Journal of Theoretical Biology 486 (2020) 110079

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtb

Identification of the relative timing of infectiousness and symptom onset for outbreak control

Robert C. Cope, Joshua V. Ross*

The University of Adelaide, Stochastic Modelling & Operations Research Group, School of Mathematical Sciences, Adelaide, SA 5005, Australia

ARTICLE INFO

Article history:
Received 23 April 2019
Revised 8 November 2019
Accepted 11 November 2019
Available online 14 November 2019

Keywords:
Bayesian model discrimination
Epidemiology
Optimal experimental design
Random forests.

ABSTRACT

In an outbreak of an emerging disease the epidemiological characteristics of the pathogen may be largely unknown. A key determinant of ability to control the outbreak is the relative timing of infectiousness and symptom onset. We provide a method for identifying this relationship with high accuracy based on data from simulated household-stratified symptom-onset data. Further, this can be achieved with observations taken on only a few specific days, chosen optimally, within each household. The information provided by this method may inform decision making processes for outbreak response. An accurate and computationally-efficient heuristic for determining the optimal surveillance scheme is introduced. This heuristic provides a novel approach to optimal design for Bayesian model discrimination.

1. Introduction

The timing of infectiousness relative to symptom onset has been identified as a key factor in ability to control an outbreak (Fraser et al., 2004). The explanation is intuitive: if symptoms appear before infectiousness, then contact tracing and isolation strategies will be effective, whereas for post-infectiousness symptom presentation, broader, non-symptom based strategies must be adopted. Consequently, identifying the relative timing as early as possible in an outbreak is imperative to assessing potential for control and selecting a measured response.

Severe acute respiratory syndrome (SARS) is a prime example of a disease in which symptoms foreshadow significant levels of infectiousness (Anderson et al., 2004). This played a critical role in limiting mortality and morbidity in outbreaks during 2003, via simple public health measures such as isolation and quarantining (Ksiazek et al., 2003; Lee et al., 2003; Fraser et al., 2004; Anderson et al., 2004; Hsieh et al., 2005; Day et al., 2006). Smallpox is most similar to SARS in this respect, but must be contrasted with HIV, where a large proportion of secondary infections occur before symptoms (Fraser et al., 2004). For influenza, the relationship is less clear, with symptoms and infectiousness likely coinciding closely, with some transmission possible before symptom onset (Patrozou and Mermel, 2009; Lau et al., 2010). For established diseases, experimental evidence (Charleston et al., 2011) or large-scale detailed case information (International Ebola Response Team et al., 2016) can provide insight into the relative timing of symptom onset and infectiousness; however, this relationship will not be known in an outbreak of an emerging pathogen. Therefore, one must turn to early outbreak surveillance data for insights.

Many jurisdictions organize their emerging disease monitoring policies around households. As an example, First Few Hundred studies are proposed as a first response surveillance scheme following the identification of a novel disease and/or strain as part of national pandemic plans (McLean et al., 2010; van Gageldonk-Lafeber et al., 2012; AHMPPI, 2014). Following the observation of a first symptomatic individual, their household is enrolled in an intensive surveillance program, so that day of symptom onset for subsequent cases within that household are recorded. Studies of this form were developed for pandemic influenza in 2009 in both the United Kingdom (McLean et al., 2010) and the Netherlands (van Gageldonk-Lafeber et al., 2012); and have been instituted in response to a lack of methods for determining disease epidemiology as required for determining a proportionate response to novel outbreaks. In the Australian Health Management Plan for Pandemic Influenza (AHMPPI), First Few Hundred studies are proposed to be implemented following the first case of a novel influenza strain, with households being tracked nationally (but managed at the state/territory level) (AHMPPI, 2014). Methods have recently been developed to characterise transmissibility and severity of a novel pathogen – other factors influencing ability to control an outbreak (Fraser et al., 2004) – based on such data (Black et al., 2017; Walker et al., 2017). Currently lacking is a method for accu-

* Corresponding author.
E-mail address: joshua.ross@adelaide.edu.au (J.V. Ross).

