SI Text

Generating function approach to link characteristic times and transmission rates

The generalized SEIRD model includes n_e number of exposed classes each of duration T_e/n_e. Given a value of $r_0 = \tau_c^{-1}$ and ρ_D, then the generating function approach of Wallinga and Lipsitch [1, 2] can be used to derive the following relationships:

$$\mathcal{R}_0 = \frac{1 + T_i r_0}{\left(\frac{n_e}{T_e n_e + r_0}\right)^{n_e} \left(1 - \rho_D + \frac{\rho_D}{1 + T_d r_0}\right)},$$

$$\beta_I = \frac{(1 + T_d r_0)(1 + T_i r_0)(\rho_D - 1)}{\left(\frac{n_e}{n_e + T_e r_0}\right)^{n_e} T_i (T_d r_0 (\rho_D - 1) - 1)},$$

$$\beta_D = \frac{(1 + T_d r_0)(1 + T_i r_0)(\rho_D)}{\left(\frac{n_e}{n_e + T_e r_0}\right)^{n_e} f T_d (T_d r_0 (\rho_D - 1) - 1)}.$$

Stochastic simulations of epidemic outbreaks

Stochastic realizations of the SEIRD model are simulated using the Gillespie framework [3], given the the following types of "reaction" events, [Pre-death infection] $S + I \to E + I$, at a rate $r_1 = \beta_I S \frac{I}{N}$, [Post-death infection] $S + D \to E + D$, at a rate $r_2 = \beta_D I \frac{S}{N}$, [Onset of infectiousness] $E \to I$, at a rate $r_3 = \sigma = 1/T_E$, [End of infectiousness] $I \to R$ with probability $1 - f$ (i.e., survival) and $I \to D$ with probability f (i.e., death), at a rate $r_4 = \gamma = 1/T_I$, [Burial] $D \to 0$, at a rate $r_5 = \rho = 1/T_D$ where $N = S + E + I + R$. Epidemics are initiated with one infectious individual in an otherwise susceptible population. Mathematically, the initial state is $\mathbf{y} = (N_0 - 1, 0, 1, 0, 0)$ at $t = 0$. The total rate of outbreak-associated events is $r_{tot} = \sum_{i=1}^{6} r_i$. The time until next event is determined randomly such that $\delta t \sim \frac{-\log \chi}{r_{tot}}$ where χ is a uniformly distributed number between 0 and 1. In this way, the time between events follows an exponential distribution with rate r_{tot}. Then, the probability of each event is r_i/r_{tot}. After selecting an event and updating the discrete number of individuals, the reaction rates are recalculated and the process continues. The same framework can be extended to include multiple n subclasses within the exposed class, to capture the peaked nature of the exposed period (centered around 11 days for EVD). Trajectories are complete when the epidemic dies out because there are no more infectious individuals. In the present context, we are interested in those trajectories that do not die out before the end of the simulation time.

Figure S1. The distribution of the characteristic time, τ_c increases with decreasing trigger population. Ensembles of SEIRD stochastic dynamics are simulated with parameters as in Fig. 2. The ensemble includes 10^4 trajectories with an epidemic that persists for 300 days. In each case, estimates of τ_c are based on fits to the CCC for 42 days after the trigger population is reached.

Pseudo-Bayesian approach for uncertainty in ρ_D

The fraction of infections of EVD in west Africa due to transmissions from the deceased has been estimated to be between .01-0.3 [4]. However an estimate from a prior EVD outbreak in Congo is even higher [5]. To address this treat ρ_D as uncertain with a uniform prior distribution between 0.1-0.4. We are interested with models that have a deterministic $\mathcal{R}_0$. For each $\mathcal{R}_0$, we randomly, uniformly sample ρ_D from the prior distribution. Each sample of $\mathcal{R}_0$ and ρ_D determine the growth rate $r_0 = \tau_c^{-1}$ which in turn determines the infection parameters, β_I and β_D.

An ensemble of 10^4 trajectories are obtained for each value $\mathcal{R}_0$ with ρ_D sampled uniformly from [.1 .4] for each trajectory. The resulting distribution of measured characteristic times $\hat{\tau}_c$ can be interpreted as a marginal distribution across ρ_D. Even with a fixed $\mathcal{R}_0$, as ρ_D varies the secondary infection distribution changes. Hence, the marginal distribution across ρ_D can also be interpreted as a marginal distribution across the time to secondary infection distributions that all correspond to the same reproductive number of the disease in the population τ_c.

Comparing results from CCC and ICC data

There are a number of potential pitfalls of using CCCs rather than ICCs. Yet, King and colleagues in their critique of CCCs noted that the summary statistics of the epidemic growth for SEIR models were functionally equivalent when fitted to an ensemble of CCCs and ICCs given the same underlying disease parameters [6]. The difference, as they pointed out, was how to interpret the *quality of the fits* in inferring CIs. Similarly, here we find that the resulting distributions, $p(\hat{\tau}_c|\tau_c)$, measured using either the CCC or ICC are similar but not equivalent (Fig. S2). Differences in the resulting CIs are minor, but the median of the ICC distribution is skewed larger than the theoretical value. These differences scale up to the overall analysis, but with minor effect on the τ_c and $\mathcal{R}_0$ CIs. The resulting τ_c CIs from our synthetic data are 16.5–28.0 given fits to cumulative data and 16.5–28.3 given fits to incident data. The resulting CIs are approximately the same size. This contrasts with prior results from profile likelihoods in which the CIs as inferred from CCCs are contained within the CIs as inferred from ICCs [6].

