A multi-tiered time-series modelling approach to forecasting respiratory syncytial virus incidence at the local level

M. C. SPAEDER[1]* AND J. C. FACKLER[2]

[1] *Division of Critical Care Medicine, Children's National Medical Center, Washington, DC, USA*
[2] *Department of Anesthesiology and Critical Care Medicine, The Johns Hopkins Hospital, Baltimore, MD, USA*

SUMMARY

Respiratory syncytial virus (RSV) is the most common cause of documented viral respiratory infections, and the leading cause of hospitalization, in young children. We performed a retrospective time-series analysis of all patients aged <18 years with laboratory-confirmed RSV within a network of multiple affiliated academic medical institutions. Forecasting models of weekly RSV incidence for the local community, inpatient paediatric hospital and paediatric intensive-care unit (PICU) were created. Ninety-five percent confidence intervals calculated around our models' 2-week forecasts were accurate to ±9·3, ±7·5 and ±1·5 cases/week for the local community, inpatient hospital and PICU, respectively. Our results suggest that time-series models may be useful tools in forecasting the burden of RSV infection at the local and institutional levels, helping communities and institutions to optimize distribution of resources based on the changing burden and severity of illness in their respective communities.

Key words: Modelling, respiratory infections, respiratory syncytial virus.

INTRODUCTION

In the USA, respiratory syncytial virus (RSV) is the most common cause of documented viral respiratory infections, and the leading cause of hospitalization, in young children [1, 2]. Caring for infants and children with RSV places a substantial burden on resources in both the outpatient and inpatient settings [3]. Children with risk factors, such as chronic lung disease, congenital heart disease and neuromuscular impairment, are at increased risk of morbidity and mortality from RSV infection [4–7].

The increased availability of highly sensitive and specific methods for the detection of viral pathogens, such as immunochromotography, direct fluorescent antibody, shell vial culture and polymerase chain reaction testing, has improved our ability to identify and survey infectious agents. RSV demonstrates a seasonal predominance in its presentation; however, it is difficult to predict with sufficient accuracy to allow prospective planning the beginning and end of the viral season as well as periods of peak incidence [8].

In the USA, voluntary reporting to the National Respiratory and Enteric Virus Surveillance System (NREVSS) of the Centers for Disease Control and Prevention is the main method by which RSV activity is tracked at the national and regional levels [9]. RSV Alert, an active surveillance programme created by MedImmune, also tracks RSV activity at the national, regional and local levels [10]. Both systems work with a network of laboratories in nearly every state to

* Author for correspondence: M. C. Spaeder, M.D., M.S., Division of Critical Care Medicine, Children's National Medical Center, 111 Michigan Avenue, NW, Washington, DC 20010, USA (Email: mspaeder@cnmc.org)

provide up to date surveillance data of RSV incidence [9, 10].

For decision makers at the local or institutional levels, surveillance that distinguishes between mild disease, which comprises most children who develop RSV, from moderate to severe disease requiring hospitalization or intensive care, could aid in the optimization of resource allocation. Furthermore, having a mechanism that not only tracks RSV activity but also provides decision makers with accurate forecasts of future activity might facilitate specific hospital or clinic staffing positions and real-time alterations of elective surgical schedules.

In recent years time-series modelling has been increasingly employed as a useful tool in infectious disease surveillance [11–14] and resource utilization [14–17]. Relatively easy to construct, time-series analysis allows not only for the modelling of time-ordered data but also provides the capability to forecast future observations. Model parameters can be updated in real time to adjust to fluctuations in the behaviour of a particular time-series and maximize the model's predictive abilities.

We hypothesized that a series of interrelated time-series models could be constructed based on analysis of historical data from our local community and hospital that would effectively capture the periodicity of RSV infection in children and forecast the incidence of RSV within our local community, hospital and paediatric intensive-care unit (PICU).

METHODS

The Institutional Review Board of Johns Hopkins Hospital approved this study. We performed a retrospective cohort study identifying all patients aged <18 years who underwent laboratory testing for RSV at one of the Johns Hopkins medical institutions in Maryland between 1 October 2002 and 30 September 2008. Laboratory-confirmed viral infection was defined as identification of RSV from a nasopharyngeal or endotracheal specimen by immunochromotography, direct fluorescent antibody, tube viral culture or shell vial culture. The date of infection was defined as the date that the specimen was obtained. Multiple positive specimens from an individual patient collected within 28 days of one another were considered a unique case.