https://doi.org/10.1016/j.jtbi.2019.110079

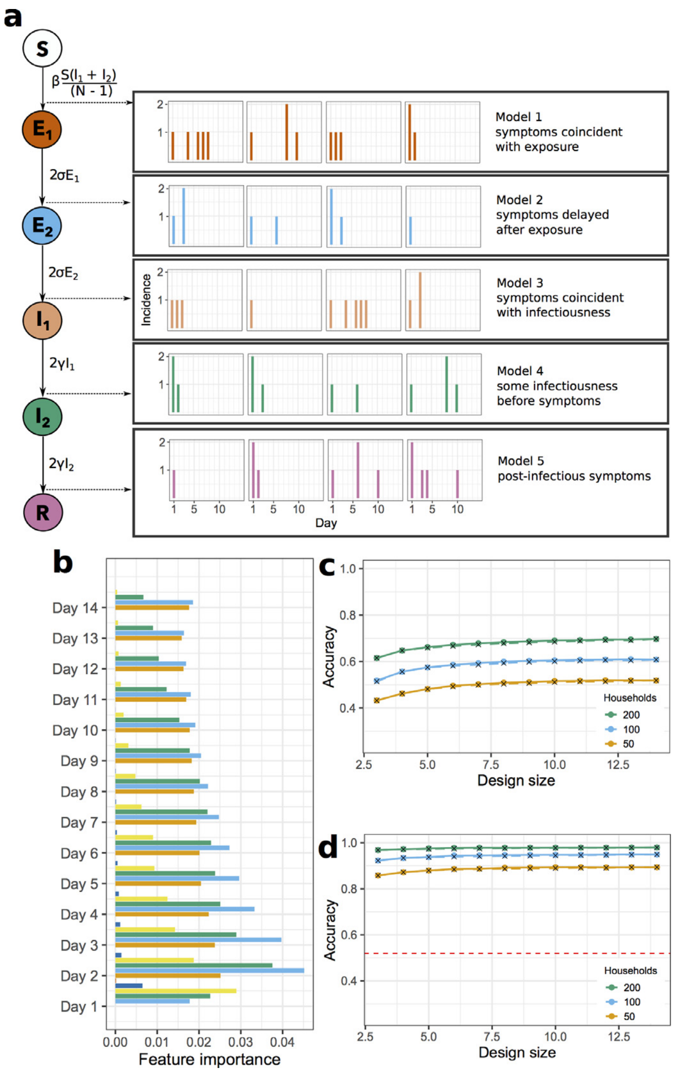

Fig. 1. (a) Model schematic describing: transitions between states within each household continuous-time Markov chain; the five observation models being discriminated between; and, the way that these household-level data are observed. The data observed in each model are the number of observations of the relevant transition each day, within each household: data from four illustrative sample households are shown here. (b) Random forest feature importance for the full 14-day design, used to construct the heuristic for smaller designs. Each bar represents a feature, so within each day there are (in this case, for households of size 5) 6 features, corresponding to the proportion of households with each incidence count, each day. (c) Resulting random forest accuracy (proportion of test simulations assigned to the correct model) as design size increases, for the true optimal design (solid lines) and heuristic solution (crosses with dashed line). In this case, random assignment would produce accuracy 0.2. (d) Two-class accuracy of random forest model discrimination: this measures the accuracy of discrimination between models with symptoms before or coincident with infectiousness, versus models with symptoms beginning after infectiousness. In this case, random assignment would produce accuracy 0.52 (red dashed line). These results correspond to households of size 5, with 10,000 training samples from each model, each with parameters drawn from the distributions displayed in Supplemental Figure S1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rate determination of relative timing of infectiousness and symptom onset using this data.

Here we introduce, and demonstrate through a simulation study, a method for identifying with high accuracy the timing of infectiousness relative to symptom onset from household-stratified symptom surveillance data (generated via simulation). Remarkably, we show this is achievable with observations taken on only a few specific days, chosen optimally, within each household. Our approach to determining the optimal surveillance scheme is based on an efficient heuristic. This heuristic provides a general, computationally-efficient approach to optimal design for Bayesian model discrimination.

2. Bayesian model discrimination for outbreak control

We model disease dynamics within each household as a continuous-time Markov chain (Keeling and Ross, 2008), that counts the number of household members that are susceptible (S), exposed (E_1 and E_2), infectious (I_1 and I_2), or recovered (and immune; R). Two compartments for exposed and infectious individuals allows for a broad range outbreak observation dynamics, and allows Erlang-distributed exposed and infectious periods. Under this model, the timing of symptom onset relative to infectiousness is mapped to which transition is observed: symptoms appear either upon infection, between infection and infectiousness, coincident with infectiousness, between infectiousness and recovery, or upon recovery. The challenge is to determine which of these five (observation) models best describes the household-stratified symptom-onset data (Fig. 1a).

There is a relatively rich literature on Bayesian model discrimination (Chopin et al., 2013; Drovandi and Cutchan, 2016; Alzahrani et al., 2018; Touloupou et al., 2018), and optimal design for such (Chaloner and Verdinelli, 1995; Ryan et al., 2015), which are the most appropriate tools and framework to address this question. A general difficulty with this theory is that practical implementation is at best difficult, and often infeasible. This has led to methods based on approximate Bayesian computation (ABC), which requires only simulation of realisations from each model, and is computationally feasible for a wide range of models. Unfortunately, there exists ‘a fundamental difficulty’ in establishing robust methods based upon summary statistics (Robert et al., 2011; Robert, 2016); however, see the recent work of Dehideniya et al. (Dehideniya et al., 2018).

Another approach to model discrimination in an ABC framework has been proposed by Pudlo et al. (Pudlo et al., 2015). They treat model discrimination as a classification problem, for which machine learning methods are ideal, and in particular propose the use of random forests to perform this task. This approach provides a highly-efficient, and importantly, robust method for model discrimination. Hainy et al. (Hainy et al., 2018) expand on this approach as specifically applied to optimal design for model discrimination.

