Identifying the Probable Timing and Setting of Respiratory Virus Infections

Justin Lessler, PhD; Ron Brookmeyer, PhD; Nicholas G. Reich, BS; Kenrad E. Nelson, MD; Derek A. T. Cummings, PhD; Trish M. Perl, MD

OBJECTIVE. Show how detailed incubation period estimates can be used to identify and investigate potential healthcare-associated infections and dangerous diseases.

METHODS. We used the incubation period of 9 respiratory viruses to derive decision rules for distinguishing between community- and hospital-acquired infection. We developed a method, implemented in a simple spreadsheet, that can be used to investigate the exposure history of an individual patient and more specifically to identify the probable time and location of infection. Illustrative examples are used to explain and evaluate this technique.

RESULTS. If the risks of hospital and community infection are equal, 95% of patients who develop symptoms of adenovirus infection within 5 days of hospital admission will have been infected in the community, as will 95% of patients who develop symptoms within 3 days for human-coronavirus infection, 2.5 days for severe acute respiratory syndrome, 1 day for influenza A, 0.5 day for influenza B, 12 days for measles, 2 days for parainfluenza, 4 days for respiratory syncytial virus infection, and 1.5 days for rhinovirus infection. Sources of infection suggested by analysis of the symptom onset times of individual patients are consistent with those from detailed investigations.

CONCLUSIONS. This work shows how a detailed understanding of the incubation period can be an effective tool for identifying the source of infection, ultimately ensuring patient safety.

Infect Control Hosp Epidemiol 2010; 31(8):809-815

Respiratory infections are a major source of morbidity and mortality among healthcare workers and patients. Influenza (including the recent pandemic H1N1 [2009]), respiratory syncytial virus, severe acute respiratory syndrome (SARS), and other respiratory viruses have circulated in healthcare settings among vulnerable patients and nonimmune staff, leading to adverse outcomes and disruption of healthcare services. To prevent and to control transmission in healthcare settings, we often want to know where infection occurred. For instance, when a patient develops influenza symptoms in the hospital, we want to know whether he or she was infected before or after admission, so that in the latter case potential sources of nosocomial transmission can be identified and can be eliminated. Multiple US states now require the public reporting of healthcare-associated infections; therefore, correct attribution of the source of infection has taken on new importance to many healthcare institutions. Often, the primary piece of data we have to identify the probable source of infection is the patient’s symptom onset time. This information, combined with knowledge of the distribution of the incubation period of the offending pathogen, can be used to select and to estimate the performance of classification rules for identifying healthcare-associated infections. In other words, by using the time of symptom onset, we can classify an infection as associated with a community or a healthcare exposure. In this article, these methods are applied using incubation period estimates for 9 respiratory viruses to provide specific recommendations for decision rules to distinguish cases infected in a healthcare or other institution from those infected elsewhere. We also present a simple method for analyzing the exposure history of any particular case, even one with multiple periods of exposure to different settings, and we determine the probability of infection having occurred in a specific period. This latter work may prove useful in areas outside of hospital infection prevention and control, such as prioritizing the investigation of possible sources of infection in outbreaks such as the present epidemic of pandemic H1N1 influenza (2009).

METHODS

General Decision Rules

The general method for distinguishing hospital- from community-acquired transmission and/or infection is to set a disease-specific cut off, such as 48 hours after admission, and
to consider patients who develop symptoms within this interval to have been infected in the community and those who develop symptoms after this interval to have been infected in the hospital. If we know accurately the distribution of the incubation period, we can use this information to estimate the performance of this approach in terms of the percentage of classifications that are correct (ie, the positive predictive value and negative predictive value). The correct cut-off to use depends not only on the incubation period but also on which classification is more important. For example, if our priority is correctly classifying hospital-acquired infections as much as possible, and if the incorrect classification of community-acquired infections is less important, we would want to use an earlier cut-off. To perform this analysis, it is necessary to know the incubation period in greater detail than it is typically reported, a problem we addressed for 9 respiratory viruses (adenovirus, human-coronavirus, SARS, influenza A, influenza B, measles, parainfluenza, respiratory syncytial virus, and rhinovirus) in previous work. On the basis of these estimates, we used the methods described in Lessler et al to make recommendations with regard to these respiratory viruses.

