Detection of outbreaks of infection or increases in bacterial resistance to antimicrobial agents is an essential component of hospital infection control surveillance. The authors applied the method of exponential smoothing to microbiology data from 1987–1992 to investigate a suspected outbreak of gentamicin resistance among Pseudomonas aeruginosa bacteria at the Department of Veterans Affairs Medical Center, San Francisco, California, in 1991–1992. The years 1987–1990 were used to develop the baseline for the forecast model. Application of the model indicated that two observed prominent peaks in the annual cumulative incidence of gentamicin-resistant P. aeruginosa were within the upper bounds of their respective 95% confidence intervals as estimated by the forecast model—i.e., that no epidemic was in progress. This prediction was supported by investigations by the hospital’s infection control team which indicated that the apparent increases were due to readmission of patients previously known to harbor these organisms. In contrast, application of a typically employed method that ignores the time series data structure indicated that there were 6 months in which incidence rates exceeded the upper bounds of their respective 95% confidence intervals, thereby erroneously suggesting that an epidemic was in progress. Recursive algorithms and some simplifying assumptions that do not affect the validity of inferences make the application of this method practical for nosocomial infection control programs.

threshold value in making decisions about whether an outbreak may be occurring, it seems reasonable to give greater weight to more recent events than to more distant events in the estimation of the threshold level. A second, related limitation of these approaches is that they ignore the time series structure of the data and the possible serial correlations in them.

The method of exponential smoothing (12, 13) allows for such differential weighting. Moreover, when implemented with a recursive algorithm, this approach is simple and practical to use in the context of regular nosocomial infection surveillance activities. We have applied exponential smoothing to 6 years of antibiotic resistance patterns derived from the computerized microbiology laboratory data of the Department of Veterans Affairs (DVA) Medical Center, San Francisco, California, to illustrate the properties of the method and the simplicity of its implementation. This analysis was stimulated by ongoing infection control surveillance activities at the medical center which suggested that there was an increase in the occurrence of gentamicin resistance among Gram-negative bacteria that began in 1989.

METHODS

Source of data

Beginning in May 1986, all antibiotic sensitivity testing for bacteria isolated from clinical specimens at the DVA Medical Center was performed with the use of the Vitek Automated Microbiological System (Biomerieux; Vitek, Inc., Hazelwood, Missouri). All results were entered manually into the microbiology component of the DVA Decentralized Hospital Computer Program, which became available in January 1987. Beginning in January 1991, all data, starting with those of January 1, 1987, were converted to the ASCII format and were transferred monthly or bi-monthly to SAS data sets (Statistical Analysis System; SAS Institute, Inc., Cary, North Carolina) for use as part of routine infection control surveillance for the medical center.

The present analysis used isolates of Pseudomonas aeruginosa obtained from January 1, 1987, through December 31, 1992. Only the first isolate of the month from a given patient was included in that month’s cumulative incidence calculation. If a patient’s hospital stay spanned 2 or more consecutive months, the first isolate from each month for that patient was counted. Thus, the monthly cumulative incidence estimates reflect estimates of the cumulative incidence of isolates for given months but not necessarily the cumulative incidence of new isolates in the hospital for that month. The percentage of total isolates contributed by patients in 2 or more consecutive months was evaluated over the period June 1992–December 1992 and was found to range from 0 to 3 percent. Thus, based on the relatively low frequency of isolates being obtained from the same patient over consecutive months, the assumption of independence of the monthly cumulative incidences that is required by the exponential smoothing technique was considered met.

Statistical analysis

Graphic analysis was used initially to evaluate the data. Plots of the monthly cumulative incidence of gentamicin resistance were produced for the entire 6-year period, as well as separately for each year. The occurrence of a seasonal trend was evaluated by computing the monthly mean values for each month over the 6 years (e.g., the mean of all January data, etc.) and subtracting from the monthly mean the average of all of the monthly means. Plots of the monthly residuals and evaluation of the autocorrelation function did not show evidence of any seasonal trends in the data. The smoothed trend in the data was obtained from the following model (13):

\[ x_t = \text{smooth}_t + \text{season}_t + \text{error}_t, \]

where \( x_t \) is the observed monthly \((t)\) frequency of gentamicin-resistant organisms; \( \text{smooth}_t \) is the smoothed monthly average based on a 13-month moving average and is equal to

\[ \{(1/24)x_{t-6} + (1/12)x_{t-5} + \ldots (1/12)x_{t+5} + (1/24)x_{t+6}\}; \]

\( \text{season}_t \) is the seasonal component (the difference between the average frequency for all months, (e.g., all Januarys) and the average of all of the monthly mean frequencies of isolates over the 6 years of data); and \( \text{error}_t \) is \( x_t - (\text{smooth}_t + \text{season}_t) \).

