Estimators for disease dynamics with imperfect surveillance (EDDIS)

Tom Ingersoll[1]
[1]U.S. Army Combat Capabilities Development Command Chemical Biological Center, Aberdeen Proving Ground, MD.

ABSTRACT:

This report presents a method for estimating infective disease-dynamics parameters when contact rates are uneven, surveillance data are not systematically sampled, and cases are underreported. An important parameter for predictive infectious disease models is the effective reproductive number (R_e). R_e determines the rate at which new infections occur and respond to intervention strategies, such as vaccination, quarantine, and social distancing. However, accurate estimation of R_e is complicated by shortcomings in surveillance data collection, and these shortcomings are difficult to mitigate through changes in sampling methods. The author proposes that estimation of R_e is not necessary to model changes in disease dynamics; rather, the basic reproductive number R_0 may be used along with contact parameters derived from network characteristics within the host population. Further, estimates of R_0 can be derived from imperfect surveillance data through application of hierarchical methods that correct for underreporting by using explicit estimates of detection probabilities. A hierarchical data-assimilative method for improving parameter estimates in predictive models when data are imperfectly collected is demonstrated in this report. Accurate estimates of changes in disease dynamics can inform management decisions and mitigation strategies.

METHODS

Figure 2. Compartments of the SEIR Model.

A set of 4 categories useful for describing the state of infection within hosts includes 1) Susceptible (*S*): hosts that can become infected, but have not yet been exposed to the disease to the disease, 2) Exposed (*E*): latent hosts that have recently acquired the disease but are not yet infectious, 3) Infective (*I*): hosts that are able to spread disease to others, and 4) Recovered (*R*): hosts that have developed immunity so can no longer transmit the disease. Models containing these categories are called SEIR models (1;2). SEIR models usually require an assumption that all hosts that have never had the disease are equally likely to become infected, the well-mixed population assumption. Models that relax this assumption are called network models. Parameterizing network models usually requires extensive contact-tracing records. EDDIS is an implicit network model, where network characteristics are inferred from surveillance data and contact tracing data is not required. Because natural populations are not well-mixed, the number of hosts within the *S* class is depends on clustering of hosts (Figure 1). These clusters are called nodes (3). Therefore, the size of class S is lower than the total population size. A system of differential equations describing SEIR dynamics is given below.

$$\frac{dS}{dT} = \nu - (\beta I + \mu)S \quad (1)$$

$$\frac{dE}{dT} = \beta SI - (\mu + \sigma)E \quad (2)$$

$$\frac{dI}{dT} = \sigma E - (\mu + \gamma)I \quad (3)$$

$$\frac{dR}{dT} = \gamma I - \mu R \quad (4)$$

Equations 1:4 System of differential equations for SEIR dynamics.

S,E,I,R are the counts of hosts in each of the compartmental model classes, as given above (Figure2). The birth rate is ν. β is the epidemic growth rate. The rate of transition from exposed to infective is σ. The death rate is μ. The recovery rate is γ. The cumulative case count is the sum of *I* and *R*, and can frequently be found in surveillance reports. β is a function of R_0,which can be estimated using EDDIS (4), which is written in R software (5), and uses the R differential equation solver deSolve (6).

Hierarchical models were used to correct underreporting, in cumulative case counts and were extended so as not to require an assumption of homogeneous detection (7). A method for estimating detection probability using predicted values from SEIR models and count data from surveillance is found using Isaac Newton's Binomial Theorem (Equation 5; 8).

$$P = \binom{n}{k}(p)^k(1-p)^{n-k} \quad (5)$$

Equation 5. Newton's Binomial Theorem.

The true cumulative case count from the dynamic model (*I+R*) is *n.*, the underreported cumulative case count from surveillance data is *k.*, the detection probability is *p*, and *P* is the probability of the parameters *n* and *p*, given the datum *k*. The product of *P* across all the data is the model likelihood. The result is a maximum likelihood method for estimating *p,* estimates of the true number of cases in the population, and a vehicle to adjust the parameters R_0 and *S* which produce the value *n* (Figure 4). Parametric adjustment using data assimilation is a method often attributed to Gauss (9).

Because the model is networked, S is the sum of uninfected hosts in the nodes of the network that are in contact with infected hosts. A temporally iterative method was used to estimate the links and node size to be added to S at various time intervals (Figure 4).

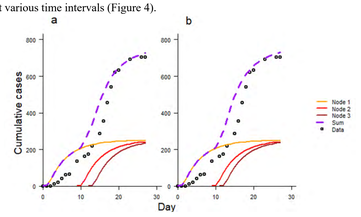

Figure 2. Adjusting parameters and link timing using maximum likelihood.

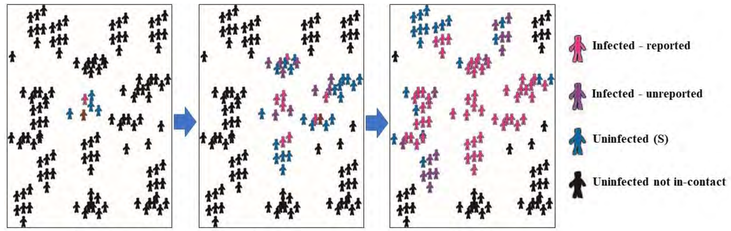

Figure 1. Transmission of a hypothetical infectious disease in a clustered population.

