Real-time modelling of the 2014 Ebola outbreak in Liberia

Master in Vaccinology and Pharmaceutical Clinical Development

Siena University

Contents

Overview and Objectives:

The purpose of this practical is to provide you with some experience of fitting models to data in real-time. By the end you should:

- Understand the relationship between serial interval and R_0.
- Be able to fit models to emerging data.
- Understand some of the problems of real-time data analysis.
- Adapt a simple model to measure the impact of interventions *a posteriori*.

Part I: Estimating R_0 for Ebola

There are many ways of estimating R_0, such as by looking at the inter-epidemic period, or the proportion left uninfected after an epidemic, or by just looking at the epidemic curve in a totally susceptible population although this is not a commonly used technique, as we rarely observe infections in a completely susceptible population. However, since the population of Liberia never experienced an outbreak of Ebola before 2014, it is a reasonable assumption. Hence, we will attempt to use this approach to estimate the R_0 of Ebola.

We first consider the situation as of 24 August 2014, when the epidemic was rapidly spreading in Liberia. At this time, several modelling work estimated R_0 using the cumulative number of cases (1, 2), which was provided on WHO website. However, it is well known that using the cumulative number of cases can bias the estimate of R_0 and that, whenever it is possible, raw incidence time-series should be preferred (3). Hence, you will compare the estimate of R_0 using both time-series.

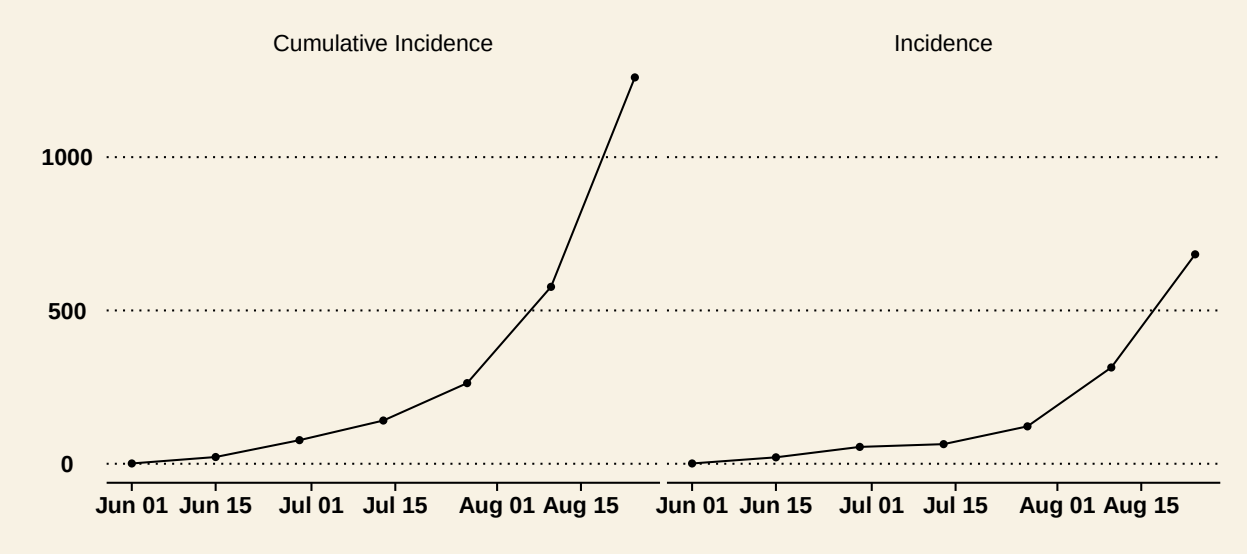

Figure 1: Cumulative and raw incidence time-series of confirmed and probable Ebola cases in Liberia up to 24 August 2014. Source: WHO.

Download and open up the spreadsheet *ebola_real_time.xls* and go to *Part I*.

The worksheet is organised as follows. We will assume that the serial interval (the time between successive cases, or generation time) is known from contact tracing and that it equals 14 days (see top left of worksheet). To make life easier, we assume that there is no variability in this, but you should be aware that this is an enormous (over)simplification.

At the very beginning of such an epidemic we can neglect to keep track of the depletion in susceptibles as the epidemic progresses, as the number of individuals in the population that are not susceptible will be negligible. Under these circumstances a single case will generate:

- R_0 cases after the first generation,
- R_0^2 cases in the 2nd generation
- R_0^3 cases in the 3rd generation, and so on.

That is, the number (I_t) of cases t generations after the initial case is given by:

$I_t = R_0^t$

And the cumulative number of cases observed up to generation t is simply:

Cumulative cases $= \sum_t I_t$

We can utilise this to estimate R_0 from the epidemic curve. In the example given we use least squares estimates to derive a best guess of R_0. That is, we choose a value of R_0, then compare the sum of the squared differences between the model and the data. We keep choosing values of R_0 until we can minimise the sum of the squared differences between the data and the model.