The subset of patients admitted to the Johns Hopkins Hospital Children's Center, a 180-bed urban academic children's hospital with approximately 8400 admissions per year, and to the Johns Hopkins Hospital PICU, a 26-bed PICU with approximately 1800 admissions per year, with laboratory-confirmed RSV were identified. We designated three source categories: community, inpatient and PICU. For each sample, cases of RSV were aggregated by week and partitioned into experimental (1 October 2002 to 30 September 2007) and validation (1 October 2007 to 30 September 2008) datasets.

For each sample, the experimental dataset was plotted as a time-series and assessed for stationarity using the Augmented Dickey–Fuller test for unit roots. Type I error was set at 0·05. The autocorrelation and partial autocorrelation functions were calculated and plotted to aid in the initial identification of base models. Multiple candidate models were constructed relative to the base models based on the minimization of Akaike's Information Criterion (AIC) [18]. To examine the relationship of RSV incidence between the community setting (predominantly mild disease), the inpatient setting (moderate to severe disease) and PICU (severe disease), we constructed inpatient and PICU models that included community incidence variables. Maximum-likelihood testing was employed to determine inclusion or exclusion of specific model parameters at a significance level of 0·10. Maximum-likelihood estimation was used to calculate model parameter coefficients.

To assess the performance of each candidate model at forecasting, 1- and 2-week forecasts were derived and plotted against the corresponding validation dataset. Root mean squared errors (RMSEs) of the forecasts were calculated to derive 95% confidence intervals about the forecasts. The models with the lowest RMSE of 2-week forecasts were considered optimal. All calculations were performed using Stata/IC v. 10.1 (Stata Corporation, USA).

RESULTS

Community sample

A total of 1499 cases of laboratory-confirmed RSV infection in children who underwent testing at one of the Johns Hopkins medical institutions were included in the analysis. Cases of RSV infection were aggregated by week and partitioned into experimental (1132 cases) and validation (367 cases) datasets. Figure 1 displays the time-series plot of the experimental dataset, which was stationary (Mackinnon approximate P value <0·001).
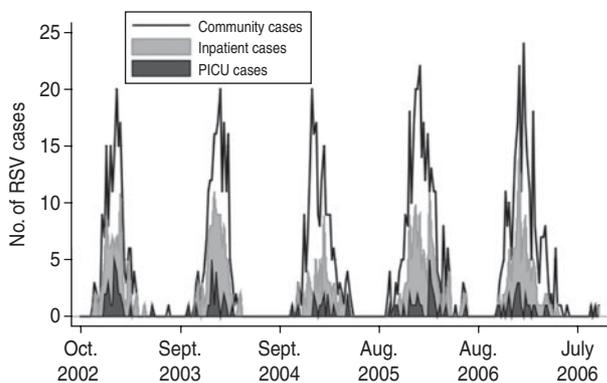
Fig. 1. Weekly incidence of respiratory syncytial virus (RSV) in children for the Johns Hopkins medical institutions (community), Johns Hopkins Hospital Children's Center (inpatient) and Johns Hopkins Hospital paediatric intensive-care unit (PICU) for 2002 to 2007.

The plot of the autocorrelation function resembled a dampened sine wave while the partial autocorrelation function cuts off at lag 8, suggesting an autoregressive process of order 8 [AR(8)] as the base model. The community base model can be expressed as:

$$Z_t = \phi_{t-1}Z_{t-1} + \phi_{t-2}Z_{t-2} + \phi_{t-3}Z_{t-4} + \phi_{t-5} + \phi_{t-6}Z_{t-6} + \phi_{t-7}Z_{t-7} + \phi_{t-8}Z_{t-8} + a_t,$$

where Z_t = community RSV cases for week t; Z_{t-i} = community RSV cases for week $t-i$; ϕ_{t-i} = weighted coefficient for community RSV cases for week $t-i$; and a_t = white noise term for week t.

The AIC for the base model was 1291. An autoregressive model with significant lags at lag 1 week ($P < 0.001$), lag 2 weeks ($P < 0.001$), lag 3 weeks ($P = 0.04$), lag 5 weeks ($P < 0.001$), lag 6 weeks ($P = 0.04$) and lag 8 weeks ($P = 0.02$) minimized the AIC (1288) and RMSE (4·67) of 2-week forecasts. The optimal community model can be expressed as:

$$Z_t = 0.5754Z_{t-1} + 0.4596Z_{t-2} - 0.1089Z_{t-3} + 0.2054Z_{t-5} - 0.1191Z_{t-6} - 0.1106Z_{t-8} + a_t,$$

where Z_t = community RSV cases for week t; Z_{t-i} = community RSV cases for week $t-i$; and a_t = white noise term for week t. Figure 2a displays the plot of community incidence of RSV cases against 2-week forecasts.

