

**Estimation for the infection curves for the spread of Severe Acute Respiratory
Syndrome (SARS) from a back-calculation approach**

Ping Yan

Modelling and Projection Section

Centre for Infectious Disease Prevention and Control

Population and Public Health Branch

Health Canada

ABSTRACT

BACKGROUND

Available data about the spread of the Severe Acute Respiratory Syndrome (SARS) often come with two types: that by dates of report and that by dates of onset of symptoms. The latter is an approximation of the epidemic curve. Statistical methods can suggest the most plausible epidemic curve by the time of infection.

METHODS

Input data are SARS cases by dates of onset. Back-calculation methods are applied. The incubation distributions used in the algorithm are three log-logistic models: (i) that with median = 4.18 days and a shape parameter 3.413 (95% quantile 9.9 days); (ii) that with a shorter incubation distribution (median= 3.453 days; 95% quantile = 6.569 days); (iii) that with a longer incubation distribution (median = 5.139 days; 95% quantile = 17.2 days.)

RESULTS

Three infection curves based on the above three incubation distribution assumptions have been reconstructed separately for Singapore, Viet Nam, Hong Kong and Canada. Where available, these infections curves are compared with the trend based on dates of report, along with documented event history of public reactions.

CONCLUSIONS

The public knowledge (e.g. media) and actions(e.g. quarantine a large number of people) are often driven by reported outbreaks, subject to a delay of approximately two weeks since the time of infection. Second, it is important to be aware of the underline mechanisms that manifest the data that are publically available. The trends by dates of report are distorted not only by a time

delay, but also by reporting patterns in different jurisdictions. The trends by dates of onset have closer resemblance of the epidemic curves, but may be biased by reporting delay. Finally, since the beginning of the multiple country outbreaks SARS, some knowledge has been gained on the distributions of time from the onset of symptoms to either death or recovery, and the incubation distribution from the point of infection to the onset of symptoms. Jointly with a plausible reconstructed infection process, then one may be able to estimate key parameters such as infectivity.

INTRODUCTION

Data about the spread of the Severe Acute Respiratory Syndrome (SARS) often come with two types. One is that based on dates of report as represented by the far right compartment in Figure 1 (top). This is what the media and the public perceive and is the driving force for immediate action upon notification of an outbreak. The other is based on dates of onset of symptoms as represented by the far left compartment in Figure 1 (top). Reported data from Beijing (1) in Figure 1 (bottom) provide a good illustration of the reporting mechanism.

