APPENDIX 1 – General Results for Branching Processes:

We view an epidemic as a discrete-time branching process beginning with I_0 infected individuals, and where time is indexed by *n*. Thus the total number of infections that have occurred by generation *n* represents the total number that have occurred once the transmission lineage stemming from each initially infected individual has passed through exactly *n* generations (or else it has gone extinct prior to this occurring). Let *N* denote a random variable representing the number of new infections generated by a single infected individual. The mean and variance of *N* are denoted by and v^2 respectively. Using T_n to denote the random variable representing the total cumulative epidemic size by generation *n*, we then have

$$E\ T_n \quad \frac{1 \quad ^{n\ 1}}{1} I_0 \tag{S1.1}$$

and

$$\text{var}[T_n] \quad v^2 \frac{1 \quad ^{2n\ 1} \quad 1 \quad 2n \quad 1 \quad ^{n}}{1 \quad ^{3}} I_0. \tag{S1.2}$$

We can then write equation (S1.1) more explicitly using the notation of the text.
The expected total number of infections that have occurred in the presence of both quarantine and isolation up until generation *n* is

$$\overline{T}_n \quad \frac{1 \quad (1 \quad q \ )\overline{R}_I \ ^{n\ 1}}{1 \quad (1 \quad q \ )\overline{R}_I} I_0, \tag{S1.3}$$

where we have just substituted $q\overline{R}_{QI} \quad (1 \quad q)\overline{R}_I$ into equation (S1.1) and simplified. With this notation, we can write the expression of the variance in T_n more explicitly by first noting that v^2 is calculated as follows:

$$v^2 \quad q\mathrm{E}\ R_{QI}^2 \quad (1 \quad q)\mathrm{E}\ R_I^2 \quad q\overline{R}_{QI} \quad (1 \quad q)\overline{R}_I^{\ 2}. \tag{S1.4}$$

Thus, we have

$$\mathrm{var}\ T_n \quad v_R \frac{1 \quad (1 \quad q\ )\overline{R}_I^{\ 2n\ 1} \quad q\ \overline{R}_I\ 2n \quad 1\ (1 \quad q\ )\overline{R}_I^{\ n}}{1 \quad (1 \quad q\ )\overline{R}_I^{\ 3}} I_0, \tag{S1.5}$$

where

$$v_R \quad q\ _{QI}^{\ 2} \quad (1 \quad q)\ _I^{\ 2} \quad q(1 \quad q)\ \overline{R}_I \quad \overline{R}_{QI}^{\ 2} \tag{S1.6}$$

is the variance in the number of infections generated by a single infected individual in the presence of both quarantine and isolation.

It is also worth noting that the above results are quite general in so far as they are not based on very many restrictive assumptions. We have allowed the possibility of transmission at all stages of the disease, including from quarantined and isolated individuals. We have also allowed the probability distribution of number of infections generated by a single infected individual to have any form, and the results require only that we know the mean and variance of this distribution. Finally, we note that identical results for the expected number of infections averted can be obtained through the use of other modeling frameworks (e.g., with a deterministic SEIQJR model; T. Day et al, unpubl. Results).

The flexibility in the form of the distribution also means that the results are valid for extreme events such as the super-spreading that occurred in the transmission of SARS (24, 25). For example, super-spreading might be represented as the mixture of two

Poisson distributions, one for super-spreading events and one for “normal” transmission events. In this case, if p is the probability of a super-spreading event occurring, and if $\overline{R}_{I,1}$ and $\overline{R}_{I,2}$ are the means of the distributions of infections occurring during a super-spreading event and during a normal transmission event respectively, then $\overline{R}_I = p\overline{R}_{I,1} + (1-p)\overline{R}_{I,2}$. Moreover, we also have $\sigma_I^2 = \overline{R}_I + p(1-p)\left(\overline{R}_{I,1} - \overline{R}_{I,2}\right)^2$.

APPENDIX 2 – Technical Results:

With the expressions from Appendix 1, the expected number of infections averted by quarantine by generation *n* is $\overline{D}_n \quad \overline{T}_n\big|_{q \ 0} \quad \overline{T}_n$, which can be written in terms of the model parameters as

$$\overline{D}_n \quad F(q, \ \ ,n,\overline{R}_I)\frac{I_0 \ 1 \quad \overline{R}_I^{n \ 1}}{1 \quad \overline{R}_I}, \tag{S2.1a}$$

where

$$F(q, \ \ ,n,\overline{R}_I) \quad \frac{\overline{R}_I \ 1 \quad \overline{R}_I^n \qquad \overline{R}_I^{n \ 1} \qquad (\overline{R}_I \ \ )^n(1 \quad \overline{R}_I)}{(1 \quad \overline{R}_I^{n \ 1})(1 \quad \overline{R}_I \ \ )} \tag{S2.1b}$$