Investigation of Individual Patients

Often the history of an individual patient involves multiple hospital admissions and discharges in the time period when infection could have occurred. Also, we may want to identify the source of infection in other contexts, such as when an individual presents with a rare or dangerous infectious disease. If the individual has recently traveled to multiple locations or has several possible sources of exposure, we may want to determine where or how they were most likely infected. To identify the probable settings where infection occurred, we can consider the patient's symptom onset time and history of exposure. Although analyzing an individual patient's exposure history may be logistically harder than applying a simple classification rule, mathematically it can be performed using a simple spreadsheet when disease incidence is low (Table 1).

To estimate the probability that a case was infected during each period of potential infection, we need to know the following pieces of information as accurately as possible: (1) the patient’s time of symptom onset, (2) the time that each period began, and (3) the distribution of the incubation period. In situations in which there may be large variations in risk between periods (eg, time spent in areas where measles is endemic versus time spent in areas where measles is not endemic), we also may want to include an estimate of the relative risk of infection. For acute respiratory viruses, incubation period distribution is approximately log-normal and can be specified by a median and dispersion.7,9

In our method, a weight is calculated for each period by multiplying the probability mass of the incubation period distribution contained in that period by the relative infection risk in that setting (Table 1). By dividing the weight for each period by the sum of all weights, we obtain an estimate of the probability of infection during that period, given the patient’s time of symptom onset and the relative risk of infection. These probabilities can be used to identify the probable source of infection or to prioritize further investigations.

For illustrative purposes, we present the application of this technique to the investigation of a case of pandemic H1N1 (2009) in Johns Hopkins Hospital (Baltimore, MD; example 1) and to 2 published case investigations (examples 2 and 3). Details on these investigations can be obtained by contacting the corresponding author or consulting the original publications.

RESULTS

General Decision Rules

The appropriate point at which to set the cut-off between community- and hospital-acquired infection varies widely, from a half-day for influenza B to 13 days for measles (Table 2). If our intention is to control the spread of disease in the hospital, we may wish to prioritize the identification of hospital-acquired infections (ie, to ensure that a designation of “community-acquired” is more often correct). In this situation, the second column of Table 2 shows the optimal cut-offs, which are generally from one-half day to 1 day earlier. These rules assume that the risk for infection is similar in the community and the hospital. If this is not the case (eg, in the context of a hospital-based outbreak), we recommend

<table>
<thead>
<tr>
<th>Table 1. Mathematical Formulae for Calculating the Probability That Infection Occurred in a Given Period (p_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>\vdots</td>
</tr>
<tr>
<td>n-1</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

**NOTE.** The analyst needs to supply the case patient's symptom onset time (s), the time that each period 1...n began (t_i), and the incident rate ratio for infection for the setting that the case patient was in during that period (r_i). These values, combined with the disease-specific incubation period distribution, F(x), are used to calculate the minimum incubation period for that period (x_i), the probability mass of the incubation period distribution contained in that period (h_i), a period-specific weight (w_i), and the probability of infection during that period (p_i). This method can be easily implemented in a simple spreadsheet (see examples 1–3 in Results).
that practitioners set situation-specific decision rules by using the methods and tools presented in Lessler et al.  

**Individual Investigations**

Techniques similar to those used for developing general decision rules can be used to investigate the timing and setting in which an individual patient was infected. Here we know the approximate time of symptom onset for the individual patient. Thus, we can more precisely estimate the probability of that patient being infected in different time periods. This contrasts with the general rules above, which must be appropriate for patients developing symptoms at any time during the “community-acquired” or “hospital-acquired” periods.

