Estimating antiviral effectiveness against pandemic influenza using household data

Kathryn Glass* and Niels G. Becker

National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australian Capital Territory 0200, Australia

Current estimates of antiviral effectiveness for influenza are based on the existing strains of the virus. Should a pandemic strain emerge, strain-specific estimates will be required as early as possible to ensure that antiviral stockpiles are used optimally and to compare the benefits of using antivirals as prophylaxis or to treat cases. We present a method to measure antiviral effectiveness using early pandemic data on household outbreak sizes, including households that are provided with antivirals for prophylaxis and those provided with antivirals for treatment only. We can assess whether antiviral drugs have a significant impact on susceptibility or on infectivity with the data from approximately 200 to 500 households with a primary case. Fewer households will suffice if the data can be collected before case numbers become high, and estimates are more precise if the study includes data from prophylaxed households and households where no antivirals are provided. Rates of asymptomatic infection and the level of transmissibility of the virus do not affect the accuracy of these estimates greatly, but the pattern of infectivity in the individual strongly influences the estimate of the effect of antivirals on infectivity. An accurate characterization of the infectiousness profile—directed by strain-specific data—is essential for measuring antiviral effectiveness.

Keywords: influenza; antiviral drugs; estimation; transmission model; pandemic

1. INTRODUCTION

In the event of an influenza pandemic, antiviral drugs such as oseltamivir and zanamivir will be used to reduce disease transmission. For the currently circulating strains of influenza, these drugs have been found to reduce the risk of infection in susceptible individuals (Hayden et al. 2000, 2004; Welliver et al. 2001; Monto et al. 2002; Jefferson et al. 2006), and to reduce the levels of virus shedding in infected individuals who are treated soon after symptom onset (Hayden et al. 1996, 1997, 1999; Nicholson et al. 2000; Treanor et al. 2000; Jefferson et al. 2006). Many countries have or will obtain a stockpile of antivirals for use in an influenza pandemic (Cheng 2005; Lett 2005; Esxeld 2006; Harrold et al. 2006) and mathematical models have been used to investigate the relative benefit of using antivirals for prophylaxis and treatment (Ferguson et al. 2005; Longini et al. 2005; Barnes & Glass 2007; McCaw & McVernon 2007) using data for the currently circulating influenza. However, the effectiveness of antivirals for a pandemic strain is not known, and strain-specific estimates of antiviral effectiveness will be urgently required to assist policy makers in making the optimal use of the stockpile. During an influenza pandemic, health care workers will be very busy caring for patients and so will have limited time to collect extensive data. Here, we propose a method for estimating the parameters describing the effectiveness of antivirals using data on household outbreak size that are relatively easy to collect.

Most existing household studies that assess antivirals for prophylaxis observe contacts of the index case only (Hayden et al. 2000, 2004; Welliver et al. 2001; Monto et al. 2002). Under this study design, it is not possible to distinguish the effects of antivirals in reducing the infectivity of cases on antiviral from the effect of antivirals in reducing susceptibility of their prophylaxed contacts. However, to compare a strategy of using antivirals for treatment with one of using them for prophylaxis, the estimates of both effects are needed. One recent approach to providing both estimates (Halloran et al. 2006) combines studies, but it is difficult to ensure consistency across studies, and the authors note the need to improve the study design in the future. By incorporating transmission into our model, we are able to include information from all individuals exposed to household cases, and so distinguish the effects of antivirals on infectivity and on susceptibility within a single study.

Data from antiviral trials indicate that the infectivity of influenza cases varies over the course of the infectious period, with a fairly early peak in shedding (Hayden et al. 1996, 1999). Thus, the potential benefit
of antivirals to reduce transmission is strongly influ-
enced by the time that they are administered. These
findings are reflected in the recommendations that
antivirals will only be provided to individuals who are
diagnosed within 2 days of symptom onset (Lett 2005;
Harrod et al. 2006). In the clinical trials of antiviral
drugs, changes in infectivity over the infectious period
are controlled by ensuring that study participants are
given antivirals at a fixed time following exposure. This
study design is likely to pose ethical and practical
difficulties during an influenza pandemic. In the
methods presented here, we assume that individuals
may be provided with antivirals prior to infection or at
various times following infection, and explicitly model
the effect of timing on the impact of the drugs. We
investigate the effect of different assumptions about
infectivity on our estimates, and determine the house-
hold data that are most useful for obtaining good
estimates of antiviral effectiveness.

