Electronic Supporting Material for "Using models to identify routes of nosocomial infection: a large hospital outbreak of SARS in Hong Kong" by Kin On Kwok, Gabriel M Leung, Wai Yee Lam and Steven Riley

Supporting Methods and Sensitivity Analyses

Construction of the case database

We constructed a hospitalisation episode database for all 1755 probable SARS cases in Hong Kong (Leung, Hedley et al. 2004), based on the Department of Health (DH) master list (containing epidemiologic and contact tracing information) and the Hospital Authority (HA) eSARS system (consisting of clinical data from the wards) (Leung, Hedley et al. 2004). We integrated 64 "snapshots" of eSARS between March 11th and July 31st, and included only cases confirmed as probable SARS as per the DH master list. This retrieved 108,596 entries, of which 3,155 had unique Hong Kong identity card (ID) numbers and hospital admission dates and so might have represented distinct SARS hospital episodes. The remainder were discarded as they were exact duplicates generated by multiple downloads from the eSARS system. Inconsistencies in admission and discharge dates were resolved by assuming that entries from later downloads were more accurate. If discharge destinations were missing for episodes in acute care hospitals, when the next episode occurred in a convalescent hospital, the discharge date from the acute hospital was set to be equal to the admission date for the convalescent hospital (approximately half of the 1453 survivors rehabilitated in a convalescent hospital after acute treatment before discharge home). After further cleaning and validation, we eventually identified 2,555 distinct hospital episodes for the 1755 SARS patients. Of these, 434 episodes occurred at Hospital P, 390 of which were first episodes for the individuals concerned. The remaining 44 were second episodes.

Source of infection classification

The source of infection for each SARS patient was identified based on the data from the SARSID database (integrated from the Department of Health master list and the eSARS system from the Hong Kong Hospital Authority), and cross-referenced with the Report of the SARS Expert Committee (SARS Expert Committee 2003) - where the epidemiological links of some of the cases where available. In combination with the first hospitalisation episode data as identified above the 390 subjects were classified as (numbers in parentheses): staff treated at hospital P (164), community acquired (108), inpatients treated at Hospital P (60), other (23), visitor (19) staff treated elsewhere (11) and inpatients treated elsewhere (5).

Waiting times for the transmission model

The expected durations for different stages in the model were calculated as follows. The average incubation period τ_E was assumed to be equal to 4.6 days (Donnelly, Ghani et al. 2003; Leung, Hedley et al. 2004), a value based on the entire Hong Kong case database. The average time that staff continued to work prior to being admitted, τ_{SW}, was assumed to be 3.3 days, a value calculated for staff in Hospital P. We did not assume that cases remained infectious for the entire duration of their stay in

hospital. Patients infected in the hospital were infectious for, on average, τ_{IP} days after the onset of symptoms; patients infected in the community for τ_{XIP} days after admission and staff for an additional τ_{SP} days after admission. We defined the generation time τ_G to be the average time from a case being infected to the infection of her infectees and we assumed that the generation time for staff and patients infected at Hospital P was equal to 8.4 days, a value obtained from analyses of contact-tracing data in the predominantly hospital-based Singapore outbreak (Lipsitch, Cohen et al. 2003). For patients, the generation time $\tau_G = \tau_E + \tau_{IP}/2$, where τ_{IP} was the average time that patients were infectious after the onset of symptoms. We assumed that patients from the community had been infectious for an average τ_{OA} days after onset before admission. Therefore, we set the average time that these patients were infectious within the hospital $\tau_{XIP} = \tau_{IP} - \tau_{OA}$. For staff,

$$\tau_G = \frac{\tau_{SW}\left(\tau_E + \tau_{SW}/2\right) + \alpha\tau_{SP}\left(\tau_E + \tau_{SW} + \tau_{SP}/2\right)}{\tau_{SW} + \alpha\tau_{SP}},$$

where τ_{SP} was the average time that staff continued to be infectious after admission. The above relationships were used to calculate vales of τ_{IP} and τ_{SP} from the other waiting times. The average duration from onset to discharge from hospital for all cases was defined to be τ_{DS} and calculated from our integrated database to be equal to 26.5 days. After infectiousness had ceased, staff remained within Hospital P for an average of τ_{SR}, patients infected in the hospital for τ_{PR} and patients infected in the community for an average time of τ_{XPR}. Values for τ_{SR}, τ_{PR} and τ_{XPR} were calculated so as to be consistent with τ_{DS} and the other waiting time parameters, so $\tau_{SR} = \tau_{DS} - \tau_{SP} - \tau_{SW}$, $\tau_{PR} = \tau_{DS} - \tau_{IP}$ and $\tau_{XPR} = \tau_{DS} - \tau_{XIP}$.

Definition of model

Here, we define a model which is slightly more general than that described in the main text. This model was used for sensitivity analyses (below in this document). The version of the model used for all results in the main paper can be recovered easily if the number of staff treated elsewhere is assumed to be zero.

