al. *Time-varying transmission heterogeneity of SARS and COVID-19 in Hong Kong*. Tech. rep. ISSN: 2693-5015 Type: article. Mar. 2022.

[2] Matthew Graham et al. "Measles and the canonical path to elimination". en. In: *Science* 364.6440 (May 2019), pp. 584–587. ISSN: 0036-8075, 1095-9203.

[3] Bryan T. Grenfell, Ottar N. Bjørnstad, and Bärbel F. Finkenstädt. "Dynamics of Measles Epidemics: Scaling Noise, Determinism, and Predictability with the TSIR Model". In: *Ecological Monographs* 72.2 (2002). Publisher: Ecological Society of America, pp. 185–202. ISSN: 0012-9615.

[4] Monte Lloyd. "'Mean Crowding'". In: *The Journal of Animal Ecology* 36.1 (Feb. 1967), p. 1. ISSN: 00218790.

[5] Bjarke Frost Nielsen, Kim Sneppen, and Lone Simonsen. "The counterintuitive implications of superspreading diseases". en. In: *Nature Communications* 14.1 (Oct. 2023), p. 6954. ISSN: 2041-1723.

[6] Di Yanming et al. "The NBP Negative Binomial Model for Assessing Differential Gene Expression from RNA-Seq". In: *Statistical Applications in*

Oregon State University

Thanks!

Oregon State University