We apply random-forest based Bayesian model discrimination, first for accurate, robust characterisation of relative timing of symptoms and infectiousness, and second, for optimal design of early outbreak surveillance for accurate model discrimination. Specifically, the aim of the latter is to select an optimal surveillance scheme, consisting of a fixed number of observations, in order to discriminate five different timings of symptom onset relative to infectiousness, within a household-stratified epidemic model. We evaluate, using simulated data, the impact of assumptions and summary statistics. Additionally, we propose a new, computationally-efficient and highly-accurate heuristic for optimal design choice, which in this application determines the optimal days upon which to perform surveillance in households.

3. Methods

3.1. Epidemic model

We demonstrate using an example system of a novel infectious disease, spreading in a population structured into households. We assume that the population is large and mixing between households is random, such that after a household is initially infected,

Table 1
Events, transitions and rates within a household. N is the (fixed) household size, β, γ and σ are the rates of infection, gaining infectiousness and recovery, respectively.

Description	Transition	Rate
Infection	$(S, E_1) \rightarrow (S-1, E_1+1)$	$\beta S(I_1+I_2)/(N-1)$
	$(E_1, E_2) \rightarrow (E_1-1, E_2+1)$	$2\sigma E_1$
Infectiousness	$(E_2, I_1) \rightarrow (E_2-1, I_1+1)$	$2\sigma E_2$
	$(I_1, I_2) \rightarrow (I_1-1, I_2+1)$	$2\gamma I_1$
Recovery	$(I_2, R) \rightarrow (I_2-1, R+1)$	$2\gamma I_2$

the remaining transmission within the household is independent of transmission outside the household (Ross et al., 2010; Black et al., 2013). Therefore, transmission dynamics within households can be modelled independently (Black et al., 2017), i.e., with infection only occurring between individuals within a household, rather than between households. Note that this is an assumption which simplifies the simulation process, but it could be modified if necessary. Given this novel etiological agent, we wish to determine if symptom onset occurs at the time of infection, between infection and infectiousness, coincident with infectiousness, after infectiousness, or coincident with recovery (i.e., these are the five candidate models we wish to discriminate). The model behaviours are otherwise assumed identical. To be emphatic, the underlying disease dynamics is identical in all five models, each differing only in when observations are made, corresponding to different timings of symptom onset (Fig. 1a). We focus on selecting between these observation models as the relative timing of infectiousness and symptom onset is critical to effective outbreak management: quarantine can be applied effectively if symptoms occur before (or possibly coincident to) infectiousness.

We model the epidemic dynamics in households as a continuous-time Markov chain (Figure 1a) (Keeling and Ross, 2008). Individuals transition from susceptible (S) to exposed (E_1, and subsequently E_2), then to infectious (I_1, and subsequently I_2), and finally to recovered (R), with rates as described in Table 1. The model dynamics are general, but explicitly resemble the dynamics of a respiratory virus such as influenza, as potential future pandemic influenza is of substantial concern globally. The collection of First Few Hundred data is included in the Australian Health Management Plan for Pandemic Influenza (AHMPPI, 2014), along with similar pandemic preparedness plans in other jurisdictions, so demonstrating the ability to discriminate models using these data for diseases resembling influenza is highly relevant. As such, prior parameter choices and the overall duration of the observation process (i.e., 14 days) also reflect influenza dynamics.

We assign a prior distribution to each parameter (Supplemental Figure S1), based on physical quantities to reflect the assumed prior knowledge of the etiological agent:

- $\frac{1}{\sigma} \sim \text{Gamma}(6, 1/2)$, representing a mean exposed duration of 3 days (mode at approximately 2.5 days);
- $\frac{1}{\gamma} \sim \text{Gamma}(6, 1/2)$, representing a mean infectious duration of 3 days (mode at approximately 2.5 days); and,
- $R_0 \sim 1 + \text{Gamma}(2, 1/2)$, representing a mean R_0 (the expected number of secondary cases caused by an infectious individual in a fully susceptible population) of 2 (mode at approximately 1.5).

These distributions are sampled per-simulation, i.e., sampled parameters are kept constant across all households within a given epidemic. We note that these priors are relatively broad, reflecting uncertainty around disease transmission dynamics, but within a range resembling the dynamics of a respiratory virus such as influenza. Prior distributions should be chosen to reflect what is known about the disease of interest.

Following the first symptomatic case in a household, the number of symptomatic cases within the household is observed daily (i.e., the unit of time considered is one day). The instant that the first individual in a household shows symptoms is time zero. Then, the number of cases seen before time 1 constitutes the first observation, between time 1 and 2 the next observation, and so on. This proceeds for 14 days, with any symptoms occurring after time 14 not observed. The 14 day duration allows time for the index case and subsequent infections to likely progress through the stages of infection given the transmission model and parameters chosen, resulting in most household transmission being observed within this time. If the disease progressed on a different timescale, the duration and frequency of observation should be varied appropriately, e.g., an infection with slower outbreak dynamics might be observed weekly rather than daily.

