Time-series modeling of epidemics in complex populations: detecting changes in incidence volatility over time

Rachael Aber[1], Yanming Di[2], and Benjamin D. Dalziel[1,3]

[1]Department of Integrative Biology, Oregon State University, Corvallis, Oregon, USA
[2]Department of Statistics, Oregon, Oregon State University, Corvallis, Oregon, USA
[3]Department of Mathematics, Oregon State University, Corvallis, Oregon, USA

Abstract

Trends in infectious disease incidence provide important information about epidemic dynamics and prospects for control. Higher-frequency variation around incidence trends can shed light on the processes driving epidemics in complex populations, as transmission heterogeneity, shifting landscapes of susceptibility, and fluctuations in reporting can impact the volatility of observed case counts. However, measures of temporal volatility in incidence, and how volatility changes over time, are often overlooked in population-level analyses of incidence data, which typically focus on moving averages. Here we present a statistical framework to quantify temporal changes in incidence dispersion and detect discrete shifts in the dispersion parameter, which may signal new epidemic phases. We apply the method to COVID-19 incidence data in 144 US counties from the January 1st, 2020 to March 23rd, 2023. Theory predicts that dispersion should be inversely proportional to incidence, however our method reveals pronounced temporal trends in dispersion that are not explained by incidence alone, but which are replicated across counties. In particular, dispersion increased around the major surge in cases in 2022, and highly overdispersed patterns became more frequent later in the time series. These findings suggest that heterogeneity in transmission, susceptibility, and reporting could play important roles in driving large surges and extending epidemic duration. The dispersion of incidence time series can contain structured information which enhances predictive understanding of the underlying drivers of transmission, with potential applications as leading indicators for public health response.

Author summary

Understanding patterns in infectious disease incidence is crucial for understanding epidemic dynamics and for developing effective public health responses. However, traditional metrics used to quantify incidence patterns often overlook variability as an important characteristic of incidence time series. Quantifying variability around incidence trends can elucidate important underlying processes, including transmission heterogeneity. We developed a statistical framework to quantify temporal changes in case count dispersion within a single time series and applied the method to COVID-19 case count data. We found that conspicuous shifts in dispersion occurred across counties concurrently, and that these shifts were not explained by incidence alone. Dispersion increased around peaks in incidence such as the major surge in cases in 2022, and dispersion also increased as the pandemic progressed. These increases potentially indicate transmission heterogeneity, changes in the susceptibility landscape, or that there were changes in reporting. Shifts in dispersion can

also indicate shifts in epidemic phase, so our method provides a way for public health officials to anticipate and manage changes in epidemic regime and the drivers of transmission.

Introduction

Time series of infectious disease incidence appear, to varying degrees, "noisy", showing higher frequency fluctuations (e.g., day-to-day or week-to-week fluctations) around trends at the broader temporal ranges typical for epidemic curves (e.g., months or years). Short-term fluctuations in incidence time series are often caused in part by variable reporting, but also reflect the population-level impacts of transmission heterogeneity, and/or changes in susceptiblility Lloyd-Smith et al. [2005], Kirkegaard and Sneppen [2021], Sun et al. [2021], Guo et al. [2023], Ko et al. [2023]. Metrics of variability in incidence time series may therefore carry information regarding underlying drivers of transmission, and offer a relatively unexplored avenue for understanding epidemic dynamics.

Contact tracing data has revealed temporal changes in the variability of individual reproductive numbers (i.e., the expected number of secondary infections that will result if a particular individual becomes infected), quantified by shifts in the dispersion parameter of the offspring distribution in branching process models Guo et al. [2023], Ko et al. [2023]. However, the scaling from individual-level transmission heterogeneity to population-level epidemic dynamics is not fully understood. In addition, traditional contact tracing is very resource intensive, and although new approaches using digital technologies may improve its speed and availability Kretzschmar et al. [2020], there is a need for complementary population-level analyses that can estimate heterogeneity using incidence data, which is more widely available. The importance of considering population-level variability and its relationship to individual-level variability is further highlighted by the finding that a combination of individual-based and population-based strategies was required for SARS-CoV-2 control Sun et al. [2021]. An important challenge therefore is to develop methods that can detect changes in population-level variability in incidence time series, and to interpret these changes in terms of underlying transmission processes.