2. METHODS

2.1. Transmission probabilities

We use the Reed–Frost assumptions (Bailey 1975) to
describe transmission within households. In the absence
of antivirals, \( \theta \) is the probability that an individual
escapes infection from an infected household member,
and \( s \) is the probability that an individual escapes
infection from outside the household over a period of
time equal to the mean generation interval. The escape
probability \( s \) applies for each generation of trans-
mission, so that if the household outbreak is observed to
last \( g \) generations, the probability of an individual
escaping infection from outside the households
throughout that period is \( s^g \).

2.2. Model of infectivity and the effect of
antivirals

We model changes in the infectiousness of individuals
over the course of the infectious period using a
deterministic birth–death process for the growth of the
virus population, with birth rate \( \lambda \) and death rate \( \mu \).

\[
\begin{align*}
  f(0) &= \frac{Y}{(\delta - \lambda)(1 - Z)} \\
  g(0) &= \sigma \\
  h(0) &= \sigma f(0) \\
  f(T_1) &= 1 - \frac{\delta X}{(\mu + \delta - \lambda)} \\
  g(T_1) &= 1 - (1 - \sigma)X \\
  h(T_1) &= 1 - X + \frac{\sigma(\mu - \lambda)}{\mu + \delta - \lambda}X \\
  f(T_1 + k) &= 1 + (f(T_1) - 1)e^{(\lambda - \mu)k} \\
  g(T_1 + k) &= 1 + (g(T_1) - 1)e^{(\lambda - \mu)k} \\
  h(T_1 + k) &= 1 + (h(T_1) - 1)e^{(\lambda - \mu)k}
\end{align*}
\]

The death rate applies once the immune system
becomes active, \( T_1 \) days after infection. We assume
that \( T_1 \) is also the time of the onset of symptoms, which
is broadly consistent with the data on viral titre and
symptom scores in clinical trials (Hayden et al. 1996).
The effect of antivirals is to introduce an additional
death rate \( \delta \) at the time \( (T_\lambda) \) they are administered.
The solid curve in figure 1 gives an example of the
infectiousness function under this model. Although a
smoother peak could be obtained with more para-
eters, a comparison of this curve with shedding data
(Hayden et al. 1996, 1999) confirms that this simple
model reflects shedding patterns fairly well. The use of
fewer parameters is attractive because it tends to
produce more precise parameter estimates.

We consider the following five possible times at
which the antivirals are administered relative to the
infection of the primary case:

(i) antivirals given prior to infection of the primary
case: \( T_\lambda = 0 \),
(ii) antivirals given upon onset of symptoms in the
primary case: \( T_\lambda = T_1 \),
(iii) antivirals given 1 day after the onset of
symptoms in the primary case: \( T_\lambda = T_1 + 1 \),
(iv) antivirals given 2 days after the onset of
symptoms in the primary case: \( T_\lambda = T_1 + 2 \), and
(v) antivirals not provided.

Figure 1 illustrates the effect of antivirals on
infectiousness for three of these scenarios.

We model the effect of antivirals on susceptibility
by assuming that an individual’s probability of
escaping infection during a single contact is changed
by a factor \( \sigma \) while on antivirals. Antivirals have no
effect on susceptibility if \( \sigma = 1 \) and antivirals prevent
the infection with certainty if \( \sigma = 0 \). The probability
that an individual who is continuously on antivirals is
not infected by a specific household case is \( \theta^\sigma \). The
inclusion of \( \sigma \) in the exponent can be understood if we
consider \( \theta \) to be of the form \( \theta = \exp(\beta I^g) \) (infectivity
function), where the protective effect of the antivirals
acts multiplicatively on \( \beta \). Similarly, the probability
that an individual on antivirals escapes infection from
outside the household during one generation time is \( s^\sigma \).