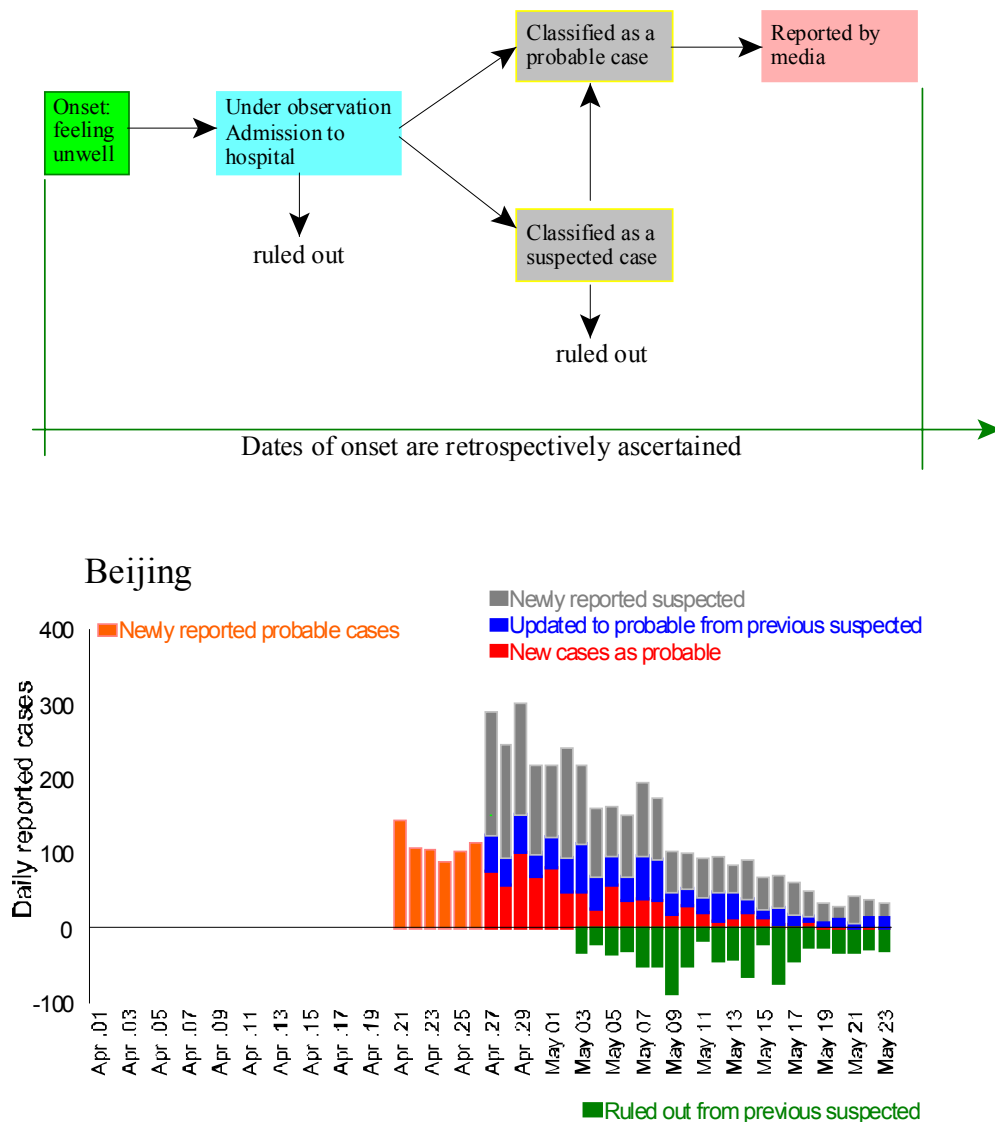


Figure 1. Illustration of mechanisms that manifest data as reported and an example from Beijing

The time trends by dates of onset are approximations to the epidemic curves. These curves are available from a number of government websites and the World Health Organization website. With some knowledge of the incubation period and the shape of its distribution, statistical methods can suggest the most plausible epidemiologic curve by the time of infection. Once one can make a plausible reconstruction of the realized infection process, there will be two immediate implications. The first is to compare the infection process with the observed trends, especially with that by dates of report, since most of the public reactions are driven by reported outbreaks. This will help to design more timely response measures based on lessons learned. The second is to use reconstructed infection process, together with available data, as well as knowledge of the incubation distribution, and knowledge of the “removal process”, characterized by the distributions of time from the onset of symptoms to either death or recovery, to estimate some of the most important epidemiology parameters, such as infectivity. All data used in this manuscript are available from public domain. The World Health Organization gives the following website, where one can download epidemiologic curves of SARS from selected regions in the world as reported by May 2, 2003

http://www.who.int/csr/sarsepicurve/2003_05_02epicurve.pdf

The Singapore Ministry of Health and Health Canada also routinely publish respective epidemiologic curves on their own websites:

<http://app.moh.gov.sg/sar/sar01.asp>
http://www.hc-sc.gc.ca/pphb-dgsp/sars-sras/prof_e.html

In order to keep consistency, only probable cases are considered.

METHOD

The back-calculation methods used here are the EMS algorithm as proposed by Becker, Watson and Carlin (2), originally developed in the HIV/AIDS context to assess the extent of the HIV epidemic and to use the reconstruction as a basis for predicting AIDS incidence. Becker and Britton (3) also pointed out the usefulness of reconstructing the infection process for other diseases for the purpose of taking the advantage of the explicit expressions available for maximum likelihood estimates of parameters when the infection process is fully observed.

The incubation distribution for the time the date of infection to the onset of symptoms must be used in back-calculation. The current consensus, based on empirical observations, seems to agree that the average incubation time is short: 2 – 10 days (3). There are three drawbacks. The first is the representativeness, because in these studies, only a subset of cases that can be ascertained to a single exposure event whereas for the majority of the cases, there are multiple exposures points or with exposures that can not be easily defined and measured. The second is relatively small number of cases that can be ascertained to a single exposure event. This implies large uncertainty in terms of confidence limits. The third is that the dates of exposure are retrospectively ascertained from diagnosed SARS patients. Data suffer from time-length bias. For each patient, the observed incubation time is limited to an observation window no longer than the length from the date of exposure to the date of analysis. This window makes one to over sample "shorter" incubation periods.

Based on observed data, there might be occasional longer observed incubation time and might be considered as "outliers". As time goes by, what one characterizes as "outliers" today, may not

look at outliers any more. So the question is, what might be the percentage of SARS cases which might have incubation period longer than 10 days. The proposed model used in the back-calculation is a log-logistic distribution. The choice is based on the following reasons:

1. Flexible: suitable to describe the distribution in a population consist of a mixture of individuals with short and long incubation periods and the capacity to accommodate “outliers”. Both log-logistic and log-normal distributions are particularly suitable for this consideration.
2. Empirical: Sartwell (5, 6) found that log-normal distributions gave good descriptions of the variation in incubation periods for a considerable number of well-known disease. It is well known that the log-logistic distribution provides a good approximation to the log-normal distribution.
3. Practical: the log-logistic distribution has a much simpler algebraic expression than the log-normal distribution. It is very practical because the log-logistic distribution can be parametrized by any given two quantiles. For example, the standard form of a log-logistic distribution can be expressed by

$$\Pr\{X > x\} = \frac{1}{1 + (\lambda x)^\beta}$$

where the inverse of scale parameter is the median: $m = \lambda^{-1}$. β is a shape parameter which, together with the median parameter m , defines other quantiles of the distribution. Therefore, the log-logistic distribution can be re-parametrized from *scale-shape* (λ, β) to median - 95% quantile: (m, t_{95}) . The conversion is

$$m = \frac{1}{\lambda}$$

$$t_{95} = \frac{1}{\lambda} \times \exp\left\{\frac{2.944438979}{\beta}\right\}$$

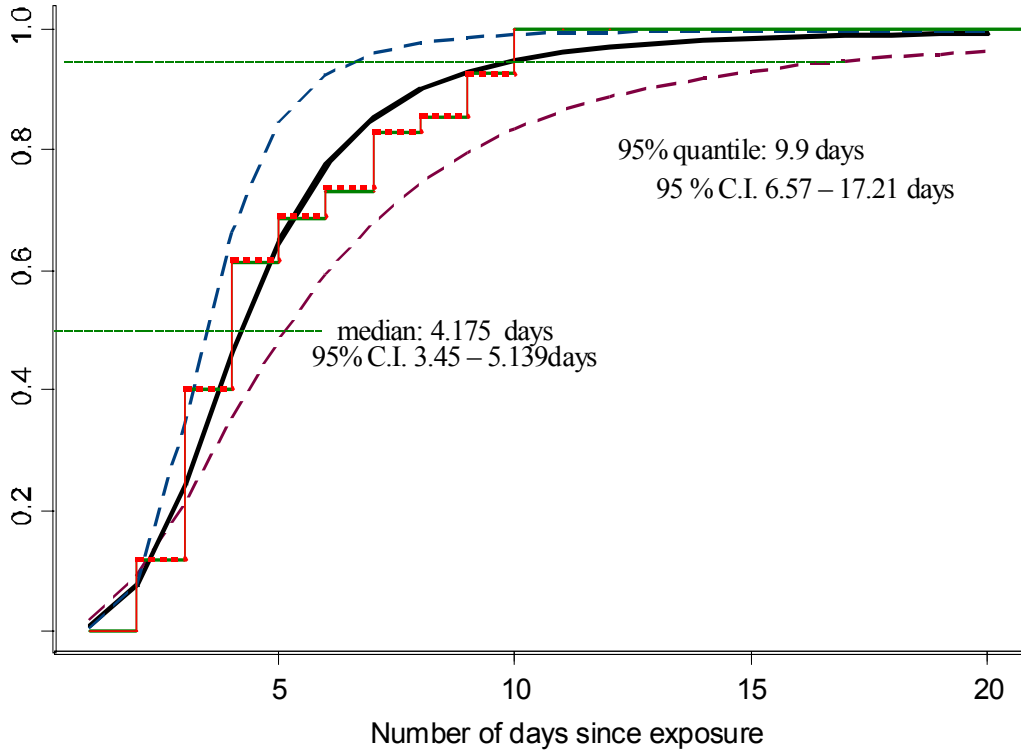


Figure 2. Incubation distributions used for back-calculation.

The parametric models for the incubation distribution in the back-calculations are represented by the three smooth curves in Figure 2. These are the cumulative probabilities from the time of infection to the time of onset measure in days. The solid middle line corresponds to a log-logistic distribution with $m = 4.18$ days and $t_{95} = 9.9$ days. The dotted line with a shorter distribution has $m = 3.45$ days and $t_{95} = 6.57$ days. The dotted line with a longer distribution has $m = 5.14$ days and $t_{95} = 17.21$ days. These three scenarios are compared with empirically estimated distribution, represented as step-function based on the 42 cases in Ontario as documented in (4).

RESULTS

For each of the following countries or regions, three reconstructed infection curves in lines are presented, corresponding to the short, medium and long incubation distributions as illustrated in Figure 2. For comparison purposes, time trends by dates of onset and by dates of report are presented in bars.

Viet Nam:

There has been no newly reported probable cases since April 14, 2003. Therefore, data presented by dates of onset are assumed as accurate. No further reporting delay adjustment is needed. The estimated infection curves using back-calculation show a clear pattern as often described in mathematical models. The short, medium and long incubation distribution assumptions do not seriously affect the interpretations. Comparing with the epidemic curve represented by dates of onset, the latter not only suffer a slight delay in time, but also a slight distortion in shape, due to the variance of the incubation distribution.