The RMSEs for the 1- and 2-week forecasts of the community model were 3·66 and 4·67, respectively. Ninety-five percent confidence intervals calculated around the community model's forecasts were accurate to ± 7.32 cases/week for the 1-week forecasts and ± 9.34 cases/week for the 2-week forecasts. On average for the validation time period, the true community
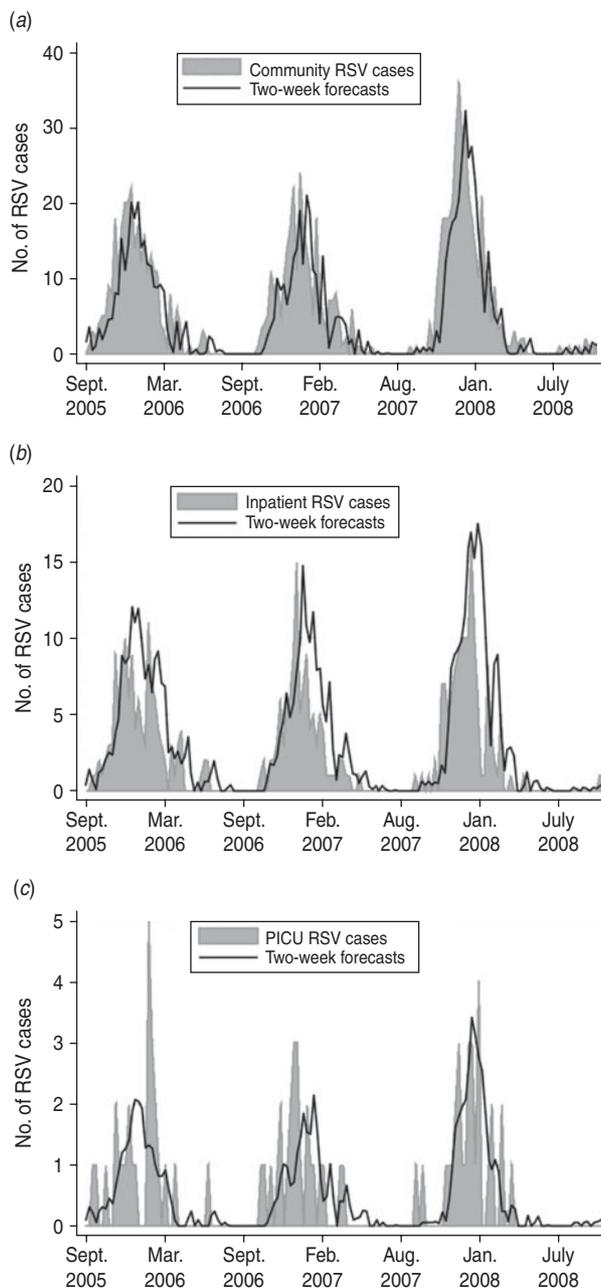
Fig. 2. Plots of (a) community incidence, (b) inpatient incidence, (c) PICU incidence of RSV cases vs. model estimates, forecasting 2 weeks into the future for 2005–2008.

RSV activity was underestimated by 0·68 case/week for the 1-week forecasts and by 1·08 cases/week for the 2-week forecasts.

Inpatient sample

A total of 631 children with laboratory-confirmed RSV infection required admission to the Johns Hopkins Hospital Children's Center. Cases of RSV infection were aggregated by week and partitioned

into experimental (499 cases) and validation (132 cases) datasets. Figure 1 displays the time-series plot of the experimental dataset, which was stationary (Mackinnon approximate P value <0·001).

The plot of the autocorrelation function resembled a dampened sine wave while the partial autocorrelation function cuts off at lag 9, suggesting an autoregressive process of order 9 [AR(9)] as the base model.

The AIC for the base model was 1010. Systematic removal of parameters from the inpatient base model did not minimize the AIC. Inclusion of a variable of community RSV incidence at lag 1 week reduced the AIC to 994. An autoregressive model with significant lags at lag 1 week (P<0·001), lag 2 weeks (P<0·001), lag 3 weeks (P<0·001), lag 6 weeks (P=0·04), lag 7 weeks (P=0·01), lag 8 weeks (P=0·03), lag 9 weeks (P=0·02) and community lag 1 week (P<0·001) minimized the AIC (991) and RMSE (3·75) of 2-week forecasts.