For simplicity, we assume that \( s \) does not depend on
the calendar time that antivirals are provided to the
individual. We discuss alternative assumptions about
transmission from outside the household below.

Table 1 shows the probability that there is only one
case in a household of size 2 (given a primary household
case), for different interventions, where \( f(t) \), \( g(t) \) and
\( h(t) \) are given by

\[
X = \frac{\lambda e^{\lambda T_1}}{\mu e^{\mu T_1} - \mu + \lambda}, \quad Y = \frac{\lambda(\mu - \lambda)}{\mu e^{\mu T_1} - \mu + \lambda}, \quad Z = \frac{\mu e^{(\lambda - \mu)T_1}}{\mu + \delta - \lambda}, \quad k = 1, 2.
\]
Antiviral effectiveness during a pandemic K. Glass and N. G. Becker 697

Figure 1. (a–c) Infectiousness over time with (dashed curves: (a) intervention prior to infection, (b) intervention upon symptom onset, (c) intervention at time $T_A$) and without (solid curves) antivirals. The plots show the effect of providing antivirals at different times in an individual’s infectious period on their infectiousness.

Table 1. The probability that there is only one case in a two-person household with a primary case under different intervention scenarios. (Parameters and functions are specified in table 2.)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Antivirals</th>
<th>Antivirals</th>
<th>Antivirals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time from infection to intervention</td>
<td>$\theta^{(I)s}$</td>
<td>$\theta^{(I)s}$</td>
<td>$\theta^{(I)s}$</td>
</tr>
<tr>
<td>No intervention</td>
<td>$\theta$</td>
<td>$\theta$</td>
<td>$\theta$</td>
</tr>
</tbody>
</table>

Note that $f_1 = f(T_1)$ is the fraction of infectiousness experienced by an individual given antivirals at symptom onset, relative to an individual not given antivirals. Alternatively, this can be interpreted as the ratio of the area under the infectiousness function of an individual given antivirals at symptom onset to the area under the infectiousness function of an individual not given antivirals. As the parameter $f_1$ reflects the impact of antivirals on transmission more directly than $\delta$, we will use $f_1$ to describe the effect of antivirals on infectiousness. A summary of all parameters in the model is given in table 2, including the values assumed for simulations. We consider two sets of values for the transmission parameters to compare relatively high and low levels of transmission. Under the low transmission rates in the absence of antivirals, there is an average of two secondary cases in a household of size 6, whereas there are 3.5 secondary cases on average under high transmission rates.

2.3. Estimating unknown parameters

We expand the two-person household scenario described above to consider larger households under two possible intervention scenarios.

— Scenario A: prophylaxis and treatment in households. Antivirals are given to the entire household $T_A$ days after infection of the primary case.
— Scenario B: treatment only in households. Antivirals are given to the primary case $T_A$ days after infection, and to each subsequent case in the household upon onset of their symptoms.

Using the above formulae for $f(t)$, $g(t)$ and $h(t)$, we can write down a likelihood function for this model, assuming that the data include household size, household outbreak size, timing of introduction of the antivirals relative to the infection of the primary case and intervention scenario as above (see the appendix in the electronic supplementary material for more details). That is, we use the final outbreak size in the household, but do not require data on the time that secondary cases become infected, nor data on who infected whom within the household. We observe households until there is a generation in which there are no new cases. In practice, there may be some difficulties translating this into calendar time. However, in our study, the aim is to determine how much data are needed for useful inferences, and for this aim the assumed setting is convenient and informative. This is confirmed by calculations for an alternative period of observation of five generation times.

We find maximum-likelihood estimates for the parameters $f_1$, $\sigma$, $\theta$ and $s$, under the case where $\lambda$, $\mu$ and $T_1$ are known, and also considering the case where all parameters are estimated simultaneously. All confidence intervals for parameters, calculated using profile likelihoods, are 95 per cent confidence intervals. Parameter values are assigned to $\lambda$, $\mu$ and $T_1$ so that the infectiousness function reflects observed data on virus shedding.