Transmission begins on Week 1, with detected (pink hosts) and undetected cases (purple hosts), within a small cluster (node) of susceptible hosts (S: blue hosts). Remaining nodes are out-of-contact (black hosts) in isolation until Week 2, when links form, spreading transmission into five nodes. Detection and transmission are spatially heterogeneous. By Week 3, most of the population is involved, with 4 nodes and some isolated hosts remaining out-of-contact (black hosts). Transmission, susceptibility, and detection are spatially and temporally autocorrelated.

RESULTS – Diamond Princess Cruise Ship CoViD-19 Outbreak (4)

Open-source data for the CoViD-19 pandemic are available for the Diamond Princess Cruise Ship (10). Diamond Princess began with a population of 2670 passengers and 1100 staff. By the end of the sampling period all those infected had been removed shoreward into quarantine (11). Additionally, by the end of the sampling period complete surveillance had been attained, so that all passengers had been repeatedly tested for CoViD-19 using polymerase chain-reaction (PCR) methods. It was determined that a well-mixed model could not adequately describe the dynamics of the observed case counts, regardless of the growth rate (Figure 3).

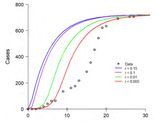

Figure 3. The failure of the well-mixed assumption.

A hierarchical network model was fit, including estimates of detection probability (Figure 4).

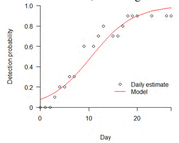

Figure 4. Logistic model for detection probability.

Because testing frequency increased, and all passengers were tested repeatedly by the end of the sampling period, a valid model with underreporting should converge on testing results towards the end of the period, and estimated rates of detection should increase across the period towards $p = 1$.

Maximum likelihood methods were used to adjust the estimate of R_0, and the true case count *n* (corrected for underreporting in the surveillance data: Figure 5).

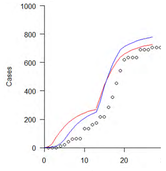

Figure 5. Adjusting R_0 using maximum likelihood.

The maximum likelihood adjustment resulted in closer fit to the data (open circles) in the preferred model (red trace), close to the end of the sampling when the detection probability was close to 1. In order to achieve likelihoods greater than 0, modeled true counts *(n)* must be greater or equal to the observed data due to underreporting.

Model terms were used to fit predictive models to hypothetical populations (Figure 6).

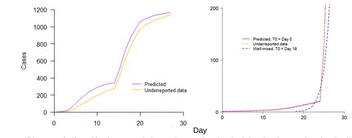

Figure 6. Predictive models a: hypothetical ship, b: hypothetical village.

Figure 6a depicts a hypothetical shipboard outbreak, with a larger population than Diamond Princess. Figure 6b, an outbreak in a hypothetical village, with slow initial growth transitioning to rapid exponential growth as typifies many network models. The well-mixed model (dashed blue trace) cannot explain the first 18 days of data.

RESULTS – 2022 US Monkeypox Outbreak

Cumulative case counts for the 2022 monkeypox outbreak were reported by the World Health Organization, and are available in open-source (12). Monkeypox is a zoonosis, with alternative human and animal hosts (13). Data for the United States were chosen because the data were uncomplicated by contacts with potential infected wildlife hosts, which would require a more complicated compartmental model (14). Data started with the first reported United States cases on June 3, 2022, and total of 79 case counts were reported, with records missing for the remainder a 118 day period. EDDIS computer codes were extended to accommodate the missing records as part of an ongoing generalization effort. A analytic hierarchical model was produced using published values for R_0 (13), and estimating the true cumulative case count *n*, the size of the susceptible class *S*, and the detection probability *p* (Figure 7).

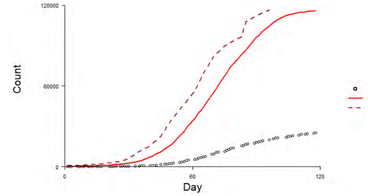

Figure 7. Analytic model for the 2022 US Monkeypox outbreak.

Detection was uniformly low at 0.2, indicating that the true case count (red trace) was 5 times higher than reported cases (data: open circles). Epidemic growth included the slow initial growth typical of network transmission, followed by a period of exponential, then linear growth, and ending with a period of logistic growth as vaccine deployment began to suppress the outbreak. The cumulative count for S is also shown (brown dashed trace).

According to the selected model, monkeypox was first reported from a node of 625 susceptible individuals. It was estimated that 105000 cases of monkeypox were present in the US by September 6, 2022, with only 19851 of these cases reported in data (12). This has consequences for the planned subsequent deployment of vaccine.

A predictive model (Figure 4) was formulated using the first 75 days of data. This model was then used to predict the observed counts for the remainder of a 110-day period (Figure 8).

