STATISTICS IN MEDICINE
Statist. Med. **18**, 3309–3319 (1999)

A STUDY OF THE AVERAGE RUN LENGTH CHARACTERISTICS OF THE NATIONAL NOTIFIABLE DISEASES SURVEILLANCE SYSTEM

LEWIS VANBRACKLE[1]*, AND G. DAVID WILLIAMSON[2]

[1]*Department of Mathematics, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144, U.S.A.*
[2]*Centers for Disease Control and Prevention, 2877 Brandywine Road, MS K-73, Atlanta, GA 30341, U.S.A.*

SUMMARY

This study examines the statistical properties (that is, false positive and negative signals) in detecting unusual patterns of reported cases of diseases from the Centers for Disease Control and Prevention's National Notifiable Diseases Surveillance System. Control charts are applied to the residuals of one-step ahead forecasts based on Box–Jenkins models of reported cases of disease. Simulation and analytical techniques are used to study the average run length characteristics of these control charts for various types of changes in the levels of the series, including spike, trend and step changes. The average run lengths for the highly correlated disease series are much longer than for the usual independent data case. This increase in the average run lengths is strongly influenced by the type of change in the level of the series and by the type of control chart. Understanding the average run length characteristics of the control charts can lead to timely detection of changes in the levels of disease series, and subsequent timely public health actions to decrease unnecessary morbidity and mortality. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

The National Notifiable Diseases Surveillance System (NNDSS) tracks weekly reported cases of 45 infectious diseases at the state and national levels. There is great interest in using the data from the NNDSS to detect significant aberrations, particularly increases, in the reported number of cases of any of the 45 diseases tracked by the system. Seventeen diseases thought to have good modelling potential were chosen from the NNDSS by Williamson and Weatherby.[1] Although state level data would be preferable for practical applications, the national level of reporting was chosen for ease of modelling. Williamson and Weatherby successfully identified and fit seasonal autoregressive integrated moving average (ARIMA) models (Box and Jenkins[2]) for seven of the 17 chosen disease series. The successfully modelled diseases were hepatitis A, hepatitis B, hepatitis non-A–non-B, legionellosis, malaria, meningococcal infections and tuberculosis.

This paper examines the application of statistical process control charts to the residuals from one-step ahead forecasts of ARIMA modelled time series in order to detect significant aberrations

* Correspondence to: Lewis VanBrackle, Department of Mathematics, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144, U.S.A. E-mail: Lvanbrac@kennesaw.edu

CCC 0277–6715/99/233309–11$17.50

in the level of the modelled process (Alwan and Roberts[3]). The average run length (ARL) characteristics are found for four common control charts and four types of changes in the level of the process.

The recursive relationship of the residuals, based on the results of Wardell *et al.*,[4] is used to simulate ARLs. The distribution of the residuals is derived and used in a Markov chain approach similar to that of Lin and Adams[5] to calculate ARLs analytically. The results of the simulation and the analytical methods agree quite well. The results show that the ARLs are substantially influenced by the seasonal behaviour of the models. Generally, the ARLs are much larger for the control charts applied to residuals from the ARIMA models than for the control charts applied to independent data, for which the charts were developed. Those models with the weakest seasonal behaviour have the largest ARLs.

2. CONTROL CHARTS

2.1. The Shewhart Control Chart

The simplest form of the control chart is the Shewhart chart. In a Shewhart chart for monitoring the level of a process, an observation at time t, x_t, is used to indicate whether the process has undergone some shift in its level. A value of $|x_t|$ exceeding certain control limits indicates that the process level has shifted from its previous level. Control limits are usually expressed in terms of the process standard deviation and are chosen to give the chart a good balance between failing to indicate a real shift in process level (a type II error) and indicating a shift when none has occurred (a type I error). An important characteristic of the Shewhart chart is its rapid detection of large shifts in the process level. However, the Shewhart chart is slow to detect small or moderate changes in the process level.

2.2. The moving average control chart

The moving average (MA) control chart uses the moving average of observations of the process as the control statistic and is more sensitive than the Shewhart chart to small shifts in the level of the process. The control statistic of the MA chart with span m is given by

$$y_t = \frac{x_{t-m+1} + x_{t-m+2} + \ldots + x_t}{m}.$$

The chart signals when $|y_t|$ exceeds control limits expressed as a multiple of the standard deviation of y_t. The particular multiple is chosen to give the chart good properties, as discussed for the Shewhart chart. The larger the span of the MA chart, the more sensitive the chart is to small shifts in the process level. In our study, we have used a span of two. Thus our MA chart may be expected to provide somewhat better detection of small shifts in the process level than the Shewhart chart.