To assess the estimates of the parameters $f_1$, $\sigma$, $\theta$ and $s$, we compute their estimates corresponding to the data generated from the above model. We exclude single-person households and those with more than six members, assuming that the remaining households are distributed in accordance with the 2001 Australian census data. That is, 44 per cent are of size 2, 21 per cent are of size 3, 21 per cent are of size 4, 10 per cent are of size 5 and 4 per cent are of size 6.

2.4. Alternative assumptions and sensitivity analysis

2.4.1. Alternative period of household observation. Under our baseline framework, we observe all households from the primary case until there is a generation in which there are no cases. Thus, a household with only the primary case is observed for one generation only, while a household with many secondary cases may be observed for up to five generations, depending on the chain of infections. We then adjust for these different observation periods in the likelihood function. In order to confirm that this method does not influence our results, we considered an alternative framework, in which all households are observed for five generations, regardless of the household size or the number of cases.
The likelihood function (see the appendix in the electronic supplementary material) is considerably more complicated under this model, as the probability of infection from outside the household must be taken into account for a full five generations, regardless of the chain of infections.

### 2.4.2. Alternative model of infectiousness

We test the sensitivity of the parameter estimates to our model of infectiousness by generating data from an alternative model in which infectiousness is constant over the infectious period. This ‘constant infectiousness’ model assumes that individuals are not infectious on day 1 or after day 4, and are equally infectious on days 2–4, with infectious period. This ‘constant infectiousness’ model in which infectiousness is constant over the infectious period. This effect on infectivity is to reduce infectiousness by a factor of 1/3.

We test the sensitivity of our estimates to some level of asymptomatic transmission by generating data in which a third of infectiousness is reduced by this alternative model, where

\[ f(T) = \begin{cases} 1, & \text{if } T > 0 \\ \frac{1}{3}, & \text{if } T = 0 \\ 0, & \text{if } T < 0 \end{cases} \]

is the day of the first household case, relative to the day of the primary case.

### 2.4.3. Sensitivity to asymptomatic cases

We test the sensitivity of our estimates to some level of asymptomatic infection by generating data in which a third of infectiousness is reduced by this alternative model, where

\[ f(T) = \begin{cases} 1, & \text{if } T > 0 \\ \frac{1}{3}, & \text{if } T = 0 \\ 0, & \text{if } T < 0 \end{cases} \]

is the day of the first household case, relative to the day of the primary case.

### 2.4.4. Alternative models of transmission from outside the household

Our baseline model of transmission assumes that over each generation of infection, a susceptible individual has a probability \( s \) of escaping infection from outside the household. In order for infection rates to be realistic over the course of a population outbreak, \( s \) must be approximately constant over time. We consider two alternative models: one in which there is no transmission from outside the household during the household outbreak, and another in which transmission from outside the household increases while data are collected, to reflect a growing epidemic.

Most household outbreaks will include at most two generations after the primary case. If the force of infection from outside the household is small—e.g. in the very early stages of an outbreak—then the chance of a household member becoming infected from outside the household over these generations is negligible. By eliminating the parameter \( s \) from the model, we may be able to obtain tighter bounds on the other parameters.

On the other hand, if case numbers increase considerably over the study period, individuals will be more likely to be infected from outside the household as time goes on. To model this, we replace \( s \) with \( s_0 \exp(-\gamma d) \), where \( \gamma \) is determined by the increase in case numbers, which can be estimated separately, and \( d \) is the day of the first household case, relative to the start of the outbreak. This model is likely to be more appropriate if the household data are collected over an extended period of time, as it takes account of the increase in the force of infection acting on the household as the outbreak progresses.