A time series is stationary if the expectation and variance are independent of time. The autocorrelation and partial autocorrelation function were used to create the correlogram, which showed that the estimated autocorrelation values were within \( \pm (2/\sqrt{N}) \) (13, 14). The Durbin-Watson statistic (15) was also used to test the coefficient of lag 1 autocorrelation, and it indicated that the series was stationary \((D \approx 1.9)\). This analysis showed that the monthly series over the 6 years fulfilled the requirements for overall stationarity.

An initial analysis evaluated the use of an autoregressive (AR) model (13) to describe the data. The best-fitting autoregressive model was an AR1 model. However, an AR1 model would not have been practi-
Exponential Smoothing in Nosocomial Infection Surveillance 639
cal for implementation by the infection control team, because the entire model would have had to be refitted as each new month’s data became available. Therefore, the method of exponential smoothing was selected to provide “one-step-ahead” forecasts of the next month’s expected frequency of gentamicin resistance.

The forecast model chosen is of the form (13)

\[ S_{t+1} = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots, \]

where \( x_1, x_2, \ldots, x_t \) = cumulative incidence for month \( t \), \( S_{t+1} \) = forecast cumulative incidence at \( t + 1 \) (the next month), and \( \beta_i = \alpha(1 - \alpha)^i \) (\( i = 0, 1, 2, \ldots, 0 < \alpha < 1 \)) is a geometric series whose summation is 1.

The formulation of the \( \beta_i's \) as a geometric series whose sum equals 1 has the effect of giving greater weight to more recent observations. For large samples, \( S_{t+1} \) is an unbiased estimator of the next month's forecast whose expected value is a constant cumulative incidence, \( \alpha \) (i.e., \( E(S_{t+1}) = \alpha \)) based on the assumption of stationarity of the series of cumulative incidence over time (Appendix 1). \( S_{t+1} \) can be expressed recursively as

\[
S_{t+1} = \alpha x_t + (1 - \alpha) \sum_{i=2}^{N} \alpha^i x_{t-i+1},
\]

This recursive formula requires only that estimates of \( \alpha \) and the current month’s cumulative incidence be available. The optimal estimate of \( \alpha \) is obtained by iteration for all \( \alpha \in (0.01, 0.99) \) and the selection of that \( \alpha \) which minimizes the sum of squared errors (SSE) around the estimate of \( S_{t+1} \) (Appendix 2).

Ninety-five percent confidence intervals around each forecast can be calculated with the forecast error variance defined (12) as follows:

\[
e_k = x_t - S_k,
\]

where \( k \) is the current month;

\[
\sigma^2_T = \frac{\sum_{i=T-N+1}^{T} (e_i - \bar{e}_T)^2}{N - 1},
\]

where \( T = \) the forecast month and \( N = \) the number of forecast months as of \( T \); and

\[
\bar{e}_T = \frac{\sum_{i=T-N+1}^{T} e_i}{N}.
\]

An approximate formula for ease of calculation is as follows. Assuming that \( E(e) = 0 \), then

\[
\sigma^2_T = \frac{\sum_{i=T-N+1}^{T} e_i^2}{N}.
\]

RESULTS

Over the 6-year period January 1, 1987, through December 31, 1992, the number of patients per month from whom \( P. \) aeruginosa was isolated ranged between a low of six (February 1990) and a high of 68 (December 1992) (table 1); and the monthly cumulative incidence of gentamicin resistance ranged between 5.6 percent (January 1988) and 62.5 percent (November 1990) (table 1, figure 1). In the years 1990–1992, the annual maximum monthly cumulative incidence of gentamicin resistance exceeded the rates observed for the previous 3 years (figure 1).