and $\quad 1 \quad q \ .$

Equation (S2.1a) is written as the product of the total expected number of infections occurring in the absence of quarantine by generation *n* (i.e., $I_0(1 \quad \overline{R}_I^{n \ 1})/(1 \quad \overline{R}_I)$) and the percentage of these that are averted through the use of quarantine, $F(q, \ \ ,n,\overline{R}_I)$. We are primarily interested in the total number of infections that can be averted in the long term (i.e., as time, *n*, gets large). In this case, provided that there is either some asymptomatic transmission or that quarantine can increase the effectiveness of isolation (so that $\quad 0$), the percentage of infections that are averted by quarantine is

$$\lim_n F \qquad \overline{R}_I \quad (1 \quad q \ )\overline{R}_I \Big/ 1 \quad (1 \quad q \ )\overline{R}_I \qquad \begin{matrix} 1 \\ \text{if } \overline{R}_I \quad 1 \\ \text{if } \overline{R}_I \quad 1 \end{matrix}. \tag{S2.2}$$

On the other hand, if $\quad 0$ then quarantine will never have any effect. Equation (S2.2) reveals that we must consider two cases separately: (i) isolation alone cannot stop the

spread of the disease (i.e., $\overline{R}_I$ 1), and (ii) isolation alone can stop the spread of disease (i.e., $\overline{R}_I$ 1).

If isolation alone cannot stop the spread of disease (i.e., $\overline{R}_I$ 1) then the total number of infections occurring in the absence of quarantine will, with nonzero probability, grow indefinitely until the pool of susceptible hosts is depleted. At the same time, however, the percentage of these that quarantine can be expected to avert approaches 100 percent (equation S2.2). As a result, quarantine is likely to be very beneficial in this case. It is only when there is very little asymptomatic transmission, and when quarantine has very little effect on the efficiency of isolation procedures, that it will have a marginal effect. In this case, q will be close to zero, and although the percentage of infections averted by quarantine still approaches 100 percent as time passes, it will do so much more slowly. In particular, if $\overline{R}_I$ is large, then the percentage of infections averted by quarantine up to the n^{th} generation is approximately $1 \quad (1 \quad q \quad)^n$. This illustrates that the number of infections averted by quarantine will be small, only when is very close to zero.

If isolation alone can stop the spread (i.e., $\overline{R}_I$ 1) then $I_0(1 \quad \overline{R}_I^{n \ 1})/(1 \quad \overline{R}_I)$ approaches $I_0/(1 \quad \overline{R}_I)$ as time passes. Combining this with results (S2.2) we obtain the expected number of infections that are averted:

$$\lim_n \overline{D}_n \quad \frac{\overline{R}_I \quad (1 \quad q \quad)\overline{R}_I}{1 \quad (1 \quad q \quad)\overline{R}_I} \frac{I_0}{1 \quad \overline{R}_I}, \tag{S2.3}$$

which is equation (1) of the text. If q is close to 1, then (S2.3) simplifies to

$$\lim_n \overline{D}_n \quad \overline{R}_I \frac{I_0}{1 \quad \overline{R}_I}, \tag{S2.4}$$

which reveals that the percentage of infections averted by quarantine is given by $F \quad \overline{R}_I$. This is equation (2) of the text.

The variance in the number of infections averted (when $\overline{R}_I \quad 1$) is given by the variance of the difference $D \quad T\big|_{q \ 0} \quad T$. Deriving an expression for this variance is non-trivial because the random variables, $T\big|_{q \ 0}$ and T , are not independent. A general formula can be obtained but it is quite complex and uninformative (unpubl. results). Therefore we focus on the simplifying case where the number of primary infections that were generated by a given quarantined individual is independent of the number of primary infections from this individual that were avoided because of the use of quarantine. Mathematically, this implies that we can express the random variable, R_I, in the form $R_I \quad R_{QI} \quad W$ where *W* is a nonnegative integer-valued random variable that is independent of R_{QI}. Recall that R_{QI} (respectively, R_I) represents the number of primary infections generated by one individual that is (respectively, is not) placed into quarantine. Thus the random variable, *W*, represents the number of infections from one infected individual that are averted by putting that individual into quarantine (we note that this independence assumption will not hold in the mixture model of super spreaders that we describe at the end of Appendix 1; for such a model, a more general formula is needed – Madras et al. in prep.). Under this independence assumption, the variance of the total number of infections averted by quarantine (under the best-case scenario in which the probability of quarantine, *q*, is equal to one, and the initial number of infected individuals also equals one) is given by (Madras et al., in prep.)

$$\text{var } D \quad \frac{\overset{2}{I}}{(1 \quad \overline{R}_I)^3} \quad \frac{\overset{2}{QI}}{(1 \quad (1 \quad)\overline{R}_I)^3} \frac{1 \quad (1 \quad)\overline{R}_I}{1 \quad \overline{R}_I}, \qquad \text{(S2.5)}$$

where σ_I^2 and σ_{QI}^2 are the variances of distributions, $p_I(\ )$ and $p_{QI}(\ )$ respectively.