### 2.4.5. Alternative assumption about mixing within households

The original likelihood function assumes that the within-household transmission rate between individuals does not vary with household size. We compare this model with one in which the escape probability varies according to household size as \( \theta^{1/(n-1)} \), where \( n \) is the household size. This formulation would arise if we assumed that every household member had a fixed number of contacts and distributed them evenly among the household members.
3. RESULTS

3.1. Impact of antiviral distribution strategies

We want to compare the precision of the parameter estimates under two different antiviral distribution strategies: one that includes prophylaxis, and the other that only includes treated cases. The plots in figure 2 show these estimates for a simulated dataset with 500 households under true parameter values that describe relatively low transmission. In each plot, the circles and lines show the estimates and confidence intervals, while the crosses show the true values. Figure 2a is for the prophylaxis and treatment strategy, figure 2b is for the treatment-only strategy and figure 2c assumes that half of the households were given antivirals according to each strategy. Each part of figure 2 illustrates estimates for a single randomly selected dataset. This allows us to demonstrate the width of the confidence intervals. Repetitions produced similar results. We see that we can get tighter bounds on the estimates of all parameters if the study includes households with both types of intervention. In the case of figure 2b, antivirals are not given to uninfected household members, so it is not possible to estimate their effect on susceptibility. Repeating these calculations using high transmission rate parameters, we find that the estimates of $s$ and $\theta$ become a little less precise, while the estimates of $\sigma$ and $f_T$ become more precise. It is still preferable to have data from both treatment-only and prophylaxis and treatment households.

3.2. Impact of antiviral timing strategies

The example in figure 2 assumes 100 households of each type, where antivirals are equally likely to be provided to the primary case at the five possible times, namely

(i) prior to infection of any member,
(ii) upon onset of symptoms of the primary case,
(iii) 1 day after symptom onset of the primary case,
(iv) 2 days after symptom onset of the primary case, and
(v) antivirals not given at all.

Antivirals are also provided to the remainder of the household under the prophylaxis and treatment scenario. In figure 3, we compare the parameter estimates under different combinations of antiviral distribution times, assuming low transmission rates, as follows:

— figure 3a includes an equal number of households of the above five types,
— figure 3b excludes households of type (i) and has an equal number of the remaining types,
— figure 3c excludes households of type (v) and has an equal number of the remaining types, and
— figure 3d excludes households of types (iii) and (iv) and has an equal number of the remaining types.

Each example has the same total number of households.

We see that the estimates of the effect of antivirals on susceptibility ($\sigma$) and infectivity ($f_T$) are more precise if the data include some households of types (i) and (v)—that is, households in which the primary case was taking antivirals as prophylaxis, and households that do not receive any antivirals. Data from the households where antivirals are provided 1 or 2 days after symptoms contribute less information, and slightly tighter bounds are possible if antivirals are provided early, or not at all. These results remain true if the transmission rates are higher.

3.3. Data requirements

In the early stages of a pandemic, when only minimal data are available, the priority will be to assess whether antivirals have a beneficial impact on susceptibility and infectiousness. That is, we will want to know whether either or both of $\sigma$ (measuring antiviral impact on susceptibility) and $f_T$ (measuring antiviral impact on infectiousness) are significantly less than 1. For figure 4, we simulated 100 datasets and show the percentage of simulations in which the confidence interval does not include 1 for each parameter, for both low and high transmission rates.

In general, it is slightly easier to distinguish an effect of antivirals if transmission rates are high, as this creates more exposures within households. There is a good chance that an effect on infectiousness and susceptibility such as that assumed here can be demonstrated if data on 200–500 households with a primary case have been collected.
3.4. Alternative period of household observation

We have assumed that a household is observed until there are no cases in a generation. This observation period is attractive because it captures most within-household transmissions with a relatively short observation period and computation is relatively shorter because expressions for the chain probabilities are simpler. However, there may be practical difficulties with this stopping rule. To assess this concern, the results were compared with those for the scenario where all households are observed for a fixed time period equal to five generation times. We found no appreciable difference in estimates or confidence intervals for the antiviral parameters. In other words, recommendations on study size would be the same.

3.5. Alternative model of infectiousness

Our analysis assumes an underlying model of infectiousness with a fairly sharp rise and fall in infectivity (figure 1). We investigate the sensitivity of our estimates to this model by generating data using an extreme alternative in which infectiousness is constant over a 3-day infectious period. We generate the data using this flat infectiousness model, and then estimate the parameters $f_T$, $\sigma$, $\theta$ and $s$, where $f_T$ under each model is the fraction of infectiousness experienced by an individual put on antivirals at symptom onset when compared with an individual not given antivirals.