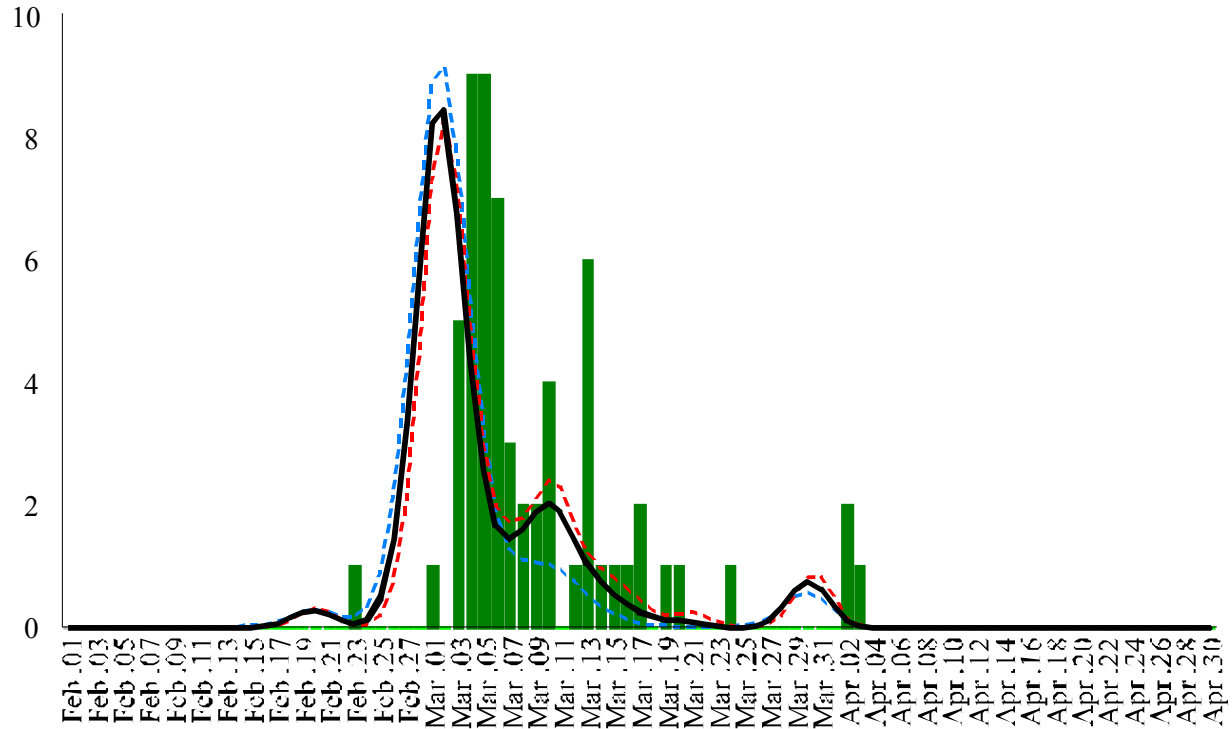


Figure 3. Reconstructed infection curves compared with trend by dates of onset in Viet Nam.

Singapore:

Back-calculated infection curves of SARS are presented in Figure 4, together with trends by dates of onset and by dates of report. Trend by dates of report was compiled from daily briefings from the Singapore Ministry of Health website:

<http://app.moh.gov.sg/sar/sar03.asp>

Documented transmission and linkages among these clusters are also available from the aforementioned website. Therefore, some major events related to public reactions are also illustrated in Figure 5. The acronyms are:

TTSH = *Tan Tock Seng Hospital*

SGH = *Singapore General Hospital*

PPWM = *Pasir Panjang Wholesale Market*

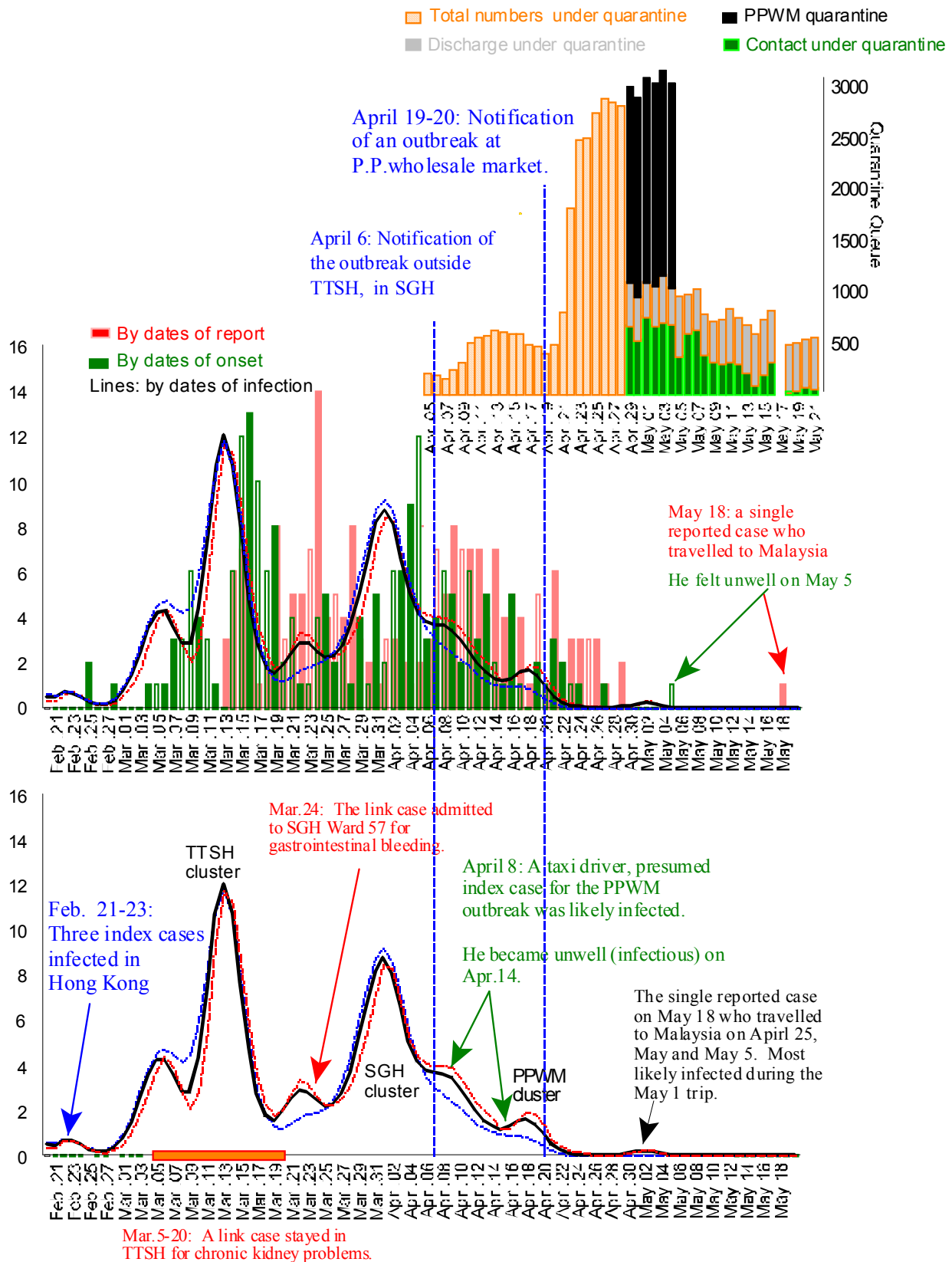


Figure 4. Singapore: Infection curves compared with trends by onset and report dates, along with document major event history on public notification and action.

In addition, Figure 4 illustrates how major actions are coincidence with the trend by dates of report, which took place immediately at the newly reported clusters of outbreaks. However, the infection time often took place approximately two weeks before the public notification.

Hong Kong:

Since Hong Kong is continuing reporting new SARS cases, one needs to adjust the numbers by dates of onset for reporting delays. The authors thanks Dr. Nigel Gay from the World Health Organization for providing the reporting delay adjusted numbers used for back-calculation.

The event history about the outbreak in the Amoy Gardens Complex is used in Figure 5 to illustrate the time lag between the infection time and the time when reported cases started to alert public and drove the action.