Figure 2 shows the relation between the SSE for the “one-month-ahead” cumulative incidence forecast and \( \alpha \). The \( \alpha \) that minimized the SSE was equal to 0.16. The figure also demonstrates that the SSE for the forecast is relatively insensitive to \( \alpha \) in the range of 0.10–0.22. Over this range, there is a \( \leq 1 \) percent increase in the SSE. The relative insensitivity of the model to the choice of \( \alpha \) over this range was further investigated by generating predicted monthly cumulative incidence for the period January 1, 1990, through December 31, 1992, based on 0.01 increments of \( \alpha \) in the range of 0.10–0.22 (figure 3). Over this range of \( \alpha = 0.10–0.22 \), the curves of predicted monthly cumulative incidence are virtually identical to the curve observed for the optimal \( \alpha \) of 0.16.

The monthly cumulative incidence data from January 1, 1987, through December 31, 1990, were used as the baseline data from which to develop the forecast, since concerns about the increased occurrence of gentamicin resistance dated back to late 1989 and early 1990 (figure 1). The forecast estimate for January 1, 1991, was based on the data from the previous 48 months. (See Appendix 3 for the application of the method to calculation of this estimate and calculation of the forecast for January–March 1991, and the upper bounds of the respective confidence intervals.) Figure 4 shows a summary of the data for the period January 1, 1991, through December 31, 1992. Two prominent peaks (February 1991 and May 1992) can be observed, although only the May 1992 peak approaches the upper bound of the 95 percent confidence interval of the forecast for the month. In both cases, the peaks represent more than a doubling of the underlying rates.
TABLE 1. Monthly cumulative incidence (%) of gentamicin resistance in selected Gram-negative bacteria, Department of Veterans Affairs Medical Center, San Francisco, California, 1987–1992

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>26.3</td>
<td>5.6</td>
<td>9.1</td>
<td>32.0</td>
<td>25.0</td>
<td>35.3</td>
</tr>
<tr>
<td>February</td>
<td>29.4</td>
<td>5.3</td>
<td>11.8</td>
<td>16.7</td>
<td>58.3</td>
<td>41.7</td>
</tr>
<tr>
<td></td>
<td>(17)</td>
<td>(19)</td>
<td>(17)</td>
<td>(6)</td>
<td>(12)</td>
<td>(12)</td>
</tr>
<tr>
<td>March</td>
<td>29.4</td>
<td>31.3</td>
<td>15.4</td>
<td>28.6</td>
<td>33.3</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>(17)</td>
<td>(16)</td>
<td>(26)</td>
<td>(14)</td>
<td>(18)</td>
<td>(24)</td>
</tr>
<tr>
<td>April</td>
<td>38.5</td>
<td>33.3</td>
<td>10.5</td>
<td>12.5</td>
<td>46.2</td>
<td>25.9</td>
</tr>
<tr>
<td></td>
<td>(13)</td>
<td>(12)</td>
<td>(19)</td>
<td>(8)</td>
<td>(13)</td>
<td>(27)</td>
</tr>
<tr>
<td>May</td>
<td>21.4</td>
<td>36.8</td>
<td>28.6</td>
<td>56.3</td>
<td>12.5</td>
<td>52.6</td>
</tr>
<tr>
<td></td>
<td>(14)</td>
<td>(19)</td>
<td>(21)</td>
<td>(16)</td>
<td>(8)</td>
<td>(19)</td>
</tr>
<tr>
<td>June</td>
<td>14.3</td>
<td>23.1</td>
<td>15.4</td>
<td>13.3</td>
<td>5.0</td>
<td>35.3</td>
</tr>
<tr>
<td></td>
<td>(14)</td>
<td>(13)</td>
<td>(13)</td>
<td>(15)</td>
<td>(20)</td>
<td>(17)</td>
</tr>
<tr>
<td>July</td>
<td>21.7</td>
<td>28.6</td>
<td>35.0</td>
<td>15.0</td>
<td>18.5</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td>(23)</td>
<td>(21)</td>
<td>(20)</td>
<td>(20)</td>
<td>(27)</td>
<td>(18)</td>
</tr>
<tr>
<td>August</td>
<td>7.1</td>
<td>13.6</td>
<td>15.0</td>
<td>36.4</td>
<td>19.1</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>(14)</td>
<td>(22)</td>
<td>(20)</td>
<td>(22)</td>
<td>(27)</td>
<td>(20)</td>
</tr>
<tr>
<td>September</td>
<td>36.8</td>
<td>17.7</td>
<td>13.8</td>
<td>43.5</td>
<td>25.0</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>(19)</td>
<td>(17)</td>
<td>(29)</td>
<td>(23)</td>
<td>(24)</td>
<td>(27)</td>
</tr>
<tr>
<td>October</td>
<td>14.3</td>
<td>36.8</td>
<td>35.0</td>
<td>25.0</td>
<td>31.8</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>(14)</td>
<td>(19)</td>
<td>(20)</td>
<td>(12)</td>
<td>(22)</td>
<td>(34)</td>
</tr>
<tr>
<td>November</td>
<td>22.2</td>
<td>18.8</td>
<td>33.3</td>
<td>62.5</td>
<td>20.0</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>(18)</td>
<td>(16)</td>
<td>(15)</td>
<td>(15)</td>
<td>(20)</td>
<td>(67)</td>
</tr>
<tr>
<td>December</td>
<td>8.3</td>
<td>11.8</td>
<td>42.9</td>
<td>40.0</td>
<td>30.4</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>(12)</td>
<td>(17)</td>
<td>(21)</td>
<td>(10)</td>
<td>(23)</td>
<td>(68)</td>
</tr>
</tbody>
</table>