From a one-tailed version of Chebeshev's inequality we then know that there is at least an 80 percent chance that the actual outcome will lie below the mean, $\overline{D}$, plus two standard deviations. If we start the epidemic with a single infected individual and have perfect quarantine (q=1) this gives the upper bound as

$$\frac{\overline{R}_I \quad (1 \qquad)\overline{R}_I}{1 \quad (1 \qquad)\overline{R}_I}\frac{1}{1 \quad \overline{R}_I} \quad 2\sqrt{\frac{\sigma_I^2}{(1 \quad \overline{R}_I)^3} \quad \frac{\sigma_{QI}^2}{(1 \quad (1 \qquad)\overline{R}_I)^3}\frac{1 \quad (1 \qquad)\overline{R}_I}{1 \quad \overline{R}_I}}. \qquad \text{(S2.6)}$$

The upper bound (S2.6) is very conservative, however, it is valid regardless of the probability distributions for R_{QI} and R_I. This upper bound is plotted in Figure 3 of the text.

APPENDIX 3 – Stochastic Simulation Results:

As a check on the robustness of the analytical results presented above, we conducted simulations of an individual-based stochastic SEIQJR model (see details below). Parameters needed to be chosen explicitly for these simulations, and therefore we have based our choices on available epidemiological data for SARS. We stress, however, that these simulations are meant for illustrative purposes only, and that the primary objective is to show that the analytical results derived from a branching process agree with a full stochastic simulation of the original model. We also used the simulations to examine how quarantine affects the duration of the epidemic.

Figure S1a reveals that the results of the analytical calculations agree very well with these simulation results. The 80^{th}-percentile upper bound obtained analytically (dashed curve) is extremely conservative as the value of $\overline{R}_I$ increases. This is revealed by the 95^{th}-percentiles of averted cases from the simulations lying well below this 80^{th}-percentile upper bound. The duration of the epidemic as a function of $\overline{R}_I$ also follows a similar qualitative pattern as that of the number of infections averted (Figure S1b). Thus, the use of quarantine in addition to isolation does not appear to produce that much of an advantage in terms of shortening the epidemic duration unless the reproduction number in the presence of isolation, $\overline{R}_I$, is close to one (Figure S1b). When $\overline{R}_I$ is close to one, quarantine appears to dramatically reduce the probability of having very long epidemics.

To obtain these results, we used a standard stochastic SIR-type model (26, 27), but with six classes; susceptible individuals (*S*), exposed (i.e., infected but asymptomatic; *E*), infected and symptomatic (*I*), quarantined (*Q*), isolated (*J*), and recovered and immune (*R*). The resulting SEIQJR model was then simulated with the rate of movement

of each individual in the population among these classes being characterized as a probability of movement per unit time. The transitions and their probabilities of occurrence are summarized in the following table.

TABLE S1: Summary of transitions in stochastic SEIQJR model

Transition	**Description**	**Rate**
$S \rightarrow S$-1 $E \rightarrow E$+1	Infection transmitted	βI
$E \rightarrow E$-1 $I \rightarrow I$+1	Exposed individual becomes infectious	κE
$I \rightarrow I$-1 $R \rightarrow R$+1	Infectious individual recovers	cI
$E \rightarrow E$-1 $Q \rightarrow Q$+1	Exposed individual is quarantined	$\gamma_1 E$
$I \rightarrow I$-1 $J \rightarrow J$+1	Infectious individual is isolated	$\gamma_2 I$
$Q \rightarrow Q$-1 $J \rightarrow J$+1	Quarantined individual becomes infectious	σQ
$J \rightarrow J$-1 $R \rightarrow R$+1	Isolated individual recovers	cJ

Parameters used in the figures are based on SARS epidemiological data for illustrative purposes. The data are summarized in the following table:

TABLE S2: Parameters used in the stochastic simulations

Parameter	**Symbol**	**Value**	**Comments**
Population size	$S+E+I+Q+J+R$	4 million	Estimate of Greater Toronto Area (an example city).
Latent period	$1/\kappa$	6.4 days*	Donnelly et al. (28); Chan-Yeung & Xu (3).
Infectious period	$1/c$	8 days	Estimate (many sources assume no more than 10 days).
Time to quarantine	$1/\gamma_1$	5 days**	Estimate (but assumed longer than $1/\gamma_2$).
Time to isolation	$1/\gamma_2$	various	The time to isolation is varied to explore a range of $\overline{R}_I$ from 0.01 to 0.99.
Transmission rate	β	0.375 days^{-1}	This corresponds to a value of $\overline{R}$ in the absence of all control measures of 3.0 (see refs. (13, 14, 29-32)).