2.3. The Exponentially Weighted Moving Average Control Chart

The exponentially weighted moving average (EWMA) control chart (Roberts[6] and Hunter[7]) uses the control statistic $y_t = (1-\lambda)y_{t-1} + \lambda x_t$ where $0 < \lambda \leqslant 1$ is a parameter of the chart. The chart signals that the process level has changed when $|y_t|$ exceeds control limits expressed as a multiple

of the asymptotic standard deviation of y_t. Again, the particular multiple is chosen to give the chart good properties.

The EWMA chart is sensitive to small shifts in the process level for small values of λ and to large shifts in the process level for large values of λ. Note that for $\lambda = 1$, the EWMA chart is identical to the Shewhart chart. In our study, we have used $\lambda = 0{\cdot}25$. This value of λ is commonly used in industrial applications and is within the range of 0·1 to 0·5 often suggested in the literature for detecting shifts of one-half to one standard deviation of the process.

2.4. The Cumulative Sum Control Chart

The cumulative sum (CUSUM) control chart (Page[8]) is based on sums of observations and can be sensitive to small shifts in the process level. The CUSUM chart uses one statistic for detecting a positive shift in the process level and another statistic for detecting a negative shift in the process level. The statistic for detecting a positive shift is $y_t = \max(0, y_{t-1}) + x_t - h$ where h is a parameter of the chart. This CUSUM statistic accumulates evidence of a positive shift in the level of the process. If there is no evidence of a positive shift, the CUSUM statistic resets to zero. The CUSUM signals that a positive shift in the process level has occurred when y_t exceeds the control limit. As usual, the control limit is chosen to give the CUSUM chart good properties.

A similar procedure using $y_t = \min(0, y_{t-1}) + x_t + h$ and signalling when y_t is less than the control limit is used to detect negative shifts in the process level. To detect both positive and negative shifts in the process level, we have employed the standard procedure of using two one-sided charts concurrently.

The CUSUM procedure can be derived from a sequential probability ratio test. This derivation indicates that the value of h should be chosen as one-half of the shift in the level which should be detected quickly. In our study, we used $h = 0{\cdot}25$, indicating that we want quick detection of a shift of one-half of the standard deviation of the process.

3. THE AVERAGE RUN LENGTH

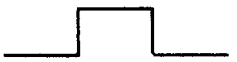
In studying the properties of control charts, the emphasis has been on determining the ARL of the chart. The ARL of a chart is the expected number of samples to be taken (in our case the expected number of weeks) before the chart indicates a shift in the process level. The ARL should be large when there has been no change in the process, but the ARL should be small when the process has undergone a change. Typically ARLs are evaluated for zero shift in the process level (in-control ARLs) and for several shift values which should be detected quickly. After consulting with epidemiologists at the Centers for Disease Control and Prevention, it was decided that in-control ARLs should range from 4 to 52 weeks, depending on the disease being studied. Control limits were chosen accordingly.

4. TYPES OF CHANGES IN THE PROCESSES

In this study, we examined four types of changes in the disease series. The four types of changes are illustrated below:

(i) The step shift in the level of the series

(ii) The impulse shift

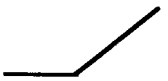


.

(iii) The spike shift

.

(iv) The trend shift



.

Step, impulse and spike shifts of height one, two and three times σ_a, the standard deviation of the residuals of the ARIMA model, were studied. Impulse shifts of 4, 8, 16 and 24 weeks duration and trend shifts ranging from 0·1 σ_a to 1·0 σ_a per week were studied.

5. SIMULATION OF AVERAGE RUN LENGTHS

5.1. The Recursive Relationship of the Residuals

The ARIMA model for a disease series contains terms relating the current week to past weeks. For example the ARIMA model for hepatitis A is

$$x_t = x_{t-1} + x_{t-52} - x_{t-53} + a_t - 0{\cdot}86a_{t-1} - 0{\cdot}79a_{t-52} + 0{\cdot}6794a_{t-53}$$

where the x's are the square roots of the number of cases of hepatitis A and the a's are random shocks, normally distributed random variables with mean zero and variance $\sigma_a^2 (a \sim N(0, \sigma_a^2))$. The square root transformation was used in the modelling process so that the normality of the random shocks would be justified. This model contains terms relating the current week, x_t, to the .previous week (x_{t-1} and a_{t-1}), as well as seasonal terms relating to the previous year (x_{t-52}, x_{t-53}, a_{t-52} and a_{t-53}).