When testing the effect of asymptomatic infections on model discrimination, we sample an additional parameter, p_{obs}, the probability that an individual shows symptoms (implemented as an independent Bernoulli trial for each individual at the time of symptom onset). Note that p_{obs} is held constant within each simulation, i.e., it varies by outbreak, but not by infected individual. We explored two scenarios: (1) p_{obs} ~ Beta(5, 5) (i.e., a mean p_{obs} of 0.5), and (2) p_{obs} ~ Beta(7.5, 2.5) (i.e., a mean p_{obs} of 0.75). Figure S1 includes a visualisation of these distributions. We emphasise that in the asymptomatic infection scenario, data collection from a household begins with the first *observed* symptomatic case in that household; the index case may be asymptomatic or symptomatic.

In preliminary studies (not reported) we estimated the accuracy of model discrimination with fixed, known parameters – the resulting accuracy was higher than with parameters sampled from the prior distribution. However, we report only the results with parameters sampled from a prior distribution here, as in an ongoing outbreak exact parameter values are likely to be unknown.

3.2. Random forest model selection

To attempt to discriminate models, we use the approximate Bayesian random forest approach of Pudlo et al. (Pudlo et al., 2015). A random forest is a popular machine learning classifier, that operates by aggregating many classification trees, each constructed on a random subset of predictors and a bootstrap sample of the training data (Hastie et al., 2009). When making a prediction the classification from each tree is determined, with the label predicted by the highest number of trees being the prediction of the random forest. Implementations of the random forest algorithm are available in most commonly-used software packages.

The process of Bayesian model discrimination using random forests proceeds as follows:

- Select a number of simulations, N_s, and a number of households, N_h.
- For each model:
  - Sample a set of parameters $\theta = (R_0, \sigma, \gamma)$ from the (prior) distributions.
  - Simulate N_h households given these parameters.
  - Repeat this process N_s times.
- Given the N_s simulations from each model, extract the data corresponding to the considered design.
- Construct a random forest that predicts the model label, given the simulations.
- Assess the accuracy of the process on a left-out test set.

Infections within each household are simulated using a standard Doob-Gillespie algorithm for simulating continuous-time Markov chain dynamics.

Random forests were constructed using the Python scikit-learn RandomForestClassifier algorithm (Pedregosa et al., 2011), with 200 trees. Note that we use a completely separate set of test simulations to determine accuracy (of the same size as the training data), rather than out-of-bag error. Out-of-bag error is an error metric commonly used with random forests, that is calculated using the training data, rather than a separate test set. It relies on the structure of the random forest: in a random forest, only a (randomly sampled) proportion of training data (i.e., simulations) are used to construct each tree, so the remaining training data may be used to test the accuracy of that tree. Aggregating the result across all trees gives the out-of-bag error. In some cases out-of-bag error is prone to bias, so to ensure we are correctly assessing accuracy we instead use a left-out test set. We report accuracy as the proportion of all test samples that are correctly assigned to their generating model. I.e., we test 10,000 left-out training simulations from each model, and count those assigned the correct label. We also count the number that were assigned correctly to pre-infectious or coincident with infectious symptoms, versus symptom onset after infectiousness has begun: we call this the *two-class accuracy*. This was tested as it is the most relevant set of models to discriminate for determining the effectiveness of quarantine for disease control.

All code necessary to produce the simulated data and perform model discrimination will be made available publicly (upon publication).

To operationalise this process during an outbreak, the observed household data (on the days corresponding to the chosen design) would be input into a pre-trained random forest model, which would result in a predicted model label. That prediction indicates which observation model the outbreak most closely resembles.

3.3. Optimal sampling design

Conducting a First Few Hundred-style study can be extremely labour intensive. Consequently, we wish to assess the potential for model discrimination when sampling is only performed on a subset of days, rather than every day. If we choose to only sample on $D < 14$ days, within the first 14 days following the first symptomatic case in each household, we must necessarily also choose the optimal days on which to sample. We call the number of days D being sampled the design size. We choose those days that produce the highest classification accuracy on a left-out test set. This design problem is small, with only $\binom{14}{D}$ designs of size D (or $2^{14} = 16,384$ total designs) to evaluate, so we apply exhaustive search in this case; however a combinatorial optimisation algorithm could be applied and would likely be necessary in a more complex design problem to search for the optimal design.

Potentially symptom onset data could be made complete for this style of study by, for example, asking each household on which day all individuals with symptoms first presented with them (rather than just the individuals who presented symptoms on the sampling days); although some loss in quality of data might be expected. In other cases it might be necessary to perform a test (e.g., virological testing) as part of the sampling program, in which case choosing optimal designs that are as small as possible can save substantial resources. Our study provides an example of the model discrimination and optimal sampling design process, that could be generalised to reflect the appropriate sampling scheme where necessary.