Emerging statistical techniques are leveraging variability in epidemic time series to enhance understanding of disease dynamics at the population level. For example, a recently-developed method estimates the dispersion parameter of the offspring distribution, which quantifies heterogeneity in secondary cases generated by an infected individual, directly from population-level incidence data Kirkegaard and Sneppen [2021]. Clustering of cases has also been estimated directly from incidence data Schneckenreither et al. [2023]. Another important application links variability in incidence to epidemic phases; for example, changes in the mean and interannual coefficient of variation of measles incidence have been used to identify a country's position on the path to elimination, providing insights into vaccination strategies and epidemiological dynamics Graham et al. [2019]. Analysis of the shape of epidemic curves for influenza in cities may identify contexts where incidence is focussed more intensely (proportionally more infections in a smaller span of time) with relevance for the sensitivity of cities to forcing from exogenous drivers of transmission (e.g., climate) and for surge capacity in the health system Dalziel et al. [2018], Wallinga [2018].

What drives indicence dispersion and how does it relate to the underlying branching process of transmission, and to observations of cases? Under a wide range of configurations for a branching process model the number of infected individuals I_t at time t will have a negative binomial distribution Kendall [1949], Grenfell et al. [2002], $I_t \sim NB\left(\mu_t, \theta_t\right)$, where μ_t is the expectation for I_t and θ_t is the dispersion parameter. The variance is related to the mean and dispersion parameters by $\mathrm{Var}[I_t] = \mu_t + \mu_t^2/\theta_t$, so smaller values of the dispersion parameter θ_t correspond to increasing amounts of dispersion, which increase the amounts by which the variance in realized

number infected I_t exceeds the expected value, μ_t. Conversely, the distribution of I_t tends to a Poisson distribution (where the variance equals mean) as θ_t becomes large. The negative binomial distribution may also accurately model a time series if there is a changing process mean within a time step: for example, if the mean of a Poisson distribution itself follows a gamma distribution, the resulting distribution is negative binomial. Negative binomial regression (in contrast to Poisson regression) can account for unobserved heterogeneity, time dependence in the rate of a process and contagion within a time step that all lead to overdispersion Barron [1992].

An interpretation of the dispersion parameter for a time series model of counts is that events are $1 + \theta^{-1}$ times as "crowded" in time relative to a Poisson process with the same mean Lloyd [1967] (see Supplemental Information). For example, $\theta = 1$ corresponds to a situation where the average number of infections in the same time step as a randomly selected case will exceed the Poisson expectation by a factor of two. In a simple example relevant to surge capacity in healthcare systems, $\theta = 1$ implies that a random infectious individual visiting the emergency department at a hospital would find it on average to be twice as crowded with other infectious individuals (infected by the same pathogen) than expected for a Poisson process with the same incidence rate.

In a sufficiently large host population, and when the infectious pathogen can be assumed to spread in nonoverlapping generations, incidence is often modeled as

$$\mathrm{I_{t+1}} = NB(\mu = R_t I_t, \theta_t = I_t) \tag{1}$$

where time-varying reproductive number R_t gives the expected number of secondary infections aqcuired from an infected indivual at time t, and the generation time is set to 1 without loss of generality Kendall [1949], Bjørnstad et al. [2002]. Setting $\theta_t = I_t$ arises from the assumption that individuals who acquire the infection at time t form independent lineages with identically distributed local rate parameters. However, this requires that susceptible depletion in one lineage does not affect another, that transmission rates are equal across lineages, and that reporting rates do not vary across lineages.