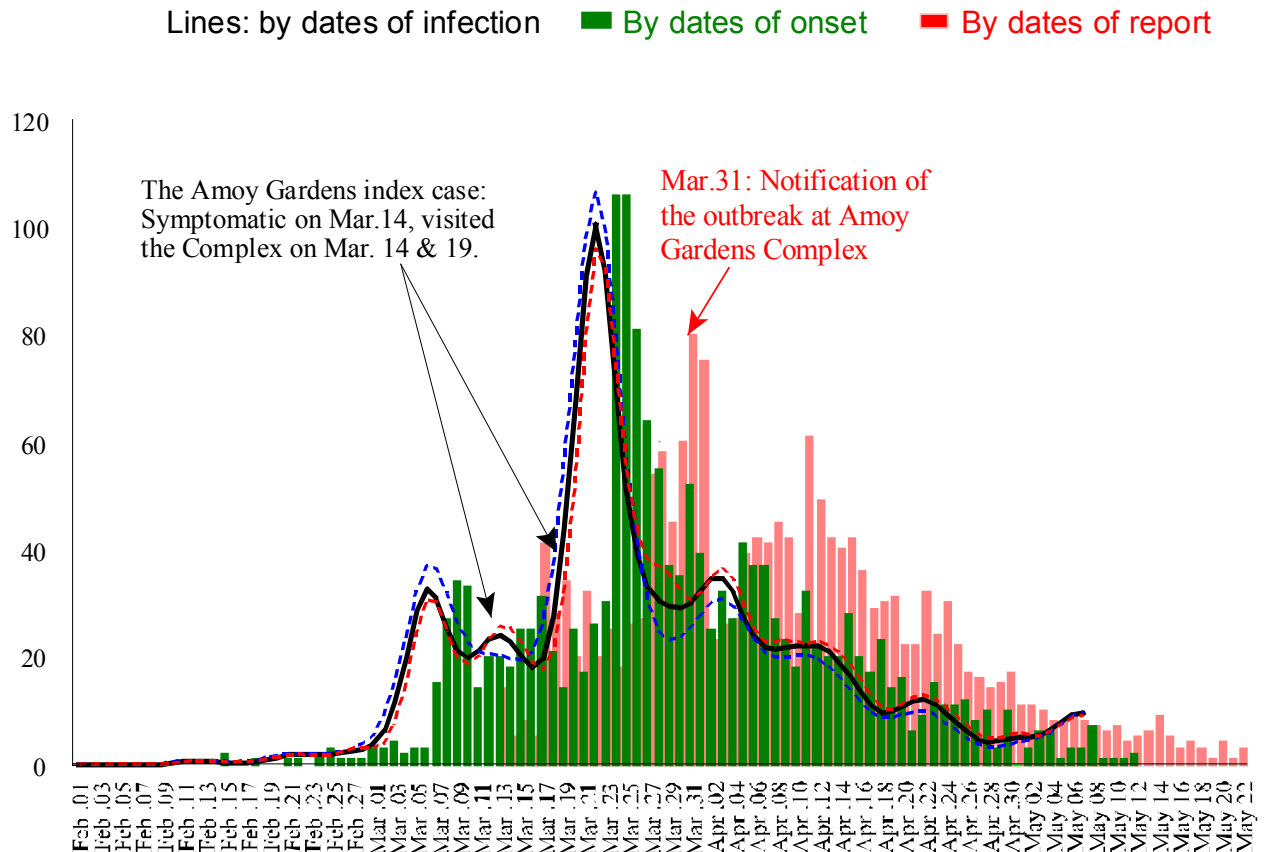


Figure 5. Hong Kong: infection curves compared with trends by onset and by report dates

Canada:

The trend analysis for Canada has the same message as what we have seen for Singapore and for Hong Kong. Figure 6 uses a single event to highlight the time lag between the earliest public notice of the outbreak in a religious group, which was driven by the elevated reported cases on April 14 and 15. The most plausible starting point of this infection is before March 30.