The variables for staff treated at hospital P, patients and the super-spreader are defined in the main text. In addition, let N_E be the number of staff who would be treated elsewhere if sufficiently symptomatic to seek medical attention. This group is made up of S_E susceptible individuals, E_E exposed but not yet infectious individuals and I_E infectious individuals.

The force of infection experienced by (those who would be treated in hospital P and those who would be treated elsewhere) was $\lambda_S(t) = \gamma_S(t)\lambda(t)$ and by patients $\lambda_P(t) = \gamma_P(t)\lambda(t)$,

where $\lambda(t) = \beta\left[\frac{I_S(t)+I_E(t)}{N_s} + \alpha_P\left\{\frac{I_P(t)+I_{SP}(t)+I_{XP}(t)}{N_p}\right\} + \alpha_{SSP}\right]$ for $t < t_{SSP}$ and

$\lambda(t) = \beta\left[\frac{I_S(t)+I_E(t)}{N_s} + \alpha_P\left\{\frac{I_P(t)+I_{SP}(t)+I_{XP}(t)}{N_p}\right\}\right]$ for later times.

If we had used the exponential distribution for all waiting times, the dynamic model would be specified for patients infected in the hospital by the following ordinary differential equations, dropping explicit time dependencies (all symbols not describing staff treated elsewhere are defined in the main text),

$$\begin{aligned}
\dot{S}_P &= \tau_{SR}^{-1}R_S + \tau_{PR}^{-1}R_P + \tau_{XPR}^{-1}R_{XP} - \tau_{SW}^{-1}I_S - \delta - \lambda_P S_P,\\
\dot{E}_P &= \lambda_P S_P - \tau_E^{-1}E_P,\\
\dot{I}_P &= \tau_E^{-1}E_P - \tau_{IP}^{-1}I_P,\\
\dot{R}_P &= \tau_{IP}^{-1}I_P - \tau_{PR}^{-1}R_P.
\end{aligned}$$

Note that the model was defined so that the number of staff, N_S, the number of patients, N_P, the number of staff treated elsewhere N_E were constant. The term $\tau_{SW}^{-1}I_S$ in the equation for $\dot{S}_P$ reflects the assumption that admitted staff were replaced by susceptible staff. Similarly, the term $-\delta$ reflects the assumption that imported SARS cases took up a space that would have been occupied by a susceptible patient otherwise. For staff infected in the hospital and treated there,

$$\begin{aligned}
\dot{S}_S &= \tau_{SW}^{-1}I_S - \lambda_S S_S,\\
\dot{E}_S &= \lambda_S S_S - \tau_E^{-1}E_S,\\
\dot{I}_S &= \tau_E^{-1}E_S - \tau_{SW}^{-1}I_S,\\
\dot{I}_{SP} &= \tau_{SW}^{-1}I_S - \tau_{SP}^{-1}I_{SP},\\
\dot{R}_S &= \tau_{SP}^{-1}I_{SP} - \tau_{SR}^{-1}R_S.
\end{aligned}$$

For patients infected outside of the hospital,

$$\begin{aligned}
\dot{I}_{XP} &= \delta - \tau_{IXP}^{-1}I_{XP},\\
\dot{R}_{XP} &= \tau_{IXP}^{-1}I_{XP} - \tau_{XPR}^{-1}R_{XP}.
\end{aligned}$$

For those staff infected but treated elsewhere,

$$\begin{aligned}
\dot{S}_E &= \tau_{SW}^{-1}I_E - \lambda_S S_E,\\
\dot{E}_E &= \lambda_S S_E - \tau_E^{-1}E_E,\\
\dot{I}_E &= \tau_E^{-1}E_E - \tau_{SW}^{-1}I_E
\end{aligned}$$

The next generation matrix and the basic reproductive number

The next generation matrix (Diekmann and Heesterbeeck 2002) for infections (excluding the super-spreader) includes only staff and patients infected in the hospital. Infection events which occurred in the community or other hospitals are not described by the model. Let the matrix

$$\mathbf{M} = \beta \begin{bmatrix} \tau_{SW}^{-1}\frac{N_s}{N_s+N_E} + \tau_{SP}^{-1}\alpha_P\gamma_S(0)\frac{N_s}{N_p} & \tau_{IP}^{-1}\alpha_P\gamma_S(0)\frac{N_s}{N_p} & \tau_{SW}^{-1}\frac{N_s}{N_s+N_E} \\ \tau_{SW}^{-1}\gamma_P(0)\frac{N_p}{N_s+N_E} + \tau_{SP}^{-1}\alpha_P\gamma_P(0) & \tau_{IP}^{-1}\alpha_P\gamma_P(0) & \tau_{SW}^{-1}\gamma_P(0)\frac{N_p}{N_s+N_E} \\ \tau_{SW}^{-1}\frac{N_E}{N_s+N_E} + \tau_{SP}^{-1}\alpha_P\gamma_S(0)\frac{N_E}{N_p} & \tau_{IP}^{-1}\alpha_P\gamma_S(0)\frac{N_E}{N_p} & \tau_{SW}^{-1}\frac{N_E}{N_s+N_E} \end{bmatrix}.$$