1) Change the values of R_0 and see what happens to the sum of the squared differences. Also observe how this affects the graph (which compares the model results with the data). Try and choose a value of R_0 that minimises the squared differences between the model and data.

Excel has an add-in program called Solver, which will run through lots of values of a cell and choose the one that fulfils certain criteria (either maximises a value in another cell, or minimises it, or sets it to zero). To

get Excel to do this we must first run Solver, tell it which cell it has to change, and which one it is trying to minimise (or maximise, or set to zero). In our case we want to minimise Cell H15 (sum of the squared differences) by changing Cell H17 (R_0).

If you never used Solver in Excel before you might have to load it first. You can find some instructions on the Solver website.

2) Run solver and estimate R_0 using alternatively the cumulative and raw incidence time-series. Which time-series leads to the highest estimate of R_0?

3) What do you think of the fit of the model to the data? In particular, pay attention to the part of the time-series that is best fitted by the model.

4) How do you expect your estimate of R_0 for Ebola to be dependent on your estimate of the serial interval? What would you expect it to be if it were:

   a) shorter (say 6 days)
   b) longer (say 24 days)

   Try and answer the above on paper. Then change the worksheet, and see whether this confirms your suspicions.

5) Use the model to forecast the number of cases up to January 2015. Were we right to be worried about Ebola?

6) Given that the population of Liberia is just above 4 million habitants, discuss what are the major limitations of this model?

Part II: Accounting for susceptible depletion

This second part of the exercise extends this simple model to try and estimate the basic reproduction number for Ebola in Liberia. Since we want to project how many cases occur over the course of the epidemic we will need to take account for the depletion of susceptibles (this is the major difference between these two models).

You should spend a few minutes working out how the model works. Again, we will be comparing the observed data with our model estimates, and we will be minimising the sum of squared residuals (Cell M12). Again, we will be using Solver to obtain the best fit (we will be trying to minimise the sum of the squared residuals, by changing the model parameters).

There are a number of parameters that we will be assuming that we know, and we will be estimating the others. We will assume that no-one is immune (almost everyone is susceptible), and (as mentioned above), that the serial interval is 14 days. We do not, however, know when the first case was introduced into this population (or indeed, whether more than one case was introduced). Thus we will have to estimate the initial number of cases, I_0 (Cell B5).

1) If $I_0 > 1$, then what does this mean?

The model, on the other hand, tracks infectious cases, irrespective of their disease status and consulting behaviour. Note that this is an important feature of transmission models, that often distinguishes them from models used in, say, health economics. In transmission dynamic models we are interested in infection, therefore we need to keep track of all infectious cases. As we do not know what proportion of Ebola cases will seek health-care and be reported in the data, we will have to estimate this as well (Cell B4).

We are interested in estimating R_0 (Cell B3).

There are 2 graphs on the right hand side of the worksheet, which show a comparison of the expected Ebola incidence, with the observed data. The first one just shows the time period for which we have data, the second graph projects the estimate over the remainder of the epidemic.

2) Spend some time changing the parameters (Cells B3:5) and seeing what effect it has on the epidemic curve.

   a. What effect does changing I_0 have, and why?
   b. What effect does changing R_0 have, and why?
   c. What effect does changing the % who seek health-care have, and why?

Estimating 3 parameters by changing them individually is close to impossible, so we will be employing Solver again, to try and get a reasonable estimate of these parameters.

3) Use Solver to estimate these 3 parameters, and look at the graphs comparing your model to the data, and the projections over time.

   a. What is your estimate of R_0? Does it differ from your estimates in Part I, and if so, then what is the explanation?
   b. When does the model predict that the epidemic peaks? How high is the peak?
   c. Does this surprise you (given what you know about the Ebola outbreak in Liberia), and if so, then what is the explanation?

4) How stable are your estimates? I.e. how sensitive are they to their initial values before using Solver?

   a. Change the initial values of Cells B3:5 and re-estimate the parameters.
   b. Keep a track of each of your best-fit estimates, by copying the value in Cell M12 and Cells B3:5 as well as the estimates of incidence (Column I) into the bottom part of the worksheet (there is the backbone of a table provided from Row 56). You will need to *Paste Values*, when you paste the values.
   c. After you have done this a few times (say, 10) plot the results, in terms of the projections. How, certain are you about the height and the timing of the peak?
   d. Which parameters are most important? Which parameters are more accurately estimated?

Mid September 2014, the CDC estimated that Ebola could affect up to 1.4 million people by January 2015 if it was not contained (1). In addition, the CDC estimated that only 40% of the Ebola cases were seeking health-care in Liberia at the time.