The optimal inpatient model can be expressed as:

$$X_t = 0.1873X_{t-1} + 0.3721X_{t-2} + 0.2017X_{t-3} - 0.1341X_{t-6} + 0.1657X_{t-7} + 0.1157X_{t-8} - 0.1536X_{t-9} + 0.2494Z_{t-1} + a_t,$$

where X_t = inpatient RSV cases for week t; X_{t-i} = inpatient RSV cases for week $t-i$; Z_{t-1} = community RSV cases for week $t-1$; and a_t = white noise term for week t. Figure 2b displays the plot of inpatient incidence of RSV cases against 2-week forecasts.

The RMSEs for the 1- and 2-week forecasts of the inpatient model were 3·3 and 3·75, respectively. Ninety-five percent confidence intervals calculated around the inpatient model's forecasts were accurate to ±6·6 cases/week for the 1-week forecasts and ±7·5 cases/week for the 2-week forecasts. On average, for the validation time period the true inpatient RSV activity was overestimated by 1·11 cases/week for the 1-week forecasts and by 1·14 cases/week for the 2-week forecasts.

PICU sample

A total of 140 children with laboratory-confirmed RSV infection required admission to the Johns Hopkins Hospital PICU. Cases of RSV infection were aggregated by week and partitioned into experimental (113 cases) and validation (27 cases) datasets. Figure 1 displays the time-series plot of the experimental dataset, which was stationary (Mackinnon approximate P value <0·001).

The plot of the autocorrelation function resembled a dampened sine wave while the partial autocorrelation function cuts off at lag 9, suggesting an autoregressive process of order 9 [AR(9)] as the base model.

The AIC for the base model was 623. Systematic removal of parameters from the PICU base model did not minimize the AIC. Inclusion of a variable of community RSV incidence at lag 1 week reduced the AIC to 574. An autoregressive model with significant lags at lag 1 week (P<0·001), lag 9 weeks (P=0·06) and community lag 1 week (P<0·001) minimized the AIC (561) and RMSE (0·76) of 2-week forecasts.

The optimal PICU model can be expressed as:

$$Y_t = 0.1745Y_{t-1} - 0.1529Y_{t-9} + 0.0899Z_{t-1} + a_t,$$

where Y_t = PICU RSV cases for week t; Y_{t-i} = PICU RSV cases for week $t-i$; Z_{t-1} = community RSV cases for week $t-1$; and a_t = white noise term for week t. Figure 2c displays the plot of PICU incidence of RSV cases against 2-week forecasts.

The RMSE for both the 1- and 2-week forecasts of the PICU model was 0·76. Ninety-five percent confidence intervals calculated around the PICU model's forecasts were accurate to ±1·52 cases/week for both the 1- and 2-week forecasts. On average for the validation time period, the true PICU RSV activity was overestimated by 0·12 case/week for the 1-week forecasts and by 0·07 case/week for the 2-week forecasts.

DISCUSSION

The impact of RSV on resource utilization during seasonal epidemics can be substantial in both the outpatient and inpatient settings [3, 19]. Isolation procedures, bed utilization, availability of human resources (physicians, nurses, respiratory therapists) and medical equipment (ventilators, nebulization systems) are just a few of factors inherent to resource utilization affected by the seasonal influx of patients with RSV. Our identified models, derived from historical data from our institution and local community produced accurate 1- and 2-week forecasts of RSV incidence in our community, hospital and PICU. We believe these models can be used prospectively to anticipate and adjust, in real-time, resource allocation.

The current surveillance systems provided by NVRESS and the RSV Alert programme provide invaluable information in tracking RSV activity in the USA. The data, however, is often aggregated over large regional or metropolitan areas and does not provide information regarding severity of illness.

Our multi-tiered approach, while specific to our community and institution, provides hospital leadership with personalized data and forecasts to address the changing burden on the institution.

RSV is an enveloped RNA paramyxovirus transmitted predominantly through direct contact although transmission through respiratory droplet can also occur [8, 20]. While helpful in reducing the spread of virus to other patients, in the hospital isolation procedures can impact resource utilization by limiting nursing ratios and closing available beds [7]. Accurate forecasting could help hospital leadership prepare for changes in resource needs brought on by the demands of a surge in viral respiratory admissions.

Our study has a number of limitations. Our model is specific to a single community and institution and we have no evidence to suggest that it can be generalized to other institutions or communities. Despite the specificity of the model, our objective was to design models particular to our own community and institution. While further work is required to assess if similar models can be constructed for institutions in other locales, our experience suggests that these models work well at the local and institutional levels.

The possibility of reaching institutional bed capacity certainly exists especially during times of peak respiratory viral illness incidence. Patients requiring admission to the hospital or PICU but who are referred to other institutions during these capacity events cause under-reporting of the true demand on the institution and have the potential to affect future forecasts. While we do not have specific information on patients with RSV referred to other institutions during high-capacity events, we surmise that the frequency of these events is rare.