We illustrate the use of our methods for investigating individual infections by using a simple spreadsheet with 3 examples that are adapted from actual patient investigations.

---

**TABLE 2.** Length of Stay in a Hospital (or Other Institution) beyond Which a Patient Developing Symptoms Should Be Considered to Have Been Infected in the Hospital

<table>
<thead>
<tr>
<th>Infection</th>
<th>Balanced Length of stay, days</th>
<th>&lt;5% Community-acquired infections falsely classified</th>
<th>&lt;5% Hospital-acquired infections falsely classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenovirus</td>
<td>6.0</td>
<td>5.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Coronavirus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human-coronavirus</td>
<td>3.0</td>
<td>3.0</td>
<td>3.5</td>
</tr>
<tr>
<td>SARS-associated coronavirus</td>
<td>5.0</td>
<td>2.5</td>
<td>9.5</td>
</tr>
<tr>
<td>Influenza</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flu A*</td>
<td>1.5</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Flu B</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Measles</td>
<td>13.0</td>
<td>12.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Parainfluenza</td>
<td>2.5</td>
<td>2.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Respiratory syncytial virus</td>
<td>4.5</td>
<td>4.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Rhinovirus</td>
<td>2.0</td>
<td>1.5</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**NOTE.** Times are listed to the nearest half-day. Table shows criteria that would misclassify an equal number of community-acquired and hospital-acquired infections (“balanced”), criteria at which less than 5% of infections classified as community acquired would actually be hospital acquired, and criteria at which less than 5% of infections classified as hospital acquired would actually be community acquired. SARS, severe acute respiratory syndrome.

* Includes pandemic H1N1 (2009).

---

**FIGURE 1.** Example 1. A triage clerk developed pandemic H1N1 (2009) after exposure to an incoming patient. The fields that are shaded represent fields in which the user must put observed or assumed information about the case in question. The case and disease fields are provided for administrative tracking purposes. The incubation period parameters of median and dispersion identify the log-normal distribution used in the table’s calculations. The user specifies the number of time periods in which infection may have occurred, the start times of each of these periods, and a name for each of these periods. The user also specifies the observed time of symptom onset. Finally, the user specifies an incident rate ratio (IRR) for each of the times on the basis of the best available information about the possible exposure settings. The table automatically calculates the minimum possible incubation period that is associated with this interval (minimum inc), the probability mass associated with this interval (P[mass]), the weight for this interval (weight), and the probability that this case was infected in the given period (P[infected in period]).
Although our technique was not used in the actual investigations, we present these vignettes to show how this technique can help to guide epidemiologic investigation into the source of infection and how the inferences from our approach compare with the results of detailed investigation. Example 1 (pandemic H1N1 [2009]) illustrates a case for which exposure risks are not available, example 2 (measles) illustrates a case for which we have data on relative incidence in exposure periods, and example 3 (SARS) illustrates a case for which there are definitive periods of possible exposure.

Example 1: a healthcare worker develops pandemic H1N1 (2009) after exposure to an incoming patient. A triage clerk developed fever, cough, myalgias, and fatigue late in the afternoon (approximately 4:00 PM) of May 28, 2009 (Figure 1). A nasopharyngeal aspirate was obtained on June 2 and gave positive test results for pandemic H1N1 (2009). An investigation was performed to identify the probable source of infection and who else may have been admitted to the hospital. On the basis of the incubation period of influenza A and the clerk’s time of symptom onset, we estimated that the most likely period of exposure was the clerk’s time away from the hospital between her May 26th shift and her May 27th shift (33.75% of the weighted probability), followed by her May 27th shift (24.6%) and the time off immediately following (23.59%).

Because we were mostly concerned about the probability of a hospital-associated infection, we prioritized the work shift on May 27th (7:00 AM–3:00 PM) for investigation. A review of the patients who were admitted during this period revealed that the triage clerk registered a patient with influenza-like symptoms during her May 27th shift at 1:00 PM who tested positive for influenza A. This patient’s infection was later proven to be pandemic H1N1 (2009), indicating that he was the most likely source of the infection. Infection in the periods immediately preceding and following this shift is equally consistent with the time of symptom onset and remains a possibility. However, pandemic H1N1 (2009) was rare in the community at this point.