Figure S2. Variation in regression-based fits of the characteristic time of case counts between estimates using cumulative or incidence case count data. The theoretical characteristic time of the underlying model is $\tau_c = 20$ as shown by the black dashed vertical line. The ICC distribution has a median of 20.3 days with 95% CIs of 16.5-28.3. The CCC distribution has a median of 19.9 days with 95% CIs of 16.5-28.0.

We emphasize that our method does not utilize the quality of any individual fit to generate the CIs for $\mathcal{R}_0$, for precisely the reasons cautioned by King and colleagues [6]. Estimating the growth rate, r_0, from linear regression is uncertain due to the fitting procedure itself. Associated with each estimate of r_0 are confidence intervals. We compare the distributions of errors due to fitting associated with the CCCs and ICCs of our simulated data in Fig. S3. The errors associated with ICCs are larger than those associated with CCCs. It remains an open question as to whether/when CCCs rather than ICCs are preferable when leveraging regression fitting to estimate r_0 given process noise alone, rather than observation noise.

Figure S3. Distributions of standard error of estimated growth rates, $r_0 = \hat{\tau}_c^{-1}$ for CCCs and ICCs arising from linearly fitting simulated data assuming deviations are Poisson distributed. The underlying deterministic model for simulated data has a theoretical characteristic time of $\tau_c = 20$ days.

Identifiability problem persists when inferring relative fraction of post-death transmission from stochastic trajectories

The measured growth rate of an epidemic, $\hat{r}_0 = \hat{\tau}_c^{-1}$ can be used to infer the basic reproductive number, $\mathcal{R}_0$. Using a generating function approach, it can be shown that $\mathcal{R}_0 = 1/M(-r_0)$ where $M(z) = \int_0^\infty e^{za} g(a) \mathrm{d}a$ [1] where $g(a)$ is the normalized fraction of all secondary cases caused by an infectious individual at "age" a since infection. A range of values of $\mathcal{R}_0$ may be compatible with a single estimate of $\hat{r}_0$ [2]. This uncertainty is a consequence of an identifiability problem given uncertainty in the relative fraction of transmission events that could be attributed to post-death transmission. Here we ask: what is the variation in the growth rate, r_0, and characteristic time, $1/r_0$, compatible with varying fraction of post-death transmission, $0 < \rho_D < 1$. Fig. S4 shows the measured variation in the characteristic time, $\langle 1/\hat{r}_0 \rangle$ for 5000 ensembles for three different expected growth rates $r_0 = 1/14$, $1/21$ and $1/28$ days^{-1}. In each case, we varied ρ_D from 0 to 1 in increments of 0.2. As in the prior section, we find that the characteristic time of an epidemic can vary substantially for a fixed value of r_0. Here, we also expect that a range of mechanistic models can all yield the same expected characteristic time. As is evident, the identifiability problem highlighted in prior analyses of deterministic models [2, 7] also applies in the case of stochastic models. For SEIRD models, the expected value of the basic reproductive number increases with ρ_D. Therefore, efforts to constrain estimates of $\mathcal{R}_0$ from EVD case data will be subject to inherent variability due, in part, to uncertainty in mechanism, e.g., the relative fraction of post-death transmission, and process, i.e., stochastic outbreak dynamics.

Figure S4. Identifiability problem persists when fitting SEIRD models to stochastic data. The three scenarios correspond to cases where the characteristic time, $1/r_0$=14, 21 and 28 days. The realized epidemic growth rates of stochastic trajectories are measured given variation in ρ_D from 0 to 1 in increments of 0.1. Circles denote the median characteristic time while triangles denote the 95% confidence intervals from an ensemble of 10^3 simulations per condition.

References

1. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B: Biological Sciences. 2007;274(1609):599–604.

2. Weitz JS, Dushoff J. Modeling Post-death Transmission of Ebola: Challenges for Inference and Opportunities for Control. Scientific reports. 2015;5.

3. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;82(25):2340–2361.

4. WHO Ebola Response Team. Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections. New England Journal of Medicine. 2014;371(16):1481–1495.

5. Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A. Understanding the dynamics of Ebola epidemics. Epidemiology and infection. 2007;135(04):610–621.

6. King AA, Domenech de Cellès M, Magpantay FMG, Rohani P. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. Proceedings of the Royal Society of London B: Biological Sciences. 2015;282(1806).

7. Eisenberg MC, Eisenberg JNS, D'Silve JP, Wells EV, Cherng S, Kao YH, et al. Forecasting and Uncertainty in Modeling the 2014-2015 Ebola Epidemic in West Africa. http://arxivorg/pdf/150105555v3pdf. 2015;.