Table 3 shows the percentage of 100 trials in which the confidence intervals for the parameters contained the true value when the data were generated using the varying infectiousness model, compared with the case when the data were generated using a model in which infectiousness was constant over the infectious period. In both cases, the analysis assumes the varying infectious model, which allows us to test the impact of this assumption if the data do not conform to this model. Under the constant infectiousness model, the confidence intervals for $\sigma$ (the impact of antivirals on

Figure 3. Estimates of parameters $f_T$ (effect of antivirals on infectivity), $\sigma$ (effect of antivirals on susceptibility), $\theta$ (quantifies within-household transmission) and $s$ (quantifies transmission from outside the household) for four different intervention strategies of a study comprising 500 households. In (a), all intervention times are included equally; in (b), none of the primary cases were given antivirals prior to infection; in (c), there were no households included in which antivirals were not provided; and in (d), individuals were either given antivirals at (or before) symptom onset or not at all. In each case, the circles and lines show the estimates and confidence intervals, while the crosses show the true values. Parameters used to generate the data were: $\lambda=4$, $\mu=5$, $T_1=2$, $f_T=0.73$, $\sigma=0.5$, $\theta=0.7$ and $s=0.98$, which correspond to fairly low transmission rates.

Figure 4. The percentage of 100 trials in which the computed 95% confidence interval did not include 1 for each parameter for (a) low and (b) high transmission rates, and for increasing numbers of households (100–1000). Black bars, 100; dark grey bars, 200; light grey bars, 500; white bars, 1000. Parameters used to generate the data were: $\lambda=4$, $\mu=5$, $T_1=2$, $f_T=0.73$, $\sigma=0.5$, $\theta=0.7$ and $s=0.98$, which correspond to fairly low transmission rates.
Table 3. The effect of assumptions about infectiousness on parameter estimates. (The table compares the percentage of 100 trials in which the 95% confidence interval contained the true value when the data were generated using the underlying model with varying infectiousness, and when the data were generated using an alternative model where infectiousness is constant over the infectious period. Parameters $f_T$ and $\sigma$ measure the effect of antivirals on infectivity and susceptibility (respectively), while $\theta$ and $s$ are the escape probabilities for infection from within and outside the household. Parameters used to generate the data were: $\lambda=4$, $\mu=5$, $T_4=2$, $f_T=0.73$, $\sigma=0.5$, $\theta=0.7$ and $s=0.98$.)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>100 households</th>
<th>200 households</th>
<th>500 households</th>
<th>1000 households</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_T$</td>
<td>100</td>
<td>98</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>96</td>
<td>95</td>
<td>93</td>
<td>97</td>
</tr>
<tr>
<td>$\theta$</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>$s$</td>
<td>100</td>
<td>98</td>
<td>99</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 3. The effect of assumptions about infectiousness on parameter estimates. (The table compares the percentage of 100 trials in which the 95% confidence interval contained the true value when the data were generated using the underlying model with varying infectiousness, and when the data were generated using an alternative model where infectiousness is constant over the infectious period. Parameters $f_T$ and $\sigma$ measure the effect of antivirals on infectivity and susceptibility (respectively), while $\theta$ and $s$ are the escape probabilities for infection from within and outside the household. Parameters used to generate the data were: $\lambda=4$, $\mu=5$, $T_4=2$, $f_T=0.73$, $\sigma=0.5$, $\theta=0.7$ and $s=0.98$.)

<table>
<thead>
<tr>
<th>data simulated assuming varying infectiousness</th>
<th>100 households</th>
<th>200 households</th>
<th>500 households</th>
<th>1000 households</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_T$</td>
<td>100</td>
<td>98</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>96</td>
<td>95</td>
<td>93</td>
<td>97</td>
</tr>
<tr>
<td>$\theta$</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>$s$</td>
<td>100</td>
<td>98</td>
<td>99</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 3. The effect of assumptions about infectiousness on parameter estimates. (The table compares the percentage of 100 trials in which the 95% confidence interval contained the true value when the data were generated using the underlying model with varying infectiousness, and when the data were generated using an alternative model where infectiousness is constant over the infectious period. Parameters $f_T$ and $\sigma$ measure the effect of antivirals on infectivity and susceptibility (respectively), while $\theta$ and $s$ are the escape probabilities for infection from within and outside the household. Parameters used to generate the data were: $\lambda=4$, $\mu=5$, $T_4=2$, $f_T=0.73$, $\sigma=0.5$, $\theta=0.7$ and $s=0.98$.)