* Numbers in parentheses, total number of patients with isolates for that month.

In terms of the forecast model, the data from each of these months would be consistent with the random variation of the series of monthly cumulative incidences. The possibility of an outbreak was investigated for each of these months. At the time that these data were actually obtained, it was not felt by the infection control team of the hospital that an outbreak was occurring, and no special actions were taken. A retrospective reassessment of the data arrived at the same conclusion.

For comparison of the forecasting method with a method that uses unweighted updating and does not
consider the time series character of the data, the upper bound thresholds were generated by the method proposed by Morrison et al. (11). The approach taken by Morrison and colleagues is typical, in principle, of methods that are currently in use for monitoring outbreak occurrence or improvements in cumulative incidence. In this method, the upper bound for any given month is generated as the upper bound of the 95 percent confidence interval based on the variance of the average rate during the "baseline" period. The baseline for each month is computed as the average frequency of patients with resistant isolates for all available data up to the year prior to the current year. (For the present analysis, this corresponds to 1987–1990 for the year 1991 and 1987–1991 for the year 1992.) In all cases except January 1991, the upper bounds of the 95 percent intervals calculated by this method were substantially lower than those produced by the forecast model (table 2). Rates for 6 months (February and April 1991 and January, February, May, and June 1992) exceeded their upper bounds by this method (figure 5). The probability of observing rates in 4 months of 1992 (January, February, May, and June) above their upper bounds is $1.58 \times 10^{-4}$. If 1991 had served as the baseline for 1992, the Morrison method would have more strongly concluded that an outbreak was in progress. Moreover, when the cumulative incidence data from December 1991 are considered along with the 1992 data, this method would have suggested a protracted period of a greater than expected monthly cumulative incidence of patients with gentamicin-resistant \textit{P. aeruginosa} (figure 5).

**DISCUSSION**

The present investigation has demonstrated that the method of exponential smoothing forecasting may be

**TABLE 2. Upper bound of the 95% confidence interval (%) as calculated by the forecast model versus the method proposed by Morrison et al.***