3.4. Summary statistics

To more effectively use the household data in training the random forest, we summarize the raw household data to produce daily distributions of counts. That is, we count the proportion of households that, on day d, observed an incidence of i, and then use the resultant (design size) × (household size + 1) data vector as the new random forest predictors. For example, with designs of size 5, households of size 5, and 200 households, the raw data would consist of $5 \times 200 = 1000$ predictors, whereas the daily summaries would consist of $5 \times 6 = 30$ predictors.

3.5. Heuristic solution

Rather than evaluating the full set of possible designs, or applying an optimisation algorithm, we propose a heuristic for efficiently finding high-quality designs of a given size. This heuristic is to perform random forest model selection on the largest possible design, extract the random forest feature importance (Fig. 1b), and use this random forest feature importance to rank design points. Specifically, days are ranked on their maximum feature importance (i.e., decrease in Gini impurity, see below); the sum of the importance of features from a day was also tested, but had inferior performance. A design of size d uses the highest-ranked d design points. The random forest feature importance metric we use is the mean decrease in Gini impurity (Raileanu and Stoffel, 2004) of a feature across the trees in the random forest. The Gini impurity at a node is the probability that a new element at that node would be assigned an incorrect label, if it was assigned a random label from the distribution of training labels at that node. This metric is calculated using the python scikit-learn random forest algorithm (Pedregosa et al., 2011).

4. Results

Random forest-based Bayesian model discrimination was able to discriminate relative timing of symptoms and infectiousness for simulated household-stratified symptom-onset data. With 200 households of size 5, accuracy was 0.6974 for discriminating the five observation models (with random parameters, and 10,000 training simulations per model). When selecting solely between pre-infectiousness and coincident symptoms versus post-infectiousness symptoms, accuracy was 0.9796 (we call this the two-class accuracy); suggesting that most model discrimination error was between similar models to which the same management decisions might be applied. Accuracy was reduced with fewer households: to 0.608 with 100 households, and 0.518 with only 50 households (Fig. 1d); these had two-class accuracy of 0.95 and 0.894, respectively. These results were robust with respect to variation in household size (Figure S2), with accuracy ranging from 0.648 with 200 households of size 3 to 0.703 with 200 households of size 7. We report results for households of size 5 for the remainder of this section.

Remarkably, model discrimination remained accurate when only a small subset of daily household data were observed, when the observations were from an optimal design: a design of size 5 with 200 households was sufficient to produce a classification accuracy of 0.662 and a two-class accuracy of 0.975 (Figs. 1d and 2a), only marginally below the accuracy of the full design (Figure S3). Accuracy increased as the design size (i.e., number of days of surveillance) and the number of households increased. The heuristic produced an effectively indistinguishable level of accuracy compared to the optimal across design sizes, both for overall accuracy (Fig. 1c) and two-class accuracy (Fig. 1d). The heuristic ensured a substantial reduction in computation time: to produce Fig. 1c, 39

