International Journal of Epidemiology 1998;27:282–288

Multivariate time series analysis in nosocomial infection surveillance: a case study

Cristina Fernández-Pérez,[a] Juan Tejada[b] and Miguel Carrasco[c]

Background The present study describes the use of time series analysis in the evaluation of the incidence of nosocomial infection. The main hypothesis analysed was that monthly occurrence of nosocomial infection in a hospital may be related to work-related factors such as the control and training of personnel imposed by a surveillance system, strikes supported by medical personnel and movement of personnel. Time series analysis was used to quantify, model and statistically evaluate these interventions.

Methods The data employed (March 1982–December 1990) were supplied by the nosocomial infection surveillance system of a primary-care general hospital. The monthly time series incidence of nosocomial infections (measured as percentage cumulative incidence) was analysed by curve fitting, autoregressive, integrated and moving average (ARIMA) modelling (Box-Jenkins) and intervention and dynamic regression analysis.

Results The imposed control and training of personnel by the surveillance system was associated with a 3.63% decrease in the accumulated monthly incidence of nosocomial infection from 7.82% to a baseline level of 4.19%. There was a strong indication that an increase of infection incidence of 4.34% corresponded to a medical strike. This increase was maintained over the following months raising the baseline level to 4.84%. An increase of 0.18% was associated with each new nursing contract. Evidence was obtained for the possible relationship between incidence of nosocomial infection and vacation periods.

Conclusions The results suggest the need for strict control of the activities of hospital personnel and for the adoption of certain preventative measures during vacation periods to avoid an undesirable increase in the incidence of nosocomial infections.

Keywords Intervention studies, nosocomial infection, surveillance, time series, ARIMA models, quality control

Accepted 28 July 1997

Nosocomial infections are those which develop in a hospital or are produced by microorganisms acquired during hospitalization.[1] Public health surveillance helps identify possible determinants of nosocomial infections and their relative importance, which may be of use in the design and implementation of preventative programmes.[2]

Incidence studies of nosocomial infection are performed to evaluate control measures, clusters of unusual cases or the surveillance system itself. To this end, information concerning the event studied such as variables measured at specific time-intervals is required.

Throughout the1980s there was a marked increase in the use of statistical techniques in public health surveillance.[3] Descriptive statistic techniques and statistical tests based on the assumption of given probability distributions were used in addition to the classical regression models.[4–7] However, temporal dependence in longitudinal data obtained from the same hospital suggests that statistical time series methods are appropriate for the analysis of epidemiological surveillance.[8–10] Statistical tools generally used for the analysis of time series include autoregressive, integrated and moving average (ARIMA) modelling using the methodology of Box and Jenkins, intervention analysis and the transfer function or dynamic regression models. These methods are based on the existence of autocorrelation within the series and try to approach the random process which generates the

[a] Servicio de Medicina Preventiva, Hospital Universitario San Carlos, C/ Martin Lagos s/n, 28040 Madrid (Spain).

[b] Department of Statistics and Operations Research, Universidad Complutense, Madrid, Spain

[c] University Center for Public Health, Universidad Autonoma, Madrid, Spain

observations as well as to isolate some deviations which are not directly observable.

Time series analysis has been used in public health surveillance studies of infectious diseases.[11–18] However, to date no application of multivariate time series analysis to nosocomial infections has been described.

The present study uses time series analysis to evaluate nosocomial infections in a primary-care general hospital from 1982 to 1990. The general aim was to quantify and model interventions hypothesized to affect nosocomial infection incidence. The more specific aims were as follows: to describe the chronological evolution of the frequency of nosocomial infection; to quantify the changes associated with control measures and personnel training implemented by the surveillance system; to establish the baseline level of incidence of infection attained by the surveillance system; to evaluate a cluster of unusual cases produced during a strike of medical personnel; and to determine the statistical relationship between infection and factors related to personnel movement (calendar effect, midweek public holidays, Easter holidays and hiring of temporary personnel). Such factors have often been associated with variation in industrial productivity.[19]

The present model considers nosocomial infection incidence as an outcome variable of hospital quality in terms of system design and care given to patients.[5,20–22] It is proposed that nosocomial infection incidence may be used as an a quantitative measure of an aspect of patient care that can be used as a guide to monitor and evaluate the quality and appropriateness of health care delivery.[22]

Materials and Methods

Description of data

The Guadalajara general hospital (National Health System of Spain) covers the medical needs of the 150 000 inhabitants of the province of Guadalajara. The hospital has 406 beds, each ward containing 28 beds distributed in twin or single rooms. A team of three or four nurses (typically two nurses and two assistant nurses) is assigned to each ward and shift.