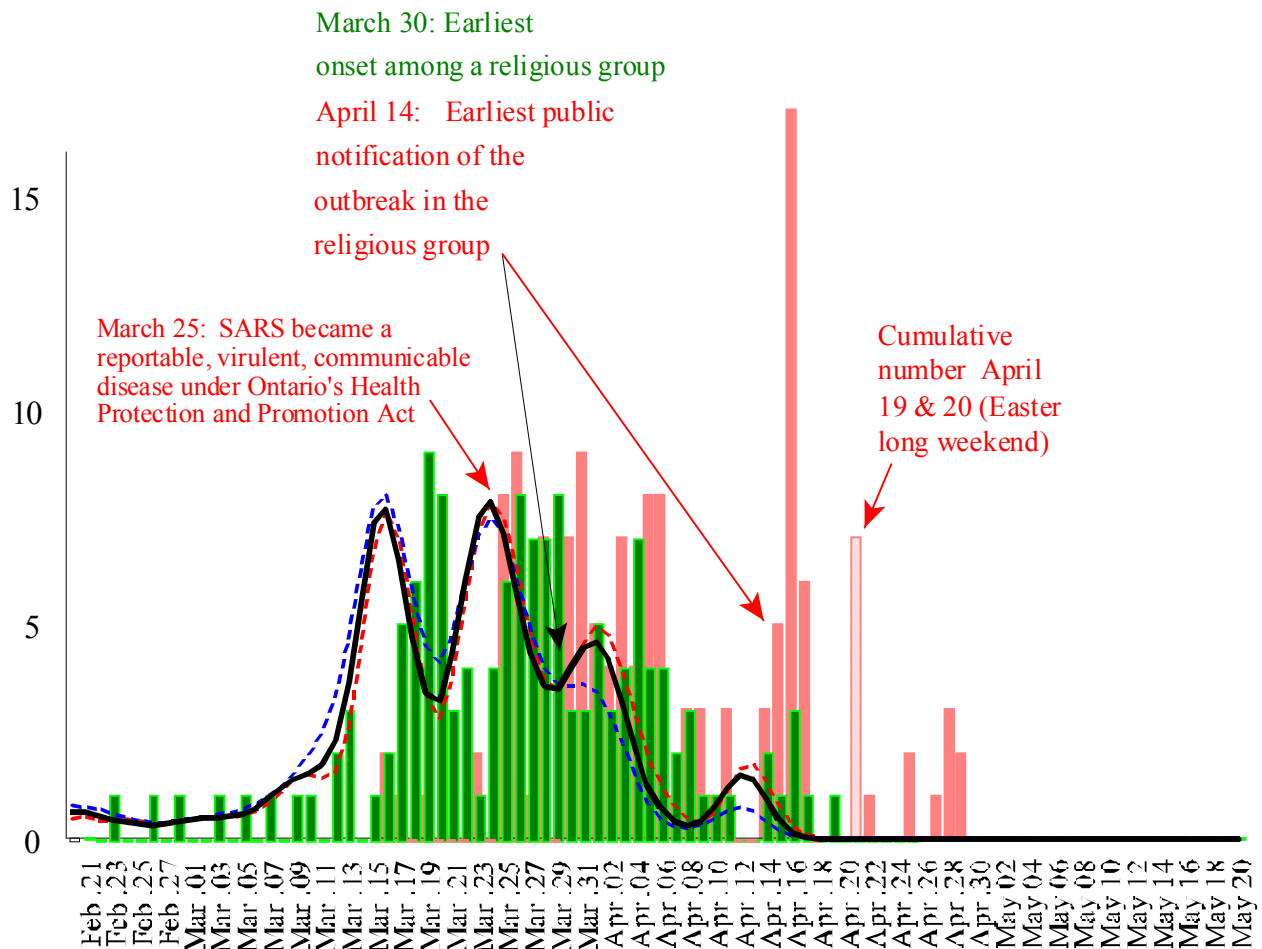


Figure 6. Canada: Infection curves compared with trends by onset and report dates, along with some document event history on public notification.

Synthesized infection curves:

Figure 7 presents the infection curves reconstruction for different regions based on the medium scenario of the incubation distribution, which is a log-logistic distribution with median = 4.18

days and 95% quantile = 9.9 days. All these regional outbreaks are linked to a single “super-spreading event” that took place between Feb. 21 and Feb. 23, 2003 at the ninth floor of the Metropole Hotel in Hong Kong. In terms of infection dynamics over time, Figure 7 demonstrates a synchronized recurrent waves, not necessarily dampening, between the infection curve in Hong Kong and the infection curve in Singapore. For all these four regions, the recurrent waves tend to share a common periodicity.

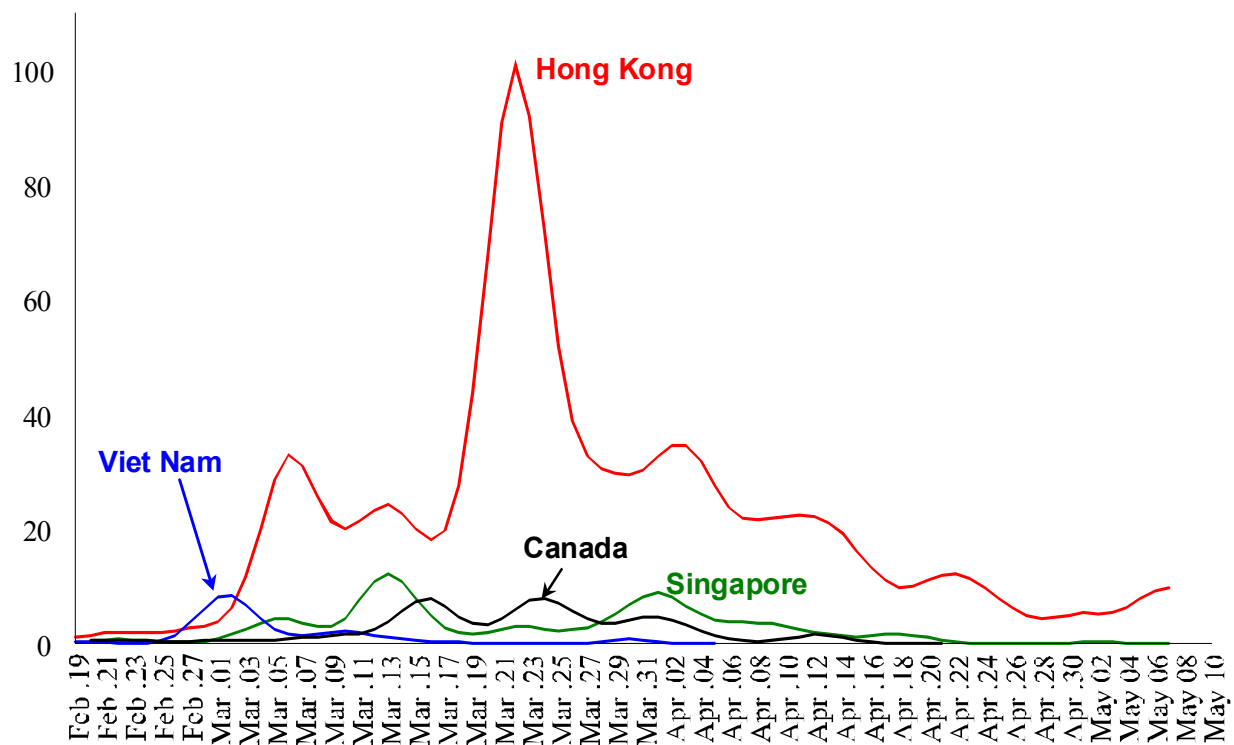


Figure 7. Infection curves reconstructed for different regions based on the medium scenario of the incubation distribution.