Following the approach of Wardell *et al.*, we can derive a recursive relationship for the residuals from one-step ahead forecasts based on the ARIMA models of Williamson and Weatherby. For example, the relationship for hepatitis A is

$$\begin{aligned} R_t = {} & s_t - s_{t-1} - s_{t-52} + s_{t-53} + a_t - 0{\cdot}86a_{t-1} - 0{\cdot}79a_{t-52} + 0{\cdot}6794a_{t-53} \\ & + 0{\cdot}86R_{t-1} + 0{\cdot}79\ R_{t-52} - 0{\cdot}6794R_{t-53} \end{aligned}$$

where the s's are the levels of the shift. This equation relates the value of the residual at the present time period, R_t, to present values of the shift and the random shock, s_t and a_t, and to past values of the shift (s_{t-1}, s_{t-52} and s_{t-53}), the random shocks (a_{t-1}, a_{t-52} and a_{t-53}) and the residuals (R_{t-1}, R_{t-52} and R_{t-53}). The presence of seasonal terms in the ARIMA model for hepatitis A leads to the presence of seasonal terms in this relationship.

5.2. The Dynamic Behaviour of the Residuals

The above expression for the residuals illustrates the dynamic response of the residuals to a shift in the process level described by Wardell *et al.* ARIMA models are adaptive forecasting models.

When an ARIMA process undergoes a shift, the expected value of the forecast converges to a new equilibrium level. The expected values of the residuals from the one-step ahead forecasts also converge to a new equilibrium level, smaller in magnitude than the level of the shift. In the above expression, the effect of the shift quickly disappears due to the $s_t - s_{t-1}$ term. The expected values of the residuals converge because the absolute values of the coefficients of the random shock and residual terms are less than one.

The seasonal terms in our ARIMA model introduce an echo effect in the behaviour of the residuals. In the recursive relationship of the residuals, the shift value from a season in the past, s_{t-52}, appears, but it is reduced by the previous value of the shift, s_{t-53}. Values of the random shocks and residuals from a season in the past also appear, but with coefficients that are less than one in absolute value. After experiencing this echo shift, the expected values of the residuals converge to a value that is smaller in magnitude than the shift.

5.3. Simulation

The normally distributed a_t's were simulated. We then made use of the recursive relationship of the residuals to generate R_t, the residual at time t, and subsequently the control statistic based on the residual. The run length, the number of steps taken until the control statistic exceeded the control limits, was noted. The procedure was repeated 10,000 times, and the mean of those 10,000 run lengths was reported as the average run length.

6. ANALYTICAL CALCULATION OF AVERAGE RUN LENGTHS

6.1. The Distribution of the Residuals

In addition to the simulation method described in the previous section, we developed an analytical method for calculating the ARLs of our control charts. Recognizing the recursive relationship for the residuals as a difference equation for the unknown residual R_t, we found a series solution for R_t. The series solution is $R_t = a_t + s_t + \sum_{j=1}^{t} \psi_j s_{t-j}$, where the ψ coefficients depend on the ARIMA model. Since $a_t \sim N(0, \sigma_a^2)$, we have $R_t \sim \mathrm{N}(\mu_t, \sigma_a^2)$, where $\mu_t = s_t + \sum_{j=1}^{t} \psi_j s_{t-j}$. We used this distribution of R_t to calculate ARLs for each of the four control charts as discussed in the next four sections.

6.2. ARLs for the Shewhart Control Chart

Since the run length (RL) of a control chart is a non-negative discrete random variable, we can write $\mathrm{ARL} = \sum_{t=1}^{\infty} P(\mathrm{RL} > t)$ where $P(\mathrm{RL} > t)$ is the probability that the run length exceeds t. For the Shewhart chart based on the one-step ahead residuals, R_t, from the ARIMA model, with control limits at $\pm$ CL, we have

$$P(\mathrm{RL} > t) = P(|R_0| \leqslant \mathrm{CL}) P(|R_1| \leqslant \mathrm{CL}) \ldots P(|R_{t-1}| \leqslant \mathrm{CL}) = \prod_{i=0}^{t-1} P(|R_i| \leqslant \mathrm{CL}).$$

Thus

$$\mathrm{ARL} = \sum_{t=1}^{\infty} \prod_{i=0}^{t-1} P(|R_i| \leqslant \mathrm{CL}).$$

Using the fact that $R_i \sim N(\mu_i, \sigma_a^2)$, where $\mu_i = s_i + \sum_{j=1}^{i} \psi_j s_{i-j}$, we can calculate the $P(|R_i| \leqslant \mathrm{CL})$ and subsequently the ARL for the Shewhart chart for each disease.