Example 1 was adapted from an unpublished hospital-associated infection. Some details have been changed for the purposes of illustration.

Example 2: an Italian teenager develops measles after travel to the United Kingdom. A teenager visited the United Kingdom from Northern Italy on a study trip September 2–15, 2007 (Figure 2). After returning to Italy, the student developed fever on September 17 and developed rash on September 20. The next week, an epidemiologic investigation was initiated to determine the source of the infection and who else
might be at risk. The incidence of measles in Italy is 0.55 cases per 100,000 population, and the incidence in the United Kingdom is 1.68 cases per 100,000 population, 3.05 times that of Italy. On the basis of the incubation period of measles and the elevated risk of infection in the United Kingdom compared with the risk of infection in Italy, we estimated that there was a 93% chance that this student was infected during her trip to the United Kingdom (Figure 2). This supposition was supported by subsequent epidemiologic investigation that showed that measles was not circulating in this region of Northern Italy before the index case returned from the United Kingdom, and an outbreak occurred subsequent to her return. Example 2 was adapted from the index case in Filia et al.

Example 3: a healthcare worker develops SARS from an unknown exposure. A healthcare worker developed SARS symptoms on March 28, 2003, with no known occupational exposure to SARS (Figure 3). An occupational exposure was suspected; however, there were no known cases of SARS in the hospital. To conduct a more thorough examination of patients for potential inpatient SARS cases, we wished to identify the shift on which the worker was most likely exposed. He had worked 6 shifts between March 13, when SARS was introduced to Canada, and March 28, when he developed symptoms. On the basis of an interview with the patient, we determined that it was highly unlikely that he could have been exposed to SARS during his time off, so the incident rate ratio was set to 0 for the times off and 1 for the times in the hospital. On the basis of this information, we determined that the worker was most likely infected during the shifts he worked on March 24 and 25, with a lower but still significant probability of exposure on March 21. Further investigation revealed that a 77-year-old hemodialysis patient that he cared for on all 3 occasions fit the SARS case definition.

Example 3 was adapted from the report by Dwosh et al of patient 9, who was found to have occupational exposure on March 21, 24, and 25, 2003. Some specifics described are elaborations created for the purpose of illustration.

A template for performing this analysis (Excel; Microsoft), as well as the spreadsheets created in the preceding examples, is available in the online supplemental material (Appendix A).

DISCUSSION

Interpretations of the time of symptom onset in investigations of infectious diseases are necessarily based on the “known” incubation period. However, our estimates have implications for prophylaxis and treatment of exposed persons and for institutions reporting infections. In this article, we have presented 2 methods for discriminating between possible sources of infection with respiratory viruses and other infectious diseases. The first yields a definitive decision rule that can be used to make recommendations and to set policy. The second allows the detailed analysis of an individual case’s exposure history and yields a probability for infection in each setting.

There is a tension between these 2 approaches. The former provides rules that can be quickly and easily applied and is perhaps more appropriate for setting general policy and for situations in which time or resources are lacking. The latter is far more flexible, should lead to more accurate results, and is appropriate if cases have complex exposure histories or if accurate results are especially important. In some situations, it may be best to use these approaches together; for instance, setting a general decision rule that identifies likely healthcare-associated infections (eg, one that has a low error rate on those classified as community-acquired infections) and then following up the possible healthcare-associated infections with an individual analysis.

Both methods are only as good as the data inputs, and uncertainty about when the case patient was in different settings, the timing of symptom onset, and the incubation period of the disease can negatively impact the usefulness of the results. The dangers of making false calls may be mitigated by performing ad hoc sensitivity analyses and by considering multiple scenarios for an individual patient by use of other plausible values for exposure times, symptom onset time, or incubation period. Unfortunately, full distributional estimates on the incubation period are not available for many diseases other than those listed here (estimates for a few others are given in Sartwell12). We hope that the existence of methods such as these will motivate more such estimates, thereby increasing the usefulness of these approaches.