<table>
<thead>
<tr>
<th>data simulated assuming constant infectiousness</th>
<th>100 households</th>
<th>200 households</th>
<th>500 households</th>
<th>1000 households</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_T$</td>
<td>93</td>
<td>85</td>
<td>57</td>
<td>28</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>98</td>
<td>97</td>
<td>92</td>
<td>88</td>
</tr>
<tr>
<td>$\theta$</td>
<td>97</td>
<td>99</td>
<td>88</td>
<td>74</td>
</tr>
<tr>
<td>$s$</td>
<td>95</td>
<td>95</td>
<td>89</td>
<td>82</td>
</tr>
</tbody>
</table>

susceptibility), $\theta$ (the within-household transmission parameter) and $s$ (the between-household transmission parameter) contain the true value in approximately 90 per cent of the trials for up to 1000 households, but the estimate of $f_T$ (the impact of antivirals on infectiousness) is much less reliable, owing to the large differences in the infectiousness functions under the two models. In order to get precise and accurate estimates of the effect of antivirals on infectiousness, the model needs to reflect the pattern of infectiousness in the individual fairly accurately.

In light of this result, we investigated the potential to estimate the three parameters defining the infectiousness function in addition to the four parameters of the model using these household data. When we try to estimate all seven parameters in the model simultaneously, more households are needed to obtain precise estimates. For example, once all parameters are included, approximately 1000 households are needed to assess whether antivirals are having an impact on susceptibility, in comparison with approximately 500 when the infectiousness function parameters are fixed. Confidence intervals for the three additional parameters are still wide with 1000 households.

3.6. Impact of asymptomatic cases

A further test of model assumptions was to assess the impact of undetected asymptomatic cases in the household outbreaks. We generated the data under an alternative model in which 33 per cent of cases were asymptomatic, considering both the scenario where asymptomatic cases are as infectious as symptomatic cases, and the scenario where they are not infectious at all. We find inclusion of asymptomatic cases has a large effect on the estimation of $\theta$, which captures the within-household transmission rate, but has no noticeable effect on the accuracy of other parameter estimates. We find that the precision of the estimate of the effect on susceptibility ($\sigma$) is slightly decreased by the presence of asymptomatic cases, but otherwise, the estimates of parameters measuring antiviral effectiveness do not seem to be affected.

3.7. Alternative model of transmission from outside the household

If it is possible to collect data from households in the very early stages of the outbreak, we can adopt a simpler model of transmission that ignores the impact of transmission from outside the household. That is, once there is a primary case in a household, we assume that the force of infection acting on individuals from outside the household is negligible for the duration of the household outbreak. Under this assumption, we fix $s=1$, and so the number of parameters to be estimated is reduced to three. Most of the results presented above continue to hold for this simpler model, except that estimates of $f_T$ and $\theta$ are considerably more precise, and the estimates of $\sigma$ are a little more precise. In particular, 100–200 households are sufficient to detect an effect of antivirals on infectivity, although at least 200 households are still needed to detect an effect of antivirals on susceptibility.

By contrast, if the household data are collected over an extended period of time during which the current case numbers increase considerably, then it may be necessary to take account of the changing force of infection acting on the household. This model requires us to estimate the increase in case numbers separately, but otherwise, the results remain similar under these assumptions, with the estimates of $f_T$, $\theta$ and $\sigma$ all becoming slightly more precise under this model.