When the hospital was opened in March 1982, a continuous, general 'Nosocomial Infection Surveillance System' was established.[23] This model has the same characteristics in terms of design and evaluation as the one developed by the Centers for Disease Control and Prevention (CDCP) and implemented by the National Nosocomial Infections Surveillance System (NNIS).[24]

The present study analyses surveillance data obtained from March 1982 to December 1990. The monthly cumulative incidence of nosocomial infections (I_t) was defined as the number of new infections produced within a month divided by the number of patients who were hospitalized for at least 24 hours during this month, expressed in per cent.[25]

In September 1982 a urinary tract infection prevention programme including personnel training was introduced. This programme was maintained until December 1986. From March to June 1987 a medical strike in the public sector involved hospital personnel. Infection data corresponding to October 1987 were not available. Numbers of temporary nursing employees during the period of study were obtained from the hospital personnel data bank.

Modelling method

A stochastic model was proposed to explain the dynamic evolution of I_t, using a stochastic difference equation with respect to its own past, qualitative interventions and other relevant quantitative variables. This model was obtained by curve fitting, ARIMA modelling (Box-Jenkins) and intervention and dynamic regression analysis techniques.[26–30]

Due to the limited size of the time series, simultaneous inclusion of all the factors and variables in the model was not performed since the lack of degrees of freedom would hinder correct interpretation of their effects. The strategy we propose (in some ways similar to stepwise regression methodology) was as follows:

(1) To build a BASIC MODEL incorporating the statistical effect of the urinary tract infection prevention programme and a medical strike over an ARIMA modelling. The model would also incorporate a component to help estimate the value of the missing data.

(2) Starting from this BASIC MODEL the statistical effects of other economic factors such as the calendar effect, Easter holidays, other vacations and hiring of temporary personnel could be incorporated into the model.

Basic Model

Descriptive analysis

An initial examination of time series data and smooth series data graphics through moving averages of a twelfth order (to isolate the tendency and eliminate possible seasonal variations and reduce noise) was performed to establish preliminary theories concerning the way in which main historical events could influence the incidence of nosocomial infection. Some descriptive statistics were also employed.

Modelling of the training period (September 1982–December 1986)

The absence of data prior to implementation of the surveillance system made it advisable to model its potential effect through a deterministic component. The logistic curve solving the differential equation 1 was proposed.

$$\frac{dD(t)}{dt} = k(M - D(t))\,(M' - D(t)) \qquad (1)$$

$$D(t_0) = \frac{M + M'}{2}$$

This equation models the decrease of the deterministic component of the infected population D(t) from an initial level M to a baseline level M'. The parameters M, M', k y t_0 are estimated by a minimum square fitting to a 12-smooth series of the data previous to the strike period.

The analysis and fitting of models was performed over the residuals of the minimum square estimation of the deterministic component termed, X_t , i.e. if D_t is the discrete time fitted deterministic component, then $X_t = I_t - D_t$. It was initially supposed that $X_t = S_t + M_t + N_t$, where S_t is the strike component, M_t is due to the missing datum and N_t is an ARIMA noise.

Modelling of the strike

From March 1987 to June 1987 a medical strike in the public sector involved 99% of hospital personnel. Modelling of the strike was performed by the introduction of an intervention variable, H_t, defined as the number of strike days expressed as a proportion with respect to the total number of days in the month: 0.27 in March and April, 0.77 in May, 0.63 in June of 1987 and zero for the remaining months. The model was constructed to fit the hypothesis that a large increase in the number of infections occurred at the start of the strike which gradually declined influencing later periods. The effect was transferred to the series by the component,

$$S_t = (\omega_1/(1 - \delta B) + \omega_2/(1 - B))\, H_t$$

where B is the backward operator ($BX_t = X_{t-1}$) and where $\omega_1 + \omega_2$ represent the direct effect on infection incidence of the strike in a given month in which the strike affects every day. δ is the residual effect of the intervention given by the rate of observed infections in a month which is translated to the next.[30]

Estimation of the missing datum

The estimation of the datum that was missing for October 1987 was obtained by substituting this by a zero in the series. A component $M_t = \omega P_t$ was then incorporated into the model, where P_t equals 1 if t corresponds to October 1987 and 0 to the other months. Finally, if $\hat{\omega}$ is the estimate of ω, then $-\hat{\omega}$ is the estimate of the missing datum.