The infinite sum in the above expression for the ARL of the Shewhart chart must be approximated by a finite sum in the actual calculation. In the program written for this calculation, the sum was terminated when $\prod_{i=0}^{t-1} P(|R_i| \leqslant \mathrm{CL}) < 10^{-5}$.

6.3. ARLs for the Moving Average Control Chart

The ARL of the MA chart with span two is calculated in a manner similar to that of the Shewhart chart. The probabilities involved are more complicated due to the nature of the MA chart statistic, but the basic approach is the same.

6.4. ARLs for the Exponentially Weighted Moving Average Control Chart

For the EWMA control chart based on the one-step ahead residuals from the ARIMA model, a Markov chain approach similar to that of Lin and Adams is used. The distribution of the residuals derived in Section 6.1 is used to calculate the transition probabilities in the Markov chain. The approach is complicated in our problem by the seasonality of our models, but a modification of the Lin and Adams technique can be used to calculate the ARLs for the EWMA chart.

6.5. ARLs for the Cumulative Sum Control Chart

The average run lengths for the CUSUM control chart were calculated using an approach similar to that used for the EWMA chart. The only difference in the approaches is the use of one CUSUM chart to detect upward shifts in the series and another CUSUM chart to detect downward shifts in the series. The overall average run length for the CUSUM chart was calculated in the usual way, using

$$\frac{1}{\mathrm{ARL}} = \frac{1}{\mathrm{ARL}_{\mathrm{up}}} + \frac{1}{\mathrm{ARL}_{\mathrm{down}}}.$$

The ARLs calculated using the above analytical methods agree quite well with those from the simulation described earlier.

7. RESULTS

The general effect of correlation of the data series on the ARLs of control charts has been described previously by Johnson and Bagshaw,[9] Bagshaw and Johnson,[10] Montgomery and Mastrangelo,[11] Wardell *et al.*[12] and VanBrackle and Reynolds.[13] In addition, the properties of control charts based on residuals from one-step ahead forecasts of some simple time series models have been described by Superville and Adams,[14] Runger *et al.* [15] and Wardell *et al.*[16] In general the effect of positive correlation of the data leads to shortened in-control ARLs (a higher false alarm rate) of control charts based on the data, if the usual control limits are used. If the control limits are adjusted to take the correlation into account and achieve the desired in-control ARL, the effect of the correlation is to delay detection of a shift in the level of the process.

Control charts based on residuals of one-step ahead forecasts of time series models exhibit somewhat different behaviour. The independence of the residuals yields in-control ARLs which are identical to those for control charts based on independent data. However, the adaptive behaviour of the residuals discussed above leads to delayed detection of shifts in the level of the