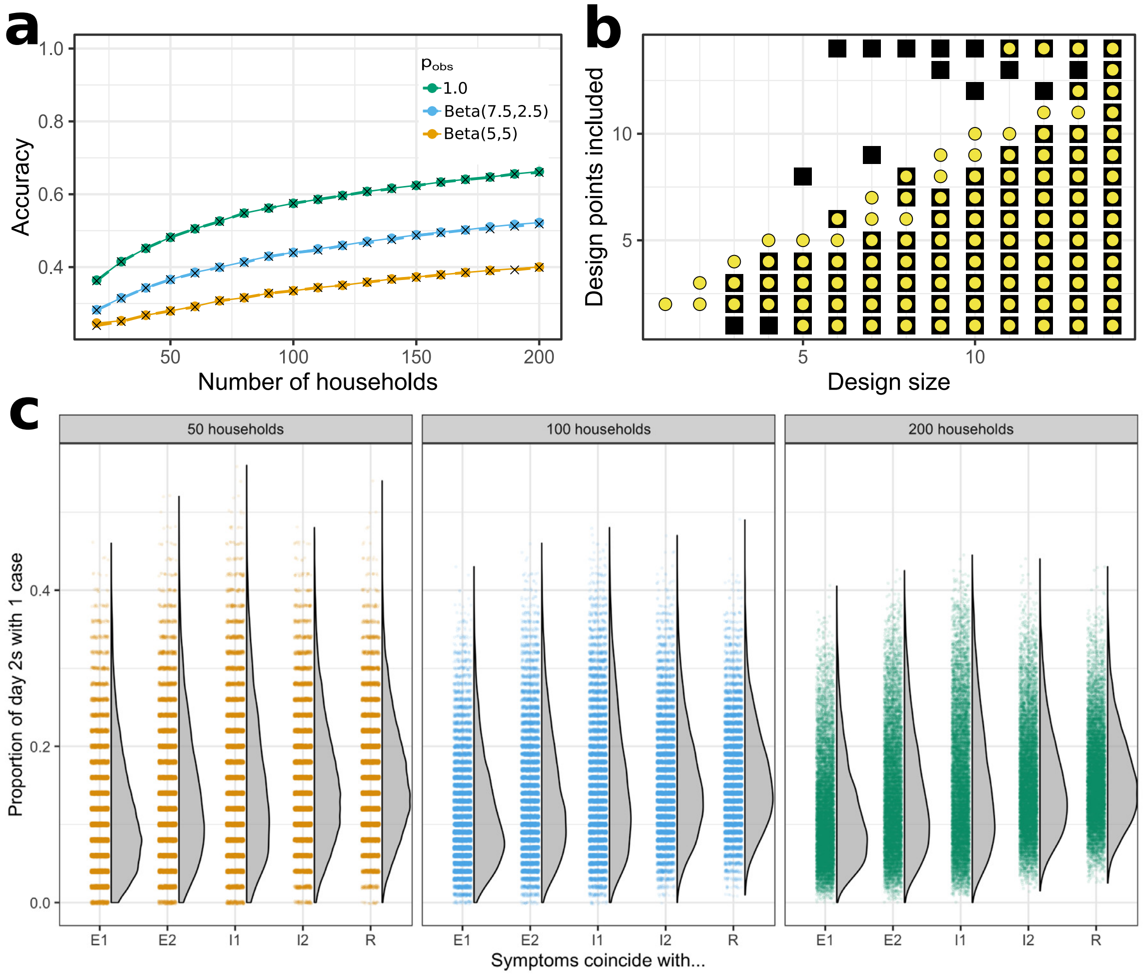

Fig. 2. (a) Accuracy of model discrimination in designs of size 5, as the number of households increases, and under partial observation. Note that p_{obs} is not a fixed parameter but is sampled from a distribution: The Beta(5,5) distribution has mean 0.5, and the Beta(7.5,2.5) distribution has mean 0.75. Figure S3 shows the equivalent result with a design of size 14. (b) Difference between heuristic designs (coloured points) and optimal designs (black boxes) as the design size increases. Note that we do not evaluate optimal designs of size 1 or 2, and so there are no optimal designs in these columns. (c) Distribution of training sample observations (under each model and number of households) for the most important feature under the heuristic: the proportion of households with 1 case observed on day 2. Each coloured point represents an observation in the training sample. These results correspond to households of size 5, with 10,000 training samples from each model, each with parameters drawn from the distributions that appear in Supplemental Figure S1.

random forests were required when using the heuristic, compared to 49,107 random forests to produce the optimal results.

The key design points (i.e., sampling days) for optimal designs were consistently the second day (Fig. 2b), followed by other days early in the outbreak (i.e., days 3–6, and day 1). Days 7–14 typically had little impact on model discrimination accuracy (i.e., optimal accuracy and two-class accuracy consistently levelled off as design size increased beyond 5; Fig. 1c/d), and the optimal combination of these days varied due to stochasticity in both training and test data. This is consistent with the feature importance used to develop the heuristic (Fig. 1b), i.e., those days that were consistently optimal were those with highest feature importance. When the most important design point is visualised (Fig. 2c) it shows a subtle but clear difference between distributions of observations from the different models; this provides intuition as to how decision trees constructed from many predictors of this form can accurately discriminate models.

To assess the impact of asymptomatic infections on model discrimination, we repeated the analysis, except with each individual only being symptomatic (at the point symptoms would otherwise appear) with probability p_{obs} (again, sampled from a prior distribution). This partial observation made model discrimination substantially more challenging: with designs of size 5 and 200 households (Fig. 2a), accuracy was 0.522 and two-class accuracy was 0.863 when p_{obs} ~ Beta(7.5, 2.5) (i.e., a mean of 0.75), and accuracy was 0.400 and two-class accuracy was 0.736 when p_{obs} ~ Beta(5, 5) (i.e., a mean of 0.5) (compared to 0.622 and 0.975 with complete observation).

5. Discussion

Identifying the relative timing of symptom onset and infectiousness in an emerging epidemic is critical to outbreak control. We have demonstrated, on simulated data, a method for identifying the relative timing based upon household-stratified data available early in an outbreak. This method produces reasonable accuracy for discriminating between five observation models, and very high accuracy for determining pre- or coincident with infectiousness symptom onset versus post-infectiousness symptom onset (i.e., two-class accuracy). This can be done without observing each household every day. Moreover, we can use random forest feature importance to inform a heuristic that vastly reduces the computation necessary to choose high-accuracy designs.

It is remarkable that it is possible to discriminate models so accurately, given that they share identical epidemic dynamics, and only differ in observation. The non-parametric nature of the random forest is able to use small but clear differences between models (e.g., Fig. 2c) to extract sufficient information to discriminate them. Combining the raw household data to form summary statistics is critical to this: if the raw household data is used rather than the summary statistics, accuracy is substantially lower. While it can be difficult to interpret the classifications made by a random forest-classifier, interrogating key individual predictors (as in Fig. 2c) provides clarity, and elucidates why feature importance provides a useful heuristic for choosing optimal designs (Molnar, 2019).