ARIMA modelling

For the noise N_t, an ARIMA model built according to Box-Jenkins methodology was proposed.[26] A general ARIMA model is given by equation 2.

$$\phi(B)\, \Phi(B^s)\, \nabla^d\, \nabla_s^D\, N_t = \Theta(B)\, \Theta(B^s)\, a_t \qquad (2)$$

where B is the backward operator $BX_t = X_{t-1}$, $B^k X_t = X_{t-k}$, $\nabla^d = (1 - B)^d$, $\nabla_s^D = (1 - B^s)^D$ and ϕ, Φ, Θ, Θ are polynomials of B and B^s, respectively, a_t is a white noise process and s is the seasonal period (in this case s = 12).

The basic model elaborated is $I_t = D_t + S_t + M_t + N_t$, provided that all the effects are significant.

Other Components

Calendar effect

Many economic variables related to production are influenced by the so-called calendar effect i.e. by the different composition of each month in terms of days of the week. For instance, the month of May in two different years obviously has the same number of days, but in general has a different composition with regard to the number of Mondays, Tuesdays, and so on.

The procedure followed in order to model this effect was that of Hillmer[19] in which the variables DM_t, DT_t, DW_t, DTH_t, DF_t, DS_t and $DMTH_t$ are built. The number of Sundays is considered to be a baseline variable and thus the first six take values equal to the difference between the number of Mondays, Tuesdays, Wednesdays, Thursdays, Fridays, Saturdays and the number of Sundays, respectively in the month t. $DMTH_t$ is the total number of days in each month. These variables are incorporated into the basic model through the additive component.

$$EC_t = c_1DM_t + c_2DT_t + c_3DW_t + c_4DTH_t + c_5DF_t + c_6DS_t + c_7DMTH_t$$

Effect due to Easter holidays

The corresponding variable E_t has a value of zero every month except for March and April. In these two months the value of E_t is established proportionally to the length of the Easter period corresponding to each month. Easter week and the following week are the two reference weeks. Each day is given a value of 1, with the exception of Saturdays and Sundays, which are given a value of 0.5. The total sum is 13. The sum of the days corresponding to each month is divided by 13.

Effect due to midweek public holidays

This effect tries to estimate the possible influence of midweek holidays (a day of the week that is not Saturday or Sunday) on the occurrence of nosocomial infection. For this purpose, the variable MPH_t each month depends on the number of midweek holidays where each day is given a different value (0.75 Monday and Friday, 1 Tuesday and Thursday and 0.5 Wednesday). In this way the relationship between certain months and incidence of infection may be accounted for. For instance, December and September (local holiday) contain long holiday periods.

Effect due to hiring

In this case, the influence of inexperience or lack of training is incorporated into the model using the number of new contracts for temporary nursing personnel per month. This variable is divided into the variable representing the number of people hired for the first time and the variable representing the number of people who had already been previously hired.

Statistical analysis

Statistical analysis of data was performed using the computer programs DBASE III+, Statgraphics 6.0 and SCA.

Results

Basic model

Descriptive analysis

The cumulative incidence of nosocomial infection during the study period was 5.43%. A total of 5025 nosocomial infections were detected. Four periods were identified by the surveillance system which are specified in Figure 1. The first one covers the implementation of the surveillance system in 1982 to the beginning of 1984. This period corresponds to the stage of personnel training and implementation of the surveillance system in the hospital and to the implementation of a urinary tract infection control programme and shows a high incidence of infection. The second, from January 1984 to May 1987 represents a period of stabilization with a minimum incidence of infection. In the third period, (May–June 1987) a sudden increase in incidence was produced which reaches a level of 10.16% and falls to 8.86% and 8.41% over the two following months respectively. From 1988 to the end of the study the level increase produced after the strike was maintained.

Modelling of the training period

The fitting of the curve was performed in the moving average 12-smooth series (Figure 2). The parameters calculated are shown in Table 1. The initial level of the system (M) showed an incidence of 7.82%, and a baseline level (M') of 4.19%. The surveillance system, therefore, is associated with a decrease of 3.63% incidence.