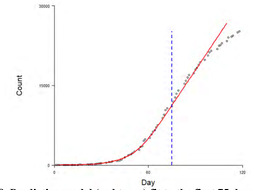

Figure 8. Predictive model (red trace) fit to the first 75 days (blue line) of data.

The data provided an exceptionally good fit to the predictive model for observations across the 75-day fitting period. The long period of slow growth (days 1:40) preceding exponential growth is characteristic of network dynamics. Predicted values following day 75 (red trace) follow the data until they begin to diverge around day 100, likely due to a decrease in daily incidence following deployment of vaccines.

Correlation between the predicted and observed counts across each time-lag from 75 through 110 days was calculated and visualized in a cross-correlogram (Figure 9).

Figure 9. Cross correlogram for predicted and observed counts, Days 75:103.

The predictive model demonstrated a very high level of reliability for 28 days following the last day of model fitting.

DISCUSSION

EDDIS development began as an effort to use imperfect data to estimate R_0, so that data-assimilation could be used to augment SEIR, and as a method for estimating the true case count when cases are underreported. EDDIS was then extended to provide temporal-stepwise iteration for the parameterization of predictive models. In these respects, EDDIS development has been fully successful. Beginning with a network SEIR model for CoVid-19, EDDIS has been generalized to analyze and predict monkeypox outbreak data.

The combination of process (SEIR in network) with observation (detection) modeling is called a hierarchical model (*7*). Data assimilation into the hierarchical model was through Newton's Binomial Theorem (8), where model output and data were combined to estimate parameter values and to provide maximum-likelihood.

EDDIS can be further expanded to accommodate surveillance data for a wider variety of disease agents. Future extensions will include increased automation, reduced reliance on expert operation, the addition of a broader variety of parameter estimates, and the addition of new test data such as surveillance records from the 2014 West African Ebolavirus Epidemic. EDDIS continues to develop into a useful tool for enhanced situational awareness, intervention strategy, and decision-making.

REFERENCES

1. Anderson, R. and R. May, Infectious Diseases of Humans, Dynamics and Control, Oxford University Press, New York, **1991**.
2. Keeling, M.J., and P. Rohani. Modeling infectious diseases in humans and animals. Princeton University Press, Princeton. **2008**.
3. Keeling, M.J., and Eames, K.T.D., Networks and epidemic models. J. R. Soc. Interface **2005**, 2, 295–307.
4. Ingersoll, T. E., Estimators for Disease Dynamics with Imperfect Surveillance. DEVCOM CBC-TR1747. **2021. (Approved for public release, distribution unlimited)**.
5. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. **2019**. https://www.R-project.org/
6. Soetaert, K., Petzoldt, T. and Woodrow, R. *Solving Differential Equations in R: Package deSolve*. Journal of Statistical Software **2010**, 33 9 1-25. http://www.jstatsoft.org/v33/i09
7. Royle, J. A., and R. M. Dorazio. Hierarchical Modeling and Inference in Ecology. Elsevier, London. **2008**
8. Kline, M. *History of Mathematical Thought*. Oxford University Press. **1972**.
9. Forbes, E.G. *Gauss and the Discovery of Ceres*. Journal for the History of Astronomy. **1971,** 2 (3): 195–199.
10. Center for Systems Science and Engineering (CSSE). *COVID-19 Dashboard*. Johns Hopkins University (JHU). **2021.** https://coronavirus.jhu.edu/map.html
11. CDC, *Public Health Responses to COVID-19 Outbreaks on Cruise Ships — Worldwide, February–March 2020*, CDC Weekly, **2020**, 69(12);347-352
12. Global.health Monkeypox. https://github.com/globaldothealth/monkeypox (accessed on 2022-09-28). **2022**.
13. Kaler, J., A. Hussain, G. Flores, S. Kheiri, and D. Desrosiers. Monkeypox: A comprehensive review of transmission, pathogenisis, and manifestation. Cureus 14(7). **2022.**
14. Reynolds, H. T., T. E. Ingersoll, H. A. Barton. Modeling the environmental growth of Psuedogymnoascus destructrans and its impact on the white-nose syndrome epidemic. Journal of Wildlife Diseases **51**: 318-331.**2015.**

DEVCOM CBC @ DTRA CBD S&T Conference
Scan the QR Code to view all of CBC's
2022 DTRA CBD S&T Conference materials
https://cbc.devcom.army.mil/cbdst-conference/

Contact the author, email: thomas.e.ingersoll.civ@army.mil

Acknowledgements: R&D funding for EDDIS was provided by DTRA (Dr. Sweta Batni) and by the Director, Combat Capabilities Development Command Chemical Biological Center (Dr. Eric Moore) under the authorities and provisions of Section 2363 of the FY 2018 NDAA to develop new technologies, engineer innovations, and introduce game-changing capabilities. Special thanks to Michael Kierzewski, and the staff of the Modeling Simulation and Analysis Branch, Combat Capabilities Development Command Chemical Biological Center. The views expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of the Department of Defense or the U.S. Government.)