The use of definitive decision rules and probabilistic aids to investigation could be of immense utility in healthcare epidemiology. Presently, the National Healthcare Safety Network does not propose any time cut-off for defining healthcare-associated infections, but as electronic patient records become increasingly ubiquitous and standardized, electronic surveillance software with general decision rules embedded in them could be used to identify possible healthcare-associated infections. For those with more intricate exposure histories, the techniques used to investigate individual cases here can be used, and these techniques are also potentially automatable. Clinicians and hospital epidemiologists can use the techniques presented here to investigate any disease for which it is important to know the source of infection and for which the incubation period is known.

In both routine infection control and potential healthcare emergencies, such as the emergence of pandemic H1N1 (2009), knowing where someone was likely infected is important in deciding the best public health response. For influenza and other respiratory viruses, the methods and recommendations presented here can help to better identify the potential source of infection, thereby giving more accuracy and focus to the infection control response.
APPENDIX A
MATHEMATICAL SUPPLEMENT

Notation

\( x \): the time of infection
\( t_i \): the start of period \( i \)
\( s \): the time of symptom onset
\( n \): the period immediately before symptom onset
\( t_{n+1} \): Let \( t_{n+1} = s \).
\( g(x) \): the probability distribution function for becoming infected at time \( x \)
\( f(x) \): the probability distribution function for the incubation period
\( F(x) \): the cumulative distribution function for the incubation period
\( q_i \): the probability of surviving to the start of period \( i \) without being infected
\( p_i \): the probability of being infected in period \( i \)

Full Solution

Assume that within each period \( i \) the rate (eg, hazard) of infection is constant, \( \lambda_x \).

The probability that a case was infected during period \( i \), given that they were not infected at the beginning of period \( i \), is

\[
Pr\{t_{i+1} \leq x \leq t_{i} \mid s, x > t_{i+1}\} = \int_{\max(t_i, s)}^{t_{i+1}} f(x) g(x \mid x > t_i) dx
\]

\[
= \int_{\max(t_i, s)}^{t_{i+1}} \frac{Pr\{s \mid x, x > t_i\} g(x \mid x > t_i)}{Pr\{s \mid x \geq t_i\}} dx
\]

\[
= \frac{1}{Pr\{s \mid x \geq t_i\}} \int_{\max(t_i, s)}^{t_{i+1}} f(s - s \lambda, e^{-\lambda x - t_i}) dx.
\]

This leads to Table A1 for calculating \( P_i \).

Solution Assuming Low Infection Rates

Assume that the rate of infection is low enough that the prior chance of being infected during any particular period is near 0 and the probability distribution function of the period is roughly constant at the rate of infection (note that this also implies that the prior probability of surviving any particular period uninfected is approximately 1). That is,

\[
\forall i, \lambda_i e^{-\lambda_i (t_{i+1} - t_i)} \approx \lambda_i.
\]

Hence,

\[
\int_{t_i}^{t_{i+1}} f(s - x) \lambda_i e^{-\lambda_i x - t_i} dx \approx \lambda_i [F(s - t) - F(s - t_{i+1})].
\]

Let \( r_i = \lambda_i / \lambda_x \), where \( x \) is some reference setting or period. Note that \( s - t_i \). This leads to Table A2.