CONCLUSION

Back-calculation methods, widely used in the study of the HIV/AIDS epidemic in early 1990's, have found use in the current fight against SARS. This analysis illustrates the following aspects:

1. Public notification and reaction are driven by reports, which tend to be approximately two weeks after the initiation of a local outbreak. Further research is necessary in

identifying ways to shorten this time lag, and to identifying ways for more efficient control measures.

2. For modellers who fit mathematical models to observed epidemic curves, it is important to know the underline mechanisms that manifest the data, including reporting patterns driven by local jurisdictions and occasional irregularities.
3. Back-calculated infection curves provide additional information on the infection process, which are not directly observable. With this additional information, one may be able to estimate some important parameters in mathematical models, such as infectivity, which may not be easily estimated only based on observed data.

The back-calculated epidemic curves provide empirical contribution to more advanced mathematical modelling. Some open challenges exist. For example, data may exist through contact tracing such that one can construct a linked data as who infects whom. Therefore, it is possible to match the reconstructed trends by dates of infection over calendar time with that observed by generation time. In classical mathematical models for infectious diseases, it is often assumed that the initial spread of an outbreak in calendar time can be approximated by an exponential curve represented by $i(t) \propto Ce^{rt}$. It is also often assumed that the initial spread by generation time can be approximated by a time-stationary branching process, defined by a mean parameter R_0 , and that the decline of the number of susceptibles in subsequent generations in the early phase can be ignored. Under certain assumptions, such as the S-I-R model under random mixing, the initial growth in calendar time and in generation time can be matched through $r = R_0 - 1$, and R_0 is the basic reproductive number.

On the other hand, the transmission of SARS has been mainly confined to hospital settings, within households of infected patients, and in some cases, within close-knit communities such as

a specific religious group in Ontario and possibly crowded university dormitories in Beijing. Such environments may contribute to the “super-spreading events”, characterized by an initial case infecting many secondary cases and leaving each case in the second and third generations infecting relatively few. Hence one may observe a non-stationary branching process in the very initial phase in generation time. Therefore, if one uses the infection curve in calendar time to estimate such a parameter R_0 , it will be useful to discuss the meaning of this parameter in a branching process setting. Intuitively, it may correspond to a “pseudo-branching process” which is time-stationary, with R_0 being the average of the number of secondary cases from the initial “super-spreader” and the number of number secondary cases from cases in the next few generations during this initial phase from the observed non-stationary branching process. In other words, the initial growth in generation time described by the observed non-stationary branching process generates the initial growth of the infection curve in calendar time, but the latter can be mimicked by the time-stationary pseudo-branching process. I take notice that such work is underway, with references to two recently published articles by Lisptich, et al. (7) and by Riley, et al. (8). In their work, detailed data from Singapore and Hong Kong have been analyzed with special treatments to the “super-spreading events”, without referencing the branching process aspects. I hope that the back-calculated infection processes which are not in the observed data may provide additional insight to their work and additional work needs to be done.

REFERENCES

- (1) The website of Ministry of Health, People's Republic of China (in Chinese language):
<http://www.moh.gov.cn/zhgl/yqfb/index.htm>.
- (2) Becker, N.G., Watson, L.F. and Carlin, J.B. (1991). A method of non-parametric back-projection and its application to AIDS data. *Statistics in Medicine*, **10**, 1527-1542.
- (3) Becker, N.G. and Britton, T. (1999) Statistical studies of disease incidence. *Journal of Royal Statistics Society (B)*, **61**, 287–307.
- (4) *Epi-Update: Interim Report on the SARS outbreak in the Greater Toronto Area, Ontario, Canada, April 24, 2003*. http://www.hc-sc.gc.ca/pphb-dgsp/sars-sras/pef-dep/gta-20030424_e.html. Health Canada.
- (5) Sartwell, P.E. (1950). The distribution of incubation periods of infectious disease. *American Journal of Hygen*, **51** , 308-318.
- (6) Sartwell, P.E. (1966). The distribution of incubation period and the dynamics of infectious disease. *American Journal of Epidemiology*, **83** , 204-216.
- (7) Lipsitch, M, Cohen, T., Robins, J.M., Ma, S., et al. (2003) Transmission dynamics and control of Severe Acute Respiratory Syndrome. *Scienceexpress/www.scienceexpress.org/23 May 2003/Page 1/10.1126/science.1086616*.
- (8) Riley, S., Fraser, C., Donnelly, C.A., Ghani, A.C., et al. (2003) Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Scienceexpress/www.scienceexpress.org/23 May 2003/Page 1/10.1126/science.1086478*.