5) Use CDC estimates to fix the proportion who seek health-care and re-fit the model. Then, forecast the number of cases up to January 2015. Can you retrieve CDC's projection?

6) On the *Data* worksheet, you can find the observed incidence for the rest of the epidemic. Include data up to 19 October 2014 into the *Part II* worksheet and re-fit your model.

   a. What can you tell about the epidemic curve? How well does your model do now?
   b. What if you include another data-point (up to 2 November 2014)?

7) Relax the assumption that 40% of the cases were seeking health-care in Liberia and re-fit your model up to 2 November 2014. Can you get a better fit? Explain why and comment on the new parameter estimates.

8) Given what you know about the Ebola outbreak in West Africa, discuss what are the major limitations of this model? What does all this make you think about claims for *real-time* model projections?

Part III: Modelling the impact of interventions

By the end of October 2014, it became apparent that the Ebola epidemic was declining in Liberia. This period coincided with the scale-up of the international response:

- Construction of new treatment centres to isolate and treat more Ebola cases.
- Systematic contact tracing to detect new suspected cases as soon as possible.
- Safe burials of dead cases to prevent super-spreading events during traditional funeral ceremonies.

These control measures had a direct impact in reducing contacts between Ebola cases and the rest of the population. In addition, evidence from past Ebola outbreak suggests that affected communities can suddenly change their contact behaviour, thereby reducing the transmission of the virus (4). Although it is difficult to disentangle the role of the international response from change in community behaviour, it is believed that both types of interventions helped to control the epidemic.

In this part of the exercise, you will try to model the impact of interventions (either due to the international response or communities) and quantify their timing and efficacy by fitting the full incidence time-series.

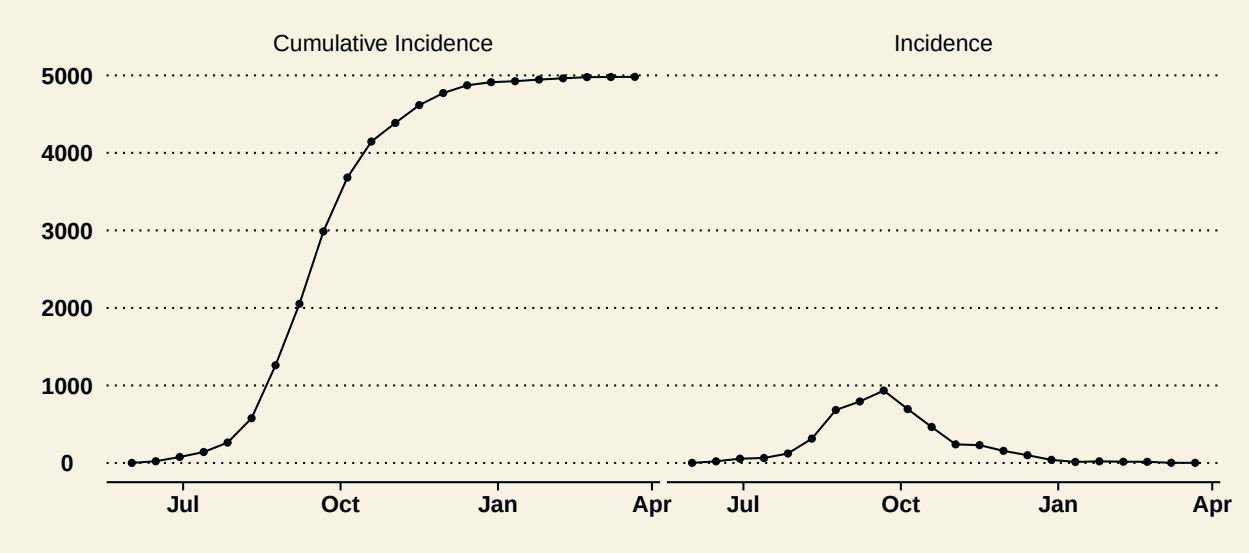

Figure 2: Cumulative and raw incidence time-series of confirmed and probable Ebola cases in Liberia up to 22 March 2014. Source: WHO.

In the *Part III* worksheet, we have included all the data points up to 22 March 2015. For simplicity, we are going to assume that interventions started at a single time point (Cell B6) and reduced the contact rate between Infectious and Susceptibles by a constant multiplicative factor (Cell B7) until the end of the epidemic. The proportion of cases seeking health-care should be fixed to 40%, as estimated by the CDC.