3.8. Alternative assumption about within-household mixing

We tested the effect of an alternative assumption in which the escape probability varies according to the household size $\theta^{1/(n-1)}$, where $n$ is the household size. Reproducing figure 4 with the corresponding new likelihood function, we find that slightly more data are required to test the impact of antivirals on infectivity $f_T$, while the other estimates are similar. When the infectivity function is changed to a constant function (as shown in table 3 for the original likelihood function), $f_T$ is still unlikely to be estimated accurately, but the other parameters are slightly more likely to be unaffected by the change to the infectivity function. Overall, the effect of this change in assumptions is fairly minor.

4. DISCUSSION

This paper shows that the data on household outbreak size can be used to estimate the effect of antiviral drugs in the early stages of an influenza outbreak. Even with relatively low levels of transmission within households, it is possible to confirm that antivirals are having a significant effect using the data from 200 to 500 households. If these data can be collected very
early in the outbreak, only 100–200 households are needed to detect whether antivirals are reducing the infectiousness of cases. In the event of an influenza pandemic, there will be limited resources for collecting data, as health workers will have many claims on their time. Our methods do not require extensive data to be collected from households, and so will not place a large burden on health care workers.

We considered a range of timing and distribution strategies to identify those that provide the most informative data for estimating antiviral effects. We find that it is useful to have a range of intervention scenarios in the study, including some households where antivirals are given as prophylaxis, as well as some households in which antivirals are only given as treatment, and some in which no antivirals are provided. It is plausible that prophylaxis may be provided to households in which one family member is at a high risk of exposure to infection—say, a health care worker. Provided they are the only family members at high risk and they are the primary cases, these households would provide valuable data for a study of antiviral effectiveness. There are also likely to be households in which antivirals are not provided because the primary case was not diagnosed sufficiently early. Including such households is also of great benefit.

In addition to the baseline transmission model, we considered the effect of alternative assumptions about transmission from outside the household and the infectiousness of individuals on the accuracy and precision of parameter estimates. If it is possible to collect very early outbreak data, a simpler model applies which increases the precision of the parameter estimates. If the data are collected over an extended period of time, then it may be necessary to adjust for the increasing force of infection acting on households. This requires slightly more data to be collected, but otherwise has little impact on the results presented here. Inclusion of asymptomatic cases has little effect on the accuracy of the estimates of parameters measuring antiviral efficacy, which is reassuring, as it may be difficult to measure the rate of asymptomatic infections in the early stages of the outbreak.

By contrast, incorrect assumptions about the infectiousness of individuals over their infectious period can lead to very inaccurate estimates of the effect of antivirals on infectivity. Shedding data from antiviral trials suggest an early peak in infectiousness and it is crucial that this is taken into account when measuring antiviral effectiveness for reducing infectivity. This is particularly important in the event of an influenza pandemic, when the timing of the administration of antivirals cannot generally be predetermined. A good understanding of the infectiousness profile of influenza cases is also needed for a number of intervention strategies, such as isolation of cases, so it is clearly a high priority to measure this for a new pandemic strain. With proper surge capacity to collect data on household outbreaks as part of the process for distributing antivirals and with the assumption that infectiousness within pandemic flu cases evolves as for seasonal flu, one can apply these methods to obtain useful estimates of antiviral effectiveness with the data on approximately 200 household outbreaks. For estimates that are robust against variations from the properties of seasonal flu, the data from approximately 1000 household outbreaks are needed. This remains feasible, particularly when the data from different jurisdictions are pooled. These methods use the data on household outbreak size, and do not require the data to be collected on who infected whom, or detailed data on the time that secondary cases develop in the household. This is a deliberate restriction to ensure that the task of data collection is not burdensome. However, if there were sufficient resources to gather data on the timing of secondary cases, it seems likely that both the antiviral impact and the shape of the infectiousness function could be estimated using fewer households. We are currently investigating the potential of this approach in comparison with a method of using linked pairs of cases to build a model of the infectiousness function. Together with the methods presented here, these techniques will allow us to estimate the impact of a number of pandemic influenza control measures from early outbreak data.

The authors gratefully acknowledge financial support from the Australian NHMRC Capacity Building grants 358425 and 224215 and the Australian Research Council grant DP0558357.

**REFERENCES**