Online Interactive Spreadsheet

The online interactive spreadsheet (Excel; Microsoft) was used to analyze the 3 examples in this article. This spreadsheet may also be used as a template for readers who wish to apply our techniques to their own data. For the spreadsheet to update properly, macros must be enabled. Each of the blue fields in the spreadsheet represents information that must be supplied by the user, whereas the nonshaded fields are calculated values. The case and disease headings are for identification purposes. The median and dispersion characterize the incubation period distribution of the disease of interest. The "number of periods" field allows the user to specify the number of periods in which infection may have occurred that they wish to examine. This can be any number of periods up to 20, and the spreadsheet will adjust the table to account for the desired number of periods. The period description column serves as an identifier for the periods of interest. Period start indicates the time at which each period begins, and the final period is considered to end at the symptom onset time. "Minimum inc" represents the calculated minimum incubation period, in days, that could have occurred if infection occurred in the given period. The incident rate ratio defaults to 1 but can be used to designate elevated or reduced risk of infection in that period, compared with some reference category. P(mass) is the probability mass associated with this interval, and "weight" is the weight for this interval. Finally, the P(infected in period) column gives the probability that the case was infected during the specified period on the basis of the time of symptom onset and input information.

https://doi.org/10.1086/655023 Published online by Cambridge University Press
**Table A1. General Tabular Solution for Calculating the Probability (P) of Having Been Infected in Any Period i**

<table>
<thead>
<tr>
<th>Period Start</th>
<th>( q_i )</th>
<th>( \rho_i )</th>
<th>( d_i )</th>
<th>( p_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(-\infty)</td>
<td>1</td>
<td>[ f(s-x) \lambda e^{-\lambda x} dx \sum_{k=0}^{\infty} \rho_k / d_0 ]</td>
<td>( p_0 / d_0 )</td>
</tr>
<tr>
<td>1</td>
<td>( t_1 )</td>
<td>1-( p_0 )</td>
<td>[ f(s-x) \lambda e^{-\lambda x} dx \sum_{k=1}^{\infty} (q_k \rho_k) / d_i ]</td>
<td>( p_1 / d_1 )</td>
</tr>
<tr>
<td>i</td>
<td>( t_i )</td>
<td>( q_{i-1} - p_{i-1} )</td>
<td>[ f(s-x) \lambda e^{-\lambda x} dx \sum_{k=1}^{\infty} (q_k \rho_k) / d_i ]</td>
<td>( p_i / d_i )</td>
</tr>
<tr>
<td>n</td>
<td>( t_n )</td>
<td>( q_{n-1} - p_{n-1} )</td>
<td>[ f(s-x) \lambda e^{-\lambda x} dx \rho_n ]</td>
<td>( p_n / d_n )</td>
</tr>
</tbody>
</table>

**Table A2. Tabular Solution for Calculating the Probability of Being Infected (p_i) in Any Period i, Assuming Low Rates of Infection**

<table>
<thead>
<tr>
<th>Period Start</th>
<th>( r_i )</th>
<th>( \rho_i )</th>
<th>( p_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(-\infty)</td>
<td>( r_0 )</td>
<td>( p_0 / d_0 )</td>
</tr>
<tr>
<td>1</td>
<td>( t_1 )</td>
<td>( r_1 )</td>
<td>( r_1 / d_1 )</td>
</tr>
<tr>
<td>i</td>
<td>( t_i )</td>
<td>( r_i )</td>
<td>( r_i / d_i )</td>
</tr>
<tr>
<td>n</td>
<td>( t_n )</td>
<td>( r_n )</td>
<td>( r_n / d_n )</td>
</tr>
</tbody>
</table>

**ACKNOWLEDGMENTS**

Financial support. National Institutes of Health (grant R01-TW008246 to D.A.T.C. and J.L.); the Bill and Melinda Gates Foundation (grant 49276 to D.A.T.C. and J.L.); Career Award at the Scientific Interface from the Burroughs Welcome Fund (to D.A.T.C.); US Department of Homeland Security (grant N00014-06-1-0991 to N.G.R. and R.B.); and the US Centers for Disease Control and Prevention (grant HK00014 to T.P.).

Potential conflicts of interest. All authors report no conflicts of interest relevant to this article.

Address reprint requests to Justin Lessler, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, E6545, Baltimore, MD 21205 (jlessler@jhsph.edu).

**REFERENCES**