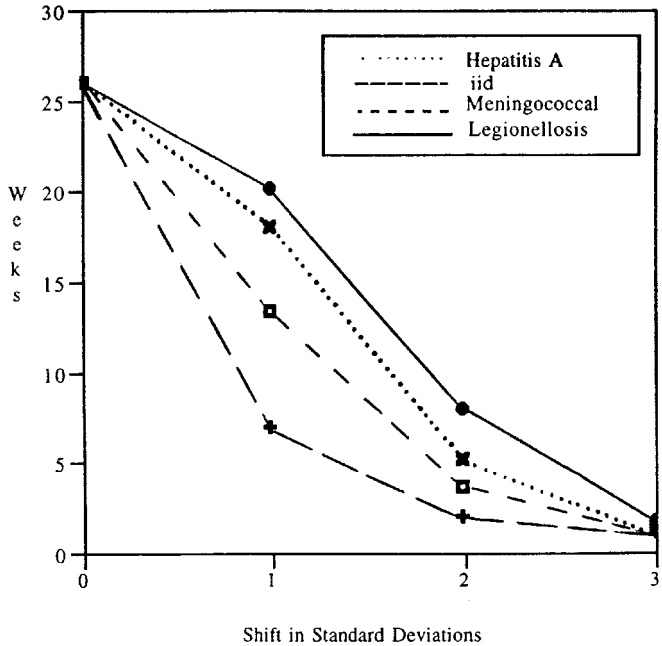


Figure 1. ARLs for the Shewhart chart for a step shift

process. Control charts based on residuals have a window of opportunity for detecting process shifts. The convergence of the expected value of the residuals to a value lower than the shift value gives charts based on residuals a decreased probability of detecting process shifts after the first few time periods following the shift. The seasonality of our ARIMA models does give the chart based on residuals a second or third chance to detect a process shift one or two seasons after the shift, but these second and third chances come too late to be of help in the rapid detection of shifts. Consequently, control charts based on residuals from ARIMA models can have out-of-control ARLs that are unacceptably high.

The results shown in the following figures illustrate the influence of the adaptive behaviour of the residuals on the ARLs of control charts based on the one-step ahead forecast residuals.

Figure 1 shows the ARLs calculated for the Shewhart chart for the independent, identically distributed (i.i.d) data for which the chart was designed and for the residuals from the models for hepatitis A, legionellosis and meningococcal infections. The control limits are chosen to give an in-control ARL of 26 for all of the models. The shift is the step shift measured in units of σ_a, the standard deviation of the residuals from the ARIMA model. The influence of the adaptive behaviour of the residuals is clear; the ARLs for each of the disease models is higher than that for the i.i.d. series at each of the shift levels. For a shift of one standard deviation, the ARLs for the disease models are roughly two to three times the ARL of the i.i.d. series. For a shift of two standard deviations, the disease model ARLs range from two to five times the i.i.d. ARL. This effect is much smaller for the shift of three standard deviations, since the likelihood of detecting such a large shift within the first few time periods after the shift is so high that the adaptive behaviour of the residuals can have little effect.

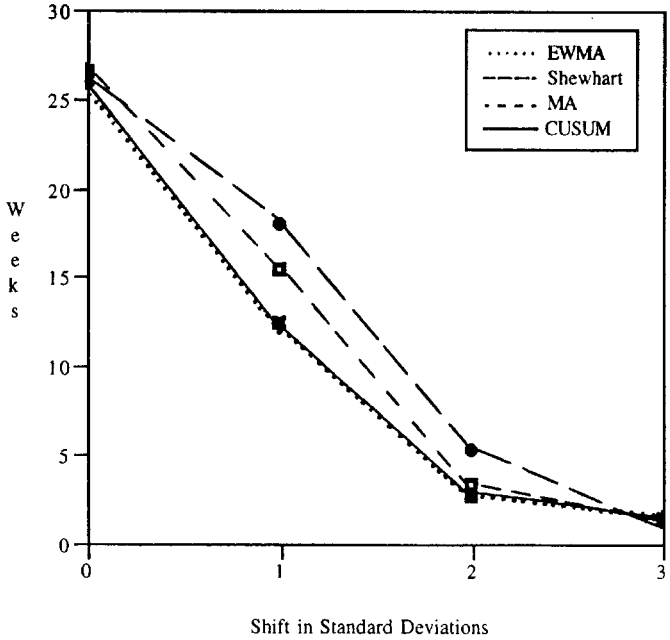


Figure 2. ARLs for hepatitis A for a step shift

It is interesting to note the differences in the behaviours of the different models. The model for legionellosis contains no seasonal terms. As a result, the adaptive behaviour of the residuals for this model is strongest, with no echo effect due to seasonality. Consequently, the Shewhart chart based on the residuals from the legionellosis model has the worst ARL behaviour of the models shown in Figure 1.

As discussed in Section 5.2, the form of the model for hepatitis A causes the residuals to adapt rather rapidly. The seasonal behaviour gives the residuals a second chance to detect the shift before the residuals again adapt. Thus, the Shewhart chart based on the residuals from the hepatitis A model has slightly lower ARLs than the chart for the legionellosis model.

The meningococcal infections model has weaker adaptive behaviour than the hepatitis A model. The residuals from this model do not adapt very rapidly until after the first season (52 weeks). Since the residuals from this model have the weakest adaptive behaviour, the Shewhart chart based on the residuals from the meningococcal infections model has the best ARL characteristics of the disease models in Figure 1.

The remaining results in Figures 2 to 4 are shown for the hepatitis A model. Results are shown for all four types of shift in the level of the process and for all four types of control charts. The other disease series have similar results, differing only in severity according to the complexity and the degree of seasonality of the model.