The accuracy of model discrimination decreases substantially as the proportion of cases that are asymptomatic increases. This can be compensated by increasing the number of households (Fig. 2a). The outbreaks in which early control is most critical are likely to be those in which most individuals are symptomatic, due to symptoms being strongly correlated with severity, for example hospitalisations and deaths. However, there also exists diseases for which outbreak control is critical, even when the proportion of symptomatic individuals is very low (e.g., poliovirus).

In some situations it may be necessary to consider more complicated surveillance schemes, in which case it may not be possible to evaluate the exact optimal design by exhaustive search. However, the heuristic proposed here should remain effective in more complicated design spaces, provided they have a similar form, i.e., designs of a given size are a subset of designs of larger sizes upon which the random forest can be trained to extract feature importance.

Assumptions impact any model-based study. Most critically, this model discrimination process assumes that the dynamics of the simulated epidemic model and observation models reflect the actual disease and observation dynamics. It is possible to use this method to select between models that differ in dynamics in addition to the observation process; however any increase in the number of models to classify will likely result in increased computation and potentially decreased accuracy. We have chosen to focus on the timing of symptom onset and infectiousness in one general disease process as an example, resembling influenza in both transmission dynamics and prior distributions on parameters. Useful future work could be to perform similar experiments on diverse disease processes, with different life histories. It would also be valuable to assess the robustness of the method for discriminating timing of symptom onset versus infectiousness when the underlying disease transmission model is misspecified. We note that selection between different transmission models (rather than observation models) for disease outbreaks has been considered in other studies, for example, assessing models of transmissibility over time for Norovirus from household data (Zelner et al., 2013).

In addition, the simulation study we present is a simplification of realistic disease dynamics. The model assumes homogeneous within-household mixing, Erlang-distributed latent and infectious durations, and constant transmission rates over the infectious period. Independence between households is also assumed (i.e., that once a household is infected, all subsequent infection events are due to transmission within that household); this is only potentially valid in the case of a large population of households and the early stages of an outbreak. Household size is uniform across households within the simulation; if household size were allowed to vary, data from each household size would need to be evaluated separately, and more households may need to be observed to obtain suitable accuracy. Assessing optimal design for model discrimination given a range of household sizes would be a valuable direction for future work.

Finally, we treat the interaction between symptom onset and infectiousness as a discrete process (i.e., symptom onset coincides exactly with transitions between states), whereas this process may be more general in practice. This paper demonstrates an example of the process of Bayesian model discrimination for outbreak control, and could be adapted to more complex disease models as desired.

In the future, the aim is to combine Bayesian model discrimination and parameter estimation in an online manner. Improving estimates of parameters improves the ability to discriminate models, and, more certainty regarding the model likely reduces variance in parameter estimates. This would allow for unified characterisation of all factors influencing the ability to control an outbreak.

Funding

R.C.C. and J.V.R. received funding from the Data To Decisions Cooperative Research Centre (D2D CRC). J.V.R. received funding from the Australian Research Council through the Future Fellowship scheme (FT130100254). J.V.R. received funding through the Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS). J.V.R. and R.C.C. received funding through the National Health and Medical Research Council (NHMRC) Centre of Research Excellence for Policy Relevant Infectious Disease Simulation and Mathematical Modelling (PRISM2). This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2019.110079.

References

Alzahrani, N., Neal, P., Spencer, S.E., McKinley, T.J., Touloupou, P., 2018. Model selection for time series of count data. Comput. Stat. Data Anal. 122, 33–44.

Anderson, R.M., Fraser, C., Ghani, A.C., Donnelly, C.A., Riley, S., Ferguson, N.M., Leung, G.M., Lam, T.H., Hedley, A.J., 2004. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 1091–1105.

Black, A., House, T., Keeling, M.J., Ross, J.V., 2013. Epidemiological consequences of household-based antiviral prophylaxis for pandemic influenza. J. R. Soc. Interf. 10, 20121019.

Black, A.J., Geard, N., Caw, J.M.M., Vernon, J.M., Ross, J.V., 2017. Characterising pandemic severity and transmissibility from data collected during first few hundred studies. Epidemics 19, 61–73.

Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: a review. Stat. Sci. 10, 273–304.

Charleston, B., Bankowski, B.M., Gubbins, S., Chase-Topping, M.E., Schley, D., Howey, R., Barnett, P.V., Gibson, D., Juleff, N.D., Woolhouse, M.E.J., 2011. Relationship between clinical signs and transmission of an infectious disease and the implications for control. Science 332 (6030), 726–729.

Chopin, N., Jacob, P.E., Papaspiliopoulos, O., 2013. SMC2: An efficient algorithm for sequential analysis of state space models. J. R. Stat. Soc.: Series B 75, 397–426.

Day, T., Park, A., Madras, N., Gumel, A., Wu, J., 2006. When is quarantine a useful control strategy for emerging infectious diseases? Am. J. Epidemiol. 163, 479–485. doi:10.1093/aje/kwj056.

Dehideniya, M.B., Drovandi, C.C., Gree, J.M.M., 2018. Optimal bayesian design for discriminating between models with intractable likelihoods in epidemiology. Comput. Stat. Data Anal. 124, 277–297. doi:10.1016/j.csda.2018.03.004.