* κ set to zero when modelling "perfect quarantine".

** γ_1 set to zero when modelling "no quarantine", and set to κ when modelling "50% chance of escaping quarantine" so that an individual spends an average of 6.4 days in the E class and exits with equal probability to either the I class or the Q class.

For initial conditions containing at least one in the E or I class, the time to the next event for a single individual is computed as t_{next} $\ln U_0^1 \Big/$, where U_0^1 is a uniform random

variate in the interval (0,1) and Σ is the sum of the rates in Table S1. Then, to determine which event has occurred another U_0^1 is computed. By simple scaling, the rate (ρ_i) corresponding to the i^{th} process in Table S1 occupies a unique segment of length ρ_i/Σ on the real line in the interval (0,1). Therefore, by computing another U_0^1 and calculating which process segment it corresponds to, we ensure that the transitions in Table S1 occur with the correct probability. The population classes (S,E,I,Q,J & R) are updated, as is time, and the process is repeated until there is no more infection or potential infection in the population. The number of infections that occurred and the duration of the outbreak are recorded.

Legend for Figure S1. Results from the stochastic simulations. (a) Number of infections averted by perfect quarantine (i.e., 1 and q = 1). Solid and dashed lines are the expected number averted and the 80th-percentile upper bound from analytical calculations. Calculation assumes that $^2_{QI}$ ^2_I and that the distribution of infections produced by a single individual in the absence of quarantine, $p_I(\ )$, is Poisson (and hence ^2_I $\overline{R}_I$). Dots are simulation results for the average number of infections averted (from 10,000 replicate runs of the simulation). Black bars indicate the 95th-percentile of the number of infections averted in the simulations for the three values $\overline{R}_I$ 0.2, $\overline{R}_I$ 0.5 and $\overline{R}_I$ 0.8. (b) Average length of epidemic (in days) from stochastic simulations for different degrees of quarantine effectiveness (as measured by the product, q).

APPENDIX 4 – Data analysis for :

The data used in Table 1 were based on the estimates from Table 3.1 of Anderson and May (23). We took the average value of the endpoints of the ranges given in Table 3.1 of Anderson and May (23) to obtain single estimates of each. We then estimated the proportion of infections that are generated by asymptomatic individuals for each disease as follows: (i) we calculated the number of days of asymptomatic infectiousness from this data by subtracting the latent period from the incubation period. (ii) we calculated the proportion of infective days that are asymptomatic by dividing the result from (i) by the infectiousness period. Instances where the result was negative imply that symptoms appear before an individual is infectious and therefore we set the result equal to zero. For Hepatitis B and Smallpox the result was greater than one. This occurred because the variability of incubation times (as indicated by the size of the range in Table 3.1 of Anderson and May (23)) was quite large relative to that of the infectious period. In these cases we set the value equal to one. These results are conditional on symptoms eventually developing, but there are some infections (e.g., Polio) for which a many infections never result in the display of symptoms. This latter effect is incorporated into the estimates by calculating the overall value of as $P\ symptoms\ _{symp}\ 1\ P\ symptoms\ 1$ where $P\ symptoms$ is the probability that an infection eventually results in the display of symptoms, and $_{symp}$ is the estimate obtained above (which is conditional on symptoms eventually developing). For smallpox and hepatitis B the value of $_{symp}$ was essentially one and therefore data for $P\ symptoms$ for these diseases was not required. Data for $P\ symptoms$ for the remaining diseases was obtained as follows:

- Influenza $P_{symptoms}$ 0.5 (a consistent and reliable estimate could not be found in the literature, but 0.5 is representative of many expert opinions);
- Diphtheria - a quantitative estimate could not be found, although asymptomatic transmission is known to be significant. Consequently we list it as unknown but likely relatively high;
- Whooping cough $P_{symptoms}$ 0.25 (J. Infectious Disease 170:873);
- Scarlet fever $P_{symptoms}$ 0.8 (http://www.merck.com/mrkshared/mmanual/section13/chapter157/157a.jsp - accessed on September 5, 2005);
- Poliomyelitis $P_{symptoms}$ 0.05 (http://www.cdc.gov/nip/publications/pink/polio.pdf - accessed on September 5, 2005);
- Chicken pox $P_{symptoms}$ 0.9 (a quantitative estimate could not be found but expert opinion suggests that it is quite high);
- Measles $P_{symptoms}$ 1 (http://www.cdc.gov/nip/publications/pink/meas.pdf - accessed on September 5, 2005);
- Rubella $P_{symptoms}$ 0.5 (http://www.cdc.gov/nip/publications/pink/rubella.pdf - accessed on September 5, 2005);
- Mumps $P_{symptoms}$ 0.8 (http://www.cdc.gov/nip/publications/pink/mumps.pdf - accessed on September 5, 2005).