<table>
<thead>
<tr>
<th>Month</th>
<th>Forecast model</th>
<th>Morrison et al. method</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 1991</td>
<td>25.4</td>
<td>26.6</td>
</tr>
<tr>
<td>February 1991</td>
<td>71.3</td>
<td>26.1</td>
</tr>
<tr>
<td>March 1991</td>
<td>68.3</td>
<td>36.6</td>
</tr>
<tr>
<td>April 1991</td>
<td>67.1</td>
<td>35.2</td>
</tr>
<tr>
<td>May 1991</td>
<td>70.4</td>
<td>46.7</td>
</tr>
<tr>
<td>June 1991</td>
<td>69.2</td>
<td>26.3</td>
</tr>
<tr>
<td>July 1991</td>
<td>62.7</td>
<td>34.2</td>
</tr>
<tr>
<td>August 1991</td>
<td>59.3</td>
<td>26.1</td>
</tr>
<tr>
<td>September 1991</td>
<td>56.4</td>
<td>37.2</td>
</tr>
<tr>
<td>October 1991</td>
<td>55.4</td>
<td>38.5</td>
</tr>
<tr>
<td>November 1991</td>
<td>55.3</td>
<td>45.0</td>
</tr>
<tr>
<td>December 1991</td>
<td>53.4</td>
<td>36.5</td>
</tr>
<tr>
<td>January 1992</td>
<td>53.8</td>
<td>27.9</td>
</tr>
<tr>
<td>February 1992</td>
<td>55.4</td>
<td>34.2</td>
</tr>
<tr>
<td>March 1992</td>
<td>56.9</td>
<td>37.0</td>
</tr>
<tr>
<td>April 1992</td>
<td>55.3</td>
<td>38.9</td>
</tr>
<tr>
<td>May 1992</td>
<td>56.6</td>
<td>41.0</td>
</tr>
<tr>
<td>June 1992</td>
<td>59.8</td>
<td>22.3</td>
</tr>
<tr>
<td>July 1992</td>
<td>59.9</td>
<td>31.6</td>
</tr>
<tr>
<td>August 1992</td>
<td>58.5</td>
<td>25.5</td>
</tr>
<tr>
<td>September 1992</td>
<td>56.3</td>
<td>35.5</td>
</tr>
<tr>
<td>October 1992</td>
<td>52.9</td>
<td>38.0</td>
</tr>
<tr>
<td>November 1992</td>
<td>53.6</td>
<td>40.7</td>
</tr>
<tr>
<td>December 1992</td>
<td>50.4</td>
<td>36.1</td>
</tr>
</tbody>
</table>

FIGURE 3. Curves for forecasting the monthly cumulative incidence of gentamicin-resistant *Pseudomonas aeruginosa* over the range $\alpha = 0.10-0.22$ with increments in $\alpha$ of 0.01: Department of Veterans Affairs Medical Center, San Francisco, California, January 1990-December 1992. The curve for $\alpha = 0.16$ represents the curve with the minimum sum of squared errors around monthly forecasts. $\bigcirc$, forecast of monthly cumulative incidence based on exponential smoothing with monthly cumulative incidence data from January 1987-December 1990 taken as the "baseline"; $\Diamond$, observed monthly cumulative incidence.
superior to the more typical methods that are used to
detect outbreaks in nosocomial infection surveillance
programs (2–4, 8–11). The principal advantages of
the method are that it takes into account the relevant
data history for the particular surveillance problem and
it weights the data such that more recent experience is
given greater weight than events further back in time.
The net effect of this approach is to provide “action
boundaries” (expressed as the upper bound of a
\((1 - (\gamma/2))\) percent confidence interval, where \(\gamma\)
denotes type I error probability) that are more realistic
(specific) than would be obtained with conventional
methods that ignore the time series quality of surveil-
ance data. Moreover, the investigation has demon-
strated that the analysis is easy to implement through
the use of a recursive forecast formula and a simplified
approach to the calculation of the desired confidence
intervals.
One potential limitation of the practical usefulness of this method relates to the selection of the optimal $\alpha$, a smoothing constant. The formal, iterative method described in Appendix 2 would be difficult to implement in infection control programs in which people have limited access to or experience with the necessary computing tools. Johnson and Montgomery (12) have suggested that the optimal $\alpha$ for minimizing the SSE of the forecast usually lies between 0.01 and 0.31. In the present analysis, the optimal $\alpha$ was 0.16. More importantly, for $\alpha$'s between 0.10 and 0.22, the differences in the SSEs relative to $\alpha = 0.16$ were $\leq 1$ percent; and the predicted cumulative incidence curves were virtually identical over this range (figure 3). Therefore, a practical alternative to the iterative determination of the optimal $\alpha$ would be the selection of two or three values of $\alpha$ over the range 0.01-0.31 to find the approximate range of $\alpha$'s over which the forecast curves are nearly identical. With the recursive formula given above (equation 1), this task can be accomplished on a hand-held calculator.