1) Adapt the model to include the impact of interventions on the contact rate:
    a. Which column of the worksheet should you edit to modify the contact rate?
    b. Edit the column by making use of the two new parameters. We suggest to use the `IF` function provided in Excel but other solutions are possible.
2) Spend some time changing the new parameters (Cells B6:7) and seeing what effect it has on the epidemic curve.
    a. What effect does changing the time of interventions have, and why?
    b. What effect does changing the efficacy of interventions have, and why?
    c. Can you think of a relationship between R_0 and the minimal efficacy required to bring the epidemic under-control? Try this value.
3) Use Solver to estimate the 4 remaining parameters. Proceed as in the question 4 of Part II to investigate the sensitivity of your estimates. **You should try different starting values and you might have to specify reasonable constraints on the parameters to help Solver find a good fit.**

Mid-September, USA president Barack Obama announced that 3,000 USA military personnel will be sent from October to Liberia to supply medical and logistical support to overwhelmed local health care systems and to boost the number of beds needed to isolate and treat victims of the epidemic.

4) Given the parameter estimates can you speculate on the impact of the USA intervention. You should also discuss the limitations of your model.

Bonus: Modelling of vaccination strategies

Modify the *Part III* worksheet to include the impact of vaccination in your model. For simplicity, you can assume that:

- The vaccine has 100% efficacy.
- One million doses of vaccine are available.
- All vaccines can be delivered in two weeks.

1) Without the interventions mentioned above, would one million doses have been enough to prevent the epidemic?

2) With the interventions mentioned above, investigate and discuss the additional benefit of the vaccine by considering different times of vaccination.

Conclusion

Mathematical modelling is a key tool to estimate parameters, make projections and evaluate alternative control strategies in infectious disease epidemiology. As such, many models were developed and used during the 2013-2015 Ebola outbreak in West Africa (5).

In the situation of an emerging disease the quality and quantity of publicly available data is usually poor and one should be cautious in over-interpreting model outputs because these are often based on strong assumptions. For instance, we saw that several parameters like the proportion of cases that seek health-care can not be estimated with accuracy during the exponential growth phase of an epidemic. As a rule of thumb, the more complex the model, the more data is required to estimate unknown parameters.

Complex models, accounting for hospitalization, funeral transmission, etc. were used during the initial phase of the Ebola epidemic in West Africa (1, 6). In addition to the lack of available data to calibrate these models, most of these did not include the effect of change in contact behaviours in affected communities (7). Overall, these models led to predictions comparable to those published by the CDC (1.4 million cases by January 2015).

This nightmare scenario did not happen in the end and, as it became clear that the epidemic was declining, modellers were eventually criticised for having overestimated the size of the Ebola outbreak (8). However, these predictions were based on the assumption that the transmission of Ebola would remain the same as in the summer of 2014, when the epidemic was growing out of control. By urging the scale-up of the international response, one can also argue that these "quick-and-dirty" modelling studies had a positive impact.

On the other hand, mathematical modelling of the Ebola outbreak at the sub-national level were also used to design the international response in real-time. For instance, the capacity of treatment centres in Freetown, Sierra Leone, was informed by a real-time modelling approach that enabled to forecast the bed demands a month ahead (9). These predictions were updated on a weekly basis and published online.

As the epidemic comes to an end and more data become available, complex model can be designed. For instance, a model that accounted explicitly for the effect of isolating Ebola cases was able to disentangle,

retrospectively, the impact of treatment centres from other interventions (community engagement, safe burial, etc.). This model demonstrated that a median of 56,600 Ebola cases were averted thanks to the scale-up of the international response and the increase of the number of beds in Sierra Leone (10). Had these beds been introduced one month earlier, the model predicted that 13,000 additional cases would have been averted.

References

1. Meltzer MI, et al. (2014) Estimating the future number of cases in the Ebola epidemic–Liberia and Sierra Leone, 2014-2015. *Morbidity and mortality weekly report Surveillance summaries (Washington, DC : 2002)* 63 Suppl 3:1–14.

2. Althaus CL (2014) Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLoS currents* 6.

3. King AA, Domenech de Celles M, Magpantay FMG, Rohani P (2015) Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences* 282(1806):20150347–20150347.

4. Camacho A, et al. (2014) Potential for large outbreaks of Ebola virus disease. *Epidemics* 9:70–78.

5. Lofgren E, et al. (2014) Opinion: Mathematical models: A key tool for outbreak response. *Proceedings of the National Academy of Sciences of the United States of America*:201421551.

6. Rivers CM, Lofgren ET, Marathe M, Eubank S, Lewis BL (2014) Modeling the impact of interventions on an epidemic of ebola in sierra leone and liberia. *PLoS currents* 6.

7. Funk S, Knight GM, Jansen VAA (2014) Ebola: the power of behaviour change. *Nature* 515(7528):492.

8. Butler D (2014) Models overestimate Ebola cases. *Nature* 515(7525):18.

9. Camacho A, et al. (2015) Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study. *PLoS currents* 7.

10. Kucharski AJ, et al. (2015) Measuring the impact of Ebola control measures in Sierra Leone. *Proceedings of the National Academy of Sciences of the United States of America*.