During periods of high incidence of other respiratory viruses, namely influenza, children with mild viral respiratory illness may be more likely to undergo testing leading to the potential discovery of an increased number of RSV cases at the community level. For our period of study, there was no correlation between seasons with unusually high incidence of influenza (e.g. 2003–2004) and the incidence of RSV at the community level.

We used methods common in time-series analysis, calculation of RMSE of forecasts, to compare our candidate models. These methods are most helpful when the loss associated with forecasting error is symmetric, in other words, when underestimating the true value is worth the same as overestimating the true value. In our study, it can be argued that the costs associated with underestimating RSV incidence are greater than the costs associated with overestimating incidence. Our optimal models for the inpatient hospital and PICU, overestimated RSV incidence, on average, by 1·14 and 0·07 cases/week, respectively. Our optimal community model, while underestimating RSV incidence on average, did so by only 1·08 cases/week.

Our results suggest that time-series models may be useful tools in forecasting the burden of RSV infection at the local and institutional levels, helping communities and institutions to optimize distribution of resources based on the changing burden and severity of illness in their respective communities.

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Shay DK, et al.** Bronchiolitis-associated hospitalizations among US children, 1980–1996. *Journal of the American Medical Association* 1999; **282**: 1440–1446.
2. **Hall CB.** Respiratory syncytial virus and parainfluenza virus. *New England Journal of Medicine* 2001; **344**: 1917–1928.
3. **Hall CB, et al.** The burden of respiratory sycytial virus infection in young children. *New England Journal of Medicine* 2009; **360**: 588–598.
4. **Welliver RC, Checchia PA, Bauman JH.** Fatality rates in published reports of RSV hospitalizations among high-risk and otherwise healthy children. *Current Medical Research and Opinion* 2010; **26**: 2175–2181.
5. **Thornburn K.** Pre-existing disease is associated with a significantly higher risk of death in severe respiratory syncytial virus infection. *Archives of Disease in Childhood* 2009; **94**: 99–103.
6. **Wilkesmann A, et al.** Hospitalized children with respiratory syncytial virus infection and neuromuscular impairment face an increased risk of a complicated course. *Pediatric Infectious Disease Journal* 2007; **26**: 485–491.
7. **Welliver RC.** Review of epidemiology and clinical risk factors for severe respiratory syncytial virus infection. *Journal of Pediatrics* 2003; **143**: S112–S117.
8. **Pickering LK, et al. (eds).** *Red Book: 2006 Report of the Committee on Infectious Diseases*, 27th edn. Elk Grove Village, IL: American Academy of Pediatrics, 2006, pp. 560–566.

9. **Centers for Disease Control and Prevention.** Brief report: respiratory syncytial virus activity – United States, July 2007–December 2008. *Morbidity and Mortality Weekly Report* 2008; **57**: 1355–1358.

10. **Boron ML, *et al.*** A novel active respiratory syncytial virus surveillance system in the United States: variability in the local and regional incidence of infection. *Pediatric Infectious Disease Journal* 2008; **27**: 1095–1098.

11. **Allard R.** Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization* 1998; **76**: 327–333.

12. **Fernandez-Perez CJ, Tejada MC.** Multivariate time series analysis in nosocomial infection surveillance: a case study. *International Journal of Epidemiology* 1998; **27**: 282–288.

13. **Luz PM, *et al.*** Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *American Journal of Tropical Medicine and Hygiene* 2008; **79**: 933–939.

14. **Spaeder MC, Fackler JC.** Time series model to predict burden of viral respiratory illness on a pediatric intensive care unit. *Medical Decision Making*. Published online: 2 December 2010. doi: 10.1177/0272989X10388042.

15. **Earnest A, *et al.*** Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Services Research* 2005; **5**: 36.

16. **Reis BY, Mandl KD.** Time series modeling for syndromic surveillance. *BMC Medical Informatics and Decision Making* 2003; **3**: 11.

17. **Upshur REG.** Time-series analysis of the relation between influenza virus and hospital admissions of the elderly in Ontario, Canada for pneumonia, chronic lung disease, and congestive heart failure. *American Journal of Epidemiology* 1999; **149**: 85–92.

18. **Wei WS.** *Time Series Analysis: Univariate and Multivariate Methods*, 1st edn. Reading, MA: Addison Wesley, 1990.

19. **Iwane MK, *et al.*** Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. *Pediatrics* 2004; **113**: 1758–1764.

20. **Sandrock C, Stollenwerk N.** Acute febrile respiratory illness in the ICU: reducing disease transmission. *Chest* 2008; **133**: 1221–1231.