A second potential practical limitation relates to the calculation of the confidence intervals. Exact specification of the variance of the forecast ($\sigma^2_T$; equation 3) requires continuous updating of the average forecast error ($\bar{\varepsilon}_T$; equation 4). This can be accomplished simply by keeping track of the running total of the $\varepsilon_T$'s, the monthly forecast errors. Alternatively, the running sum of the monthly forecast errors alone can be used to estimate the variance (equation 5) with the assumption that $E(\varepsilon_T) = 0$. Figure 6 shows the differences between the upper bounds of the 95 percent confidence intervals obtained with equation 3 (the “standard method”) and those obtained with equation 5 (the “abbreviated method”). Equation 5 has limited usefulness for short-term predictions. However, for such short-term predictions, the use of equation 3 is not particularly burdensome. Over the long term, equation 5 provides an acceptable alternative to equation 3, in that the differences between the variances estimated by the two methods approach zero (figure 6). Application of equation 5 to the data did not alter any of the monthly inferences presented in figure 4.

A critical assumption of the forecast method is that of a stable underlying cumulative incidence (12)—i.e., stationarity of the time series. Forecast estimates are unbiased only if this assumption is met. No evidence was found for nonstationarity in the data used for this example. However, if the method is applied to diverse data sets, it is quite probable that changes in infection control practices and/or changes in the underlying risk profile of patients could lead to monthly cumulative incidences (of some nosocomial infection or the isolation of a particular strain of bacteria) that are changing over time. With the first-order exponential smoothing of the type used herein, the effect of a linear, time-dependent trend (increase or decrease) in the underlying cumulative incidence would lead to a bias in each forecast of magnitude $\{(1 - \alpha)/\alpha\} \lambda$, where $\lambda$ is the linear trend slope (12). If the linear trend is positive (increasing cumulative incidence with time), the forecast will be biased toward a lower value than would be observed for the unbiased estimate of the forecast (underestimate of the trend). If the trend is
negative (decreasing cumulative incidence), the forecast will be biased toward an overestimation of the cumulative incidence. In both cases, the variance of the forecast estimate will be increased. The overall result of nonstationarity in this situation would be a tendency for the cumulative incidences to be biased to an unknown degree (dependent on the actual trend) and to have a higher upper-bound (less precise and specific) confidence limit relative to the unbiased estimate. Visual analysis of the data may in some cases permit the detection of such trends, in which case a new baseline that reflects overall stationarity can be computed as the basis for future forecasts.

Application of a typical method (11) used for analysis of infection control surveillance data could have led to a substantially different interpretation of the data presented, if this method had been in use at the time the data were being evaluated by the infection control staff. The forecast model suggests that the observed variations in the data in early 1991 and late 1991 and through the first half of 1992 are consistent with the temporal fluctuations of the monthly cumulative incidence of patients with gentamicin-resistant \textit{P. aeruginosa} observed over the years 1987–1990 that preceded these periods of apparent increases in cumulative incidence. The failure of methods such as that used by Morrison et al. (11) to account for this month-to-month fluctuation of the data in terms of its time series results in narrower confidence intervals around the estimated monthly cumulative incidence and a false perception of data that are “extreme” relative to the baseline. In fact, investigations by the infection control team concluded that there was no evidence to support an outbreak and that the peaks in the cumulative incidence seen in February and May of 1992 were the result of readmission to the hospital of patients hospitalized several months previously who were known to harbor these organisms. Surveillance for gentamicin-resistant \textit{P. aeruginosa} has continued with laboratory-based reports. Monthly reports to the hospital’s infection control committee were suspended early in 1993 after no increase in isolates was detected over the previous quarter.

The method demonstrated here has been applied to surveillance in general public health (16). Other potential applications include the monitoring of recruitment into multicenter clinical trials where recruitment over time for any center is considered a stationary process and is set to remain above some lower bound of a confidence interval. Based on previous recruitment requirements for any ensuing period (e.g., month) could be calculated.

In summary, the forecast method presented here appears to be a more realistic method with which to analyze surveillance activities for hospital-acquired infections, and it makes more efficient use of the extensive data now routinely available.

**ACKNOWLEDGMENTS**

The authors thank John Harvey of the Information Resources Management Service, Department of Veterans Affairs Medical Center, San Francisco, for assistance with data management.