In practice, these assumption will not often hold, and our aim in this paper is to develop, test and apply an alternative approach, which makes data-driven estimates of θ_t, including identifying probably timepoints when θ is changing rapidly, which may help to reveal how heterogeneity in transmission, susceptibility and reporting, are impacting incidence volatlity.

Methods

Acknowledgments

RA work on this project was supported by X and Y. BDD work on this project was supported by the National Science Foundation (NSF) grants (X, Y) and by the David and Lucile Packard Foundation.

References

References

David N Barron. The analysis of count data: Overdispersion and autocorrelation. *Sociological methodology*, pages 179–220, 1992.

Ottar N Bjørnstad, Bärbel F Finkenstädt, and Bryan T Grenfell. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series sir model. *Ecological monographs*, 72 (2):169–184, 2002.

Benjamin D Dalziel, Stephen Kissler, Julia R Gog, Cecile Viboud, Ottar N Bjørnstad, C Jessica E Metcalf, and Bryan T Grenfell. Urbanization and humidity shape the intensity of influenza epidemics in us cities. *Science*, 362(6410):75–79, 2018.

Matthew Graham, Amy K Winter, Matthew Ferrari, Bryan Grenfell, William J Moss, Andrew S Azman, C Jessica E Metcalf, and Justin Lessler. Measles and the canonical path to elimination. *Science*, 364(6440):584–587, 2019.

Bryan T Grenfell, Ottar N Bjørnstad, and Bärbel F Finkenstädt. Dynamics of measles epidemics: scaling noise, determinism, and predictability with the tsir model. *Ecological monographs*, 72(2): 185–202, 2002.

Zihao Guo, Shi Zhao, Shui Shan Lee, Chi Tim Hung, Ngai Sze Wong, Tsz Yu Chow, Carrie Ho Kwan Yam, Maggie Haitian Wang, Jingxuan Wang, Ka Chun Chong, et al. A statistical framework for tracking the time-varying superspreading potential of covid-19 epidemic. *Epidemics*, 42:100670, 2023.

David G Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–282, 1949.

Julius B Kirkegaard and Kim Sneppen. Superspreading quantified from bursty epidemic trajectories. *Scientific Reports*, 11(1):24124, 2021.

Yura K Ko, Yuki Furuse, Kanako Otani, Masato Yamauchi, Kota Ninomiya, Mayuko Saito, Takeaki Imamura, Alex R Cook, Tadayuki Ahiko, Shunji Fujii, et al. Time-varying overdispersion of sars-cov-2 transmission during the periods when different variants of concern were circulating in japan. *Scientific Reports*, 13(1):13230, 2023.

Mirjam E Kretzschmar, Ganna Rozhnova, Martin CJ Bootsma, Michiel van Boven, Janneke HHM van de Wijgert, and Marc JM Bonten. Impact of delays on effectiveness of contact tracing strategies for covid-19: a modelling study. *The Lancet Public Health*, 5(8):e452–e459, 2020.

Monte Lloyd. Mean crowding'. *The Journal of Animal Ecology*, pages 1–30, 1967.

James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

Günter Schneckenreither, Lukas Herrmann, Rafael Reisenhofer, Niki Popper, and Philipp Grohs. Assessing the heterogeneity in the transmission of infectious diseases from time series of epidemiological data. *Plos one*, 18(5):e0286012, 2023.

Kaiyuan Sun, Wei Wang, Lidong Gao, Yan Wang, Kaiwei Luo, Lingshuang Ren, Zhifei Zhan, Xinghui Chen, Shanlu Zhao, Yiwei Huang, et al. Transmission heterogeneities, kinetics, and controllability of sars-cov-2. *Science*, 371(6526):eabe2424, 2021.

Jacco Wallinga. Metropolitan versus small-town influenza. *Science*, 362(6410):29–30, 2018.