The first column contains the expected number of new infections generated by a single infectious staff member. The first row of the first column is the expected number of new staff infections, the second row of the first column is expected number of new patient infections and the last row of the first column is expected number of infections of new staff treated elsewhere. The second term of each of these expressions contains a factor α_P because staff spent the second phase of their infectious stage, after admission, as patients. Similarly, the second column contains the expected number of secondary cases of staff (first row), patients (second row) and staff treated elsewhere (last row) generated by each infectious patient. We defined the basic reproductive number excluding super-spreaders R_0^{XSS} to be equal to the dominant eigenvalue of $\mathbf{M}$. This definition is consistent with R_0^{XSS} being the average number of secondary infections generated by a single typically infectious non-super-spreading individual in an otherwise susceptible population. Therefore, without loss of generality, we used R_0^{XSS} rather than β to define the basic level of transmission in the model. We tested empirically that for a special case of large population sizes and a single seed, the critical point implied by $\mathbf{M}$ was equal to that observed in the version of the model used for the results in the main paper (with non-exponential waiting times).

For the two-class model used in the main article, R_0^{XSS} and β are related explicitly by the formula

$$R_0^{XSS} = \frac{\beta}{2}[(a+d) + \sqrt{4bc + (a-d)^2}]$$

where

$$a = \tau_{SW}^{-1}\frac{N_s}{N_s+N_E} + \tau_{SP}^{-1}\alpha_P\gamma_S(0)\frac{N_s}{N_p}$$

$$b = \tau_{IP}^{-1}\alpha_P\gamma_S(0)\frac{N_s}{N_p}$$

$$c = \tau_{SW}^{-1}\gamma_P(0)\frac{N_p}{N_s + N_E} + \tau_{SP}^{-1}\alpha_P\gamma_P(0)$$

$$d = \tau_{IP}^{-1}\alpha_P\gamma_P(0)$$

Sensitivity analyses

In Table S1, we show that our results are not sensitive to changing the sizes of the populations of staff or patients, until the populations are small enough that saturation of susceptibles may have become a factor. Also, we show that our results are not sensitive to the inclusion of visitors (as patients) and staff-treated-elsewhere as a 3rd type of infectious class.

Table S1. Sensitivity analyses. Estimates for the parameters presented in Table 2 are recalculated under different assumptions. Firstly, we keep the structure of the model the same and vary the size of the susceptible staff and patient populations. These changes reflect uncertainty over the effective population size within which the pathogen had the opportunity to circulate. In the final row of the table, we use extended model structure (as described above in this document) to test the sensitivity of our conclusions to the inclusion of visitors (n=19, treated as patients) and staff treated elsewhere (n=11, included as 3rd type in extended model).

Model	Np	Ns	Hypothesis	R_0^{XSS}	α_{SSP}	t_{SSP}	α_P	$\gamma_P(0)$	t_I	Δ_P^{γ}	Δ_S^{γ}	ΔAIC
			H_1	174	-	-	0.085	0.302	3	0.010	3.40×10^{-3}	0
Basic	1315	2250	H_2	0.660	76.8	4	0.112	0.645	-	-	-	14.3
			H_3	0.595	48.6	4	0	0.382	25	4.99	1.29	2.21
			H_1	175	-	-	0.087	0.291	3	0.011	3.44×10^{-3}	0
Basic	1125	657	H_2	0.674	73.5	4	0.112	0.719	-	-	-	16.22
			H_3	0.608	48.1	4	0	0.391	25	5.78	1.27	1.81
			H_1	175	-	-	0.093	0.266	3	0.016	3.45×10^{-3}	0
Basic	562	328	H_2	0.708	71.4	4	0.114	0.776	-	-	-	20.68
			H_3	0.638	47.1	4	0	0.413	25	8.87	1.24	0.91
			H_1	170	-	-	0.112	0.231	3	0.027	3.51×10^{-3}	0
Basic	350	205	H_2	0.820	40.8	4	0.028	0.915	-	-	-	28.62
			H_3	0.785	45.4	4	0.028	0.428	25	5.56	0.823	2.85
			H_1	170	-	-	0.096	0.252	3	0.01	3.50×10^{-3}	0
Extended	1315	2250	H_2	0.665	84.6	4	0.119	0.536	-	-	-	14.18
			H_3	0.598	58.4	4	0	0.316	25	5.00	1.28	1.81

References

Diekmann, O. and J. A. P. Heesterbeeck (2002). Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. New York, Wiley.

Donnelly, C. A., A. C. Ghani, et al. (2003). "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong." Lancet **361**(9371): 1761-6.

Leung, G. M., A. J. Hedley, et al. (2004). "The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients." Ann Intern Med **141**(9): 662-73.

Lipsitch, M., T. Cohen, et al. (2003). "Transmission dynamics and control of severe acute respiratory syndrome." Science **300**(5627): 1966-70.

SARS Expert Committee (2003). "SARS in Hong Kong: from Experience to Action." www.sars-expertcom.gov.hk.