**REFERENCES**

**APPENDIX 1**

*S*<sub>t+1</sub> as an Unbiased Estimator

\[ S_{t+1} = \alpha x_t + (1 - \alpha)S_t \]

\[ S_{t+1} = \alpha x_t + (1 - \alpha)[\alpha x_{t-1} + (1 - \alpha)S_{t-1}] \]

\[ S_{t+1} = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k x_{t-k} + (1 - \alpha)^t S_1 \]

\[ E(S_{t+1}) = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k + (1 - \alpha)^t S_1 \]

As the sample size gets large, the expected value of the forecast at time *t* + 1 is therefore unbiased.

\[ E(S_{t+1}) = \lim_{t \to \infty} a\alpha \sum_{k=0}^{t-1} (1 - \alpha)^k + \lim_{t \to \infty} (1 - \alpha)^t S_1 \]

\[ = a\alpha \frac{1}{1 - (1 - \alpha)} = \alpha, \]

since \((1 - \alpha) < 1\).

**APPENDIX 2**

Iterative Estimation of Alpha

The present analysis makes use of the 1987–1990 monthly cumulative incidence data to forecast the 1991–1992 cumulative incidences. Alpha is estimated by iteration as follows:

**Step 1**: Set \(\alpha = 0.01\).

**Step 2**: Let \(S_1 = x_1\) = cumulative incidence for January 1987 = 26.3 (5/19).

Let \(x_2\) = cumulative incidence for February 1987 = 29.4 (5/17).

Let \(x_3\) = cumulative incidence for March 1987 = 29.4 (5/17).

**Step 3**: Calculate the forecast \((t = 2)\):

\[ S_2 = \alpha x_1 + (1 - \alpha)S_1 = (0.01)(26.3) + (0.99)(26.3) = 26.3. \]

**Step 4**: Calculate the residual forecast: \(e_2 = x_2 - S_2 = 29.4 - 26.3 = 3.1. \)

**Step 5**: Calculate the next forecast and residual \((t = 3)\):

\[ S_3 = \alpha x_2 + (1 - \alpha)S_2 = (0.01)(29.4) + (0.99)(26.3) = 26.33. \]

\[ e_3 = x_3 - S_3 = 29.40 - 26.33 = 3.07. \]

**Step 6**: Repeat step 5 for all \(t\) up to 48 months.
Step 7: Calculate the sum of squared errors (SSE) as

\[ SSE_\alpha = \sum_{r=1:1\ Jan.'87}^{r=48:Dec.'90} e_r^2. \]

Step 8: Increment \( \alpha \) by 0.01 and repeat steps 2–7.

Step 9: Choose \( \alpha \) such that the sum of squared errors is the minimum (\( \alpha = 0.16 \)).

APPENDIX 3

Application of the Forecast Method to Data from January, February, and March, 1991

The forecast estimate for January 1991 is the average of the monthly cumulative incidences over the previous 48-month baseline period (January 1987–December 1990).

Forecast cumulative incidence

\[ S_{jan91} = \frac{1}{48} \sum_{r=1:jan87}^{r=48:dec90} x_r = 24.5833 \]

January 1991 forecast residual

\[ e_{jan91} = x_{jan91} - S_{jan91} = 25 - 24.5833 = 0.4167 \]

Forecast cumulative incidence

\[ S_{feb91} = \alpha x_{jan91} + (1 - \alpha)S_{jan91} = 0.16(25) + (1 - 0.16)24.5833 = 24.65 \]

February 1991 forecast residual

\[ e_{feb91} = x_{feb91} - S_{feb91} = 58.30 - 24.65 = 33.65 \]

The upper bound of the 95 percent confidence interval for the forecast is calculated as follows:

January 1991:

\[ \hat{\sigma}_{jan91} = \sqrt{\frac{e_{jan91}^2}{1}} = 0.4167 \]

Upper bound:

\[ S_{jan91} + 1.96 \hat{\sigma}_{jan91} = 24.5833 + 1.96(0.4167) = 25.40 \]

February 1991:

\[ \hat{\sigma}_{feb91} = \sqrt{\frac{\sum e_r^2}{2}} = \sqrt{\frac{e_{feb91}^2}{2}} = \sqrt{\frac{0.4167^2 + 33.65^2}{2}} = 23.79 \]

Upper bound:

\[ S_{feb91} + 1.96 \hat{\sigma}_{feb91} = 24.65 + 1.96(23.79) = 71.30 \]

March 1991:

\[ \hat{\sigma}_{mar91} = \sqrt{\frac{0.4167^2 + 33.65^2 + 3.27^2}{3}} = 19.52 \]

Upper bound:

\[ S_{mar91} + 1.96 \hat{\sigma}_{mar91} = 30.4 + 1.96(19.52) = 68.30 \]