Drovandi, C.C., Cutchan, R.A.M., 2016. Alive SMC2: Bayesian model selection for low-count time series models with intractable likelihoods. Biometrics 72, 344–353.

Fraser, C., Riley, S., Anderson, R.M., Ferguson, N.M., 2004. Factors that make an infectious disease outbreak controllable. Proc. Natl. Acad. Sci. U.S.A. 101 (16), 6146–6151.

Hainy, M., Price, D. J., Restif, O., Drovandi, C., 2018. Optimal bayesian design for model discrimination via classification. https://arxiv.org/abs/1809.05301.

Hastie, T., Tibsherani, R., Friedman, J., 2009. The Elements of Statistical Learning, second ed Springer series in statistics, New York.

Hsieh, Y.H., King, C.C., Chen, C.W., Ho, M.S., Lee, J.Y., Liu, F.C., Wu, Y.C., et al., 2005. Quarantine for SARS, taiwan. Emerging Infect. Dis. 11, 278–282.

International Ebola Response Team, Agua-Agum, J., Ariyarajah, A., Aylward, B., Bawo, L., Bilivogui, P., Blake, I.M., et al., 2016. Exposure patterns driving ebola transmission in west africa: a retrospective observational study. PLoS Med. 13 (11), e1002170.

Keeling, M.J., Ross, J.V., 2008. On methods for studying stochastic disease dynamics. J. R. Soc. Interf. 5, 171–181.

Ksiazek, T.G., Erdman, E., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al., 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1953–1966.

Lau, L.L., Cowling, B.J., Fang, V.J., Chan, K.H., Lau, E.H., et al., 2010. Viral shedding and clinical illness in naturally acquired influenza virus infections. J. Infect. Dis. 201, 1509–1516.

Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G.M., Ahuja, A., Yung, M.Y., Leung, C.B., To, K.F., et al., 2003. A major outbreak of severe acute respiratory syndrome in hong kong. N. Engl. J. Med. 348, 1986–1994.

McLean, E., Pebody, R.G., Campbell, C., Champerland, M., et al., 2010. Pandemic (h1n1) 2009 influenza in the UK: clinical and epidemiological findings from the first few hundred (FF100) cases. Epidemiol. Infect. 138, 1531–1541.

Molnar, C., 2019. Interpretable machine learning: a guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/.

Patrozou, E., Mermel, L.A., 2009. Does influenza transmission occur from asymptomatic infection or prior to symptom onset? Public Health Rep. 124, 193–196.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Pudlo, P., Marin, J.-M., Estoup, A., et al., 2015. Reliable ABC model choice via random forests. Bioinformatics 32, 859–866.

Raileanu, L.E., Stoffel, K., 2004. Theoretical comparison between the gini index and information gain criteria. Ann. Math. Artif. Intell. 41 (1), 77–93.

Robert, C.P., Cornuet, J.-M., Marin, J.-M., Pillai, N.S., 2011. Lack of confidence in approximate Bayesian computation model choice. Proc. Natl. Acad. Sci. 108, 15112–15117. doi:10.1073/pnas.1102900108.

Robert, C.P., 2016. Approximate Bayesian Computation: A Survey on Recent Results. In: Cools, R., Nuyens, D. (Eds.), Springer Proceedings in Mathematics & Statistics. Monte Carlo and Quasi-monte Carlo Methods, Vol. 163. Springer, Cham..

Ross, J.V., House, T., Keeling, M.J., 2010. Calculation of disease dynamics in a population of households. PLoS ONE 5, e9666.

Ryan, E.G., Drovandi, C.C., Gree, J.M.M., Pettitt, A.N., 2015. A review of modern computational algorithms for bayesian optimal design. Int. Stat. Rev. 84, 128–154.

Touloupou, P., Alzahrani, N., Neal, P., Spencer, S.E.F., Kinley, T.J.M., 2018. Efficient model comparison techniques for models requiring large scale data augmentation. Bayesian Anal. 13, 437–459.

van Gageldonk-Lafeber, A.B., van der Sande, M.A., Meijer, A., Friesema, I.H., Donker, G.A., Reimerink, J., et al., 2012. Utility of the first few100 approach during the 2009 influenza a(h1n1) pandemic in the netherlands. Antimicrob. Resist. Infect. Control. 1 (30).

Walker, J.N., Ross, J.V., Black, A.J., 2017. Inference of epidemiological parameters from household stratified data. PLoS ONE 12 (10), e0185910.

Zelner, J.L., Lopman, B.A., Hall, A.J., Ballesteros, S., Grenfell, B.T., 2013. Linking time–varying symptomatology and intensity of infectiousness to patterns of norovirus transmission. PLoS ONE 8 (7), e68413.

Australian Health Management Plan for Pandemic Influenza, 2014. http://www.health.gov.au/internet/main/publishing.nsf/content/519F9392797E2DDCCA257D47001B9948/24File/AHMPPI.pdf (accessed 22/02/19).