Figure 2 shows the ARLs for the step shift. Note that the ARLs are all greater than those for the i.i.d. series in the previous figure. It is clear from this figure that the control charts which are designed to accumulate evidence of a shift in the process, the EWMA and the CUSUM charts, have superior detection ability in this situation. The EWMA and CUSUM ARL characteristics

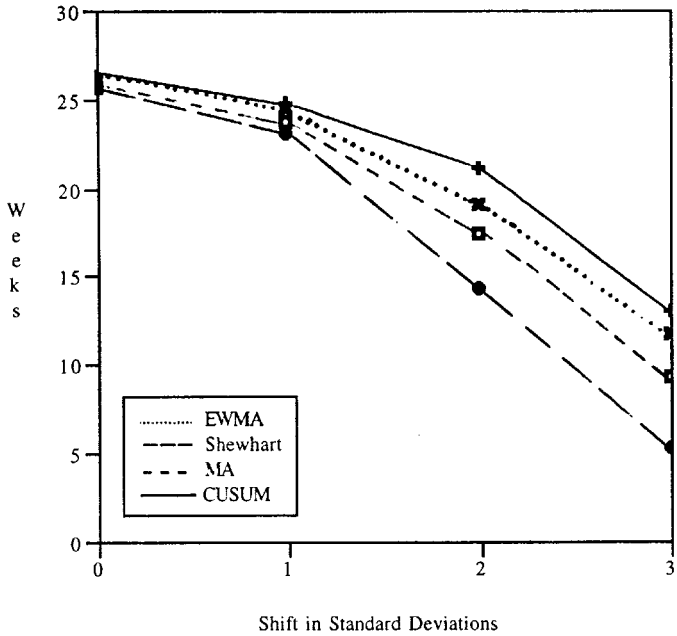


Figure 3. ARLs for hepatitis A for a spike shift

are almost identical, and their lines on the figure are nearly indistinguishable. The MA chart, with its short term accumulation of evidence, outperforms the Shewhart chart. The Shewhart chart has the longest detection times of all, since the adaptive behaviour of the residuals quickly disguises the shift in the process level.

The relative ARL characteristics of the control charts for an impulse shift are similar to those of Figure 1. The ARLs are slightly lower than those for the step shift. The impulse shift consists of two shifts in the process level, a shift up followed by a shift back down to the original level some weeks later. This double shift gives the control charts two opportunities to detect the shift in the process level. The residuals must adapt themselves to both shifts, and the control charts detect the change slightly more quickly than for the step shift.

Figure 3 shows the ARL behaviour for the four control charts for a spike shift, a shift of one week duration. Note the reversal of the ordering of the control charts in this figure. While all of the charts perform rather badly for this type of shift, the accumulating charts (MA, EWMA and CUSUM) perform worse than the Shewhart chart. The spike shift gives no opportunity for the accumulation of evidence of a shift in the process level. As the size of the shift increases, the difference in the detection ability of the charts becomes even more pronounced. The region of large shift is where the Shewhart chart typically performs best. With no evidence to accumulate, the other charts cannot compete with the Shewhart chart.

Finally, Figure 4 illustrates the ARL characteristics of the control charts for a trend shift. The shift level is measured from 0·1 σ_a per week to 1 σ_a per week. For the smaller values of the trend shift, less than 0·5 σ_a per week, the ordering of the charts is as it was for the step shift. The EWMA and CUSUM charts are again nearly identical in their ARL characteristics. The difference

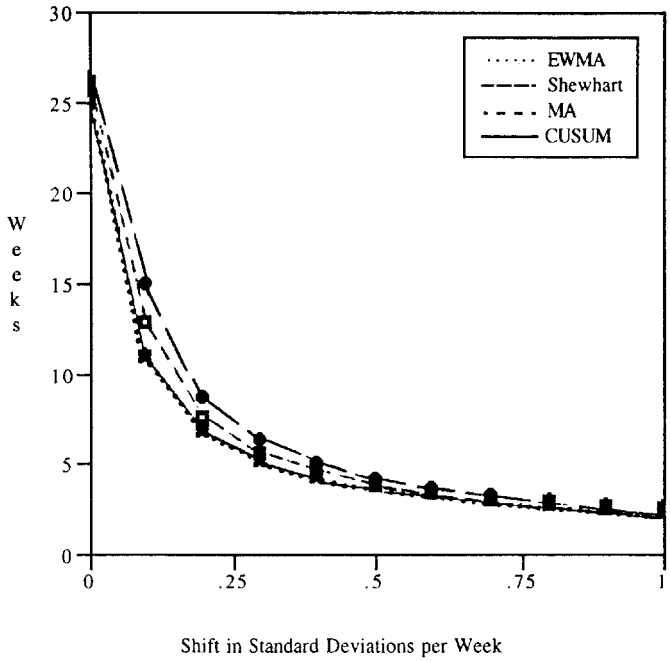


Figure 4. ARLs for hepatitis A for a trend shift

between the ARLs of all the charts is small, and they all perform well. For larger values of the trend shift, the Shewhart chart performs slightly better than the others, but all of the charts detect such large trend shifts quite well.

8. DISCUSSION

The ARL behaviour described above applies to the seven successfully modelled national level disease series only. Much work remains to be done, both for those diseases already modelled and for the other diseases monitored in the NNDSS. For the diseases already modelled, the effects of varying the control chart parameters remain to be studied. The combined Shewhart–EWMA chart was shown by Lin and Adams to be more effective than the traditional control charts for simpler non-seasonal models. The usefulness of this chart for our more complicated seasonal models remains to be evaluated. In addition, the characteristics of control charts applied to state level data and to those disease series for which no good model could be found need to be examined.

Other approaches to the modelling and detection problem also remain to be tried. A combined spatial and time series modelling approach may yield useful results by detecting not only the time of the aberration in a disease series, but also the location of that aberration. Bayesian modelling, change point theory and neural net modelling approaches have yet to be tried on this problem, but may have much to offer in this context.

REFERENCES

1. Williamson, G. D. and Weatherby Hudson, G. 'A monitoring system for detecting aberrations in public health surveillance reports', *Statistics in Medicine*, **18**, 3283–3298 (1999).
2. Box, G. E. P. and Jenkins, G. M. *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976.
3. Alwan, L. C. and Roberts, H. V. 'Time-series modeling for statistical process control', *Journal of Business and Economic Statistics*, **6**, 87–95 (1988).
4. Wardell, D. G., Moskowitz, H. and Plante, R. D. 'Run-length distributions of special-cause charts for correlated processes', *Technometrics*, **36**, 3–16 (1994).
5. Lin, W. S. W. and Adams, B. M. 'Combined control charts for forecast-based monitoring schemes', *Journal of Quality Technology*, **28**, 289–301 (1996).
6. Roberts, S. W. 'Control chart tests based on geometric moving averages', *Technometrics*, **1**, 239–250 (1959).
7. Hunter, J. S. 'The exponentially weighted moving average', *Journal of Quality Technology*, **18**, 203–210 (1986).
8. Page, E. S. 'Continuous inspection schemes', *Biometrika*, **41**, 100–114 (1954).
9. Johnson, R. A. and Bagshaw, M. 'The effect of serial correlation on the performance of CUSUM tests', *Technometrics*, **16**, 103–112 (1975).
10. Bagshaw, M. and Johnson, R. A. 'The effect of serial correlation on the performance of CUSUM tests II', *Technometrics*, **17**, 73–80 (1976).
11. Montgomery, D. C. and Mastrangelo, C. M. 'Some statistical process control methods for autocorrelated data', *Journal of Quality Technology*, **23**, 179–193 (1991).
12. Wardell, D. G., Moskowitz, H. and Plante, R. D. 'Control charts in the presence of data correlation', *Management Science*, **38**, 1084–1105 (1992).
13. VanBrackle, L. N. and Reynolds, M. R., Jr. 'EWMA and CUSUM control charts in the presence of correlation', *Communications in Statistics – Simulation and Computation*, **26**, 979–1008 (1997).
14. Superville, C. R. and Adams, B. M. 'An evaluation of forecast-based quality control schemes', *Communications in Statistics – Simulation and Computation*, **23**, 645–661 (1994).
15. Runger, G. C., Willemain, T. R. and Prabhu, S. 'Average run lengths for CUSUM control charts applied to residuals', *Communications in Statistics – Theory and Methods*, **24**, 273–282 (1995).
16. Wardell, D. G., Moskowitz, H. and Plante, R. D. 'Run-length distributions of residuals control charts for autocorrelated processes', *Journal of Quality Technology*, **26**, 308–317 (1994).