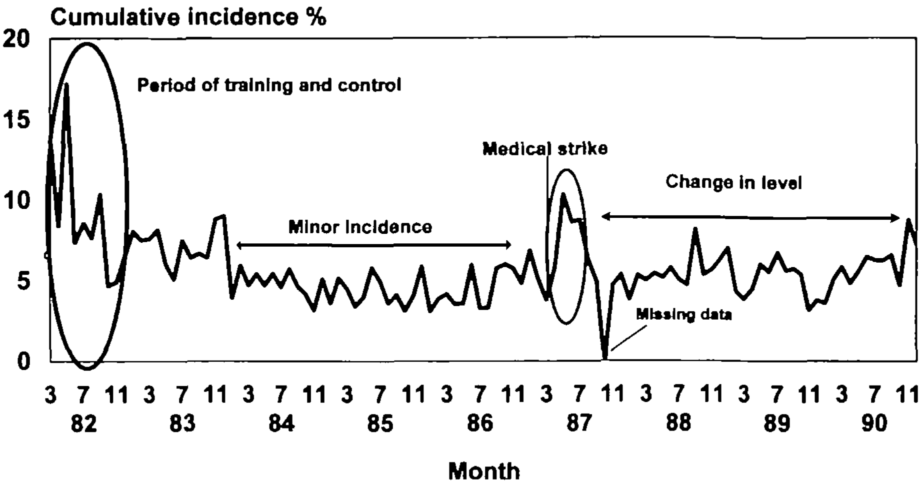


Figure 1 Monthly cumulative incidence of nosocomial infection in Hospital General de Guadalajara, Spain, between March 1982 and December 1990 (n = 106 months)

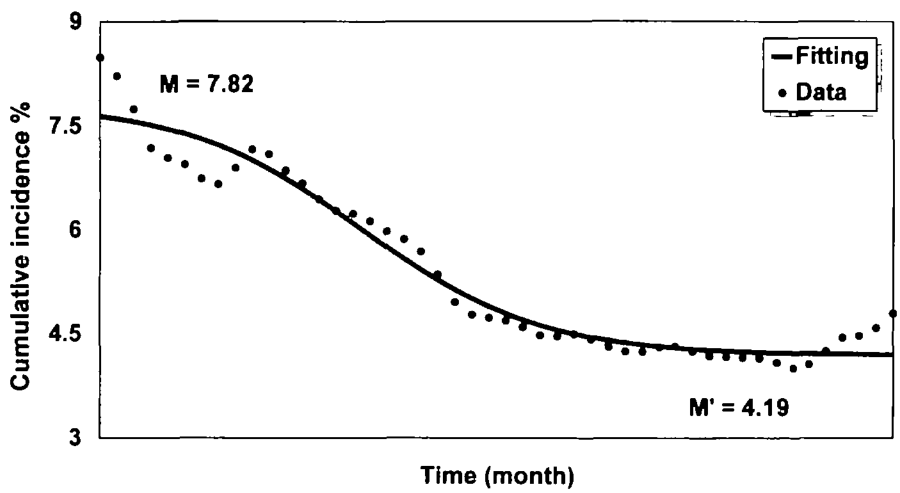


Figure 2 Modelling of the training period. Nosocomial infection smooth series data graphics through moving averages of a twelfth order and modelling with a logistic curve. The parameters M (initial level), M' (baseline level), k and t_0 are estimated by a minimum square fitting to a 12-smooth series of the data previous to the strike period (n = 48 months). Hospital General de Guadalajara, Spain

ARIMA, strike and missing datum model

If one considers the residuals generated by the fitting throughout the period analysed, it can seen that there is still some information left since there was significant variation in their level. The ARIMA model identified was:

$$X_t = (1 - \theta B)\,(1 - \Theta_1 B^{12})\,(1 - \Theta_2 B^{24})\,a_t + (\omega_1/(1 - \delta B) + \omega_2/(1 - B))H_t + \omega P_t$$

including the strike and missing datum model.

The results of the estimation of the modelling parameters are shown in Table 2. The statistical effect of a month in which the strike affects every day ($\omega_1 + \omega_2$) shows an incidence increase of 4.34%. The residual effect of the intervention given by the rate (δ) of observed incidence in a month which is translated to the next is 49%.

The estimated value of the infection incidence in October 1987 (ω) was 5.01%. All the parameters of the modelling are significant.

The autocorrelation (ACF) and partial autocorrelation (PACF) functions of the residuals showed no significant values (Figure 3). Furthermore, the mean of the residuals was not significantly different to 0 and their variance was constant, confirming the adequacy of the model.

Other Components

The basic model obtained allowed comparison of the statistical effect of other variables on nosocomial infection. Final estimates, including only significant variables are shown in Table 3. In this case, an AR(12) model was used for the noise.

Table 1 Modelling of the training period. Results of fitting 12-smooth series of the data previous to the strike period. Hospital General de Guadalajara. Spain. (n = 48 months)

Parameter	Incidence (%)	Standard error	*P*-value
M	7.82	0.23	<0.001
M′	4.19	0.09	<0.001
K	0.05	0.01	<0.001
t_0	16.53	0.96	<0.001

M, initial level, M′, minimum level; K, variation rate and t_0, the time at which the process is in level (M + M′/2)

Table 2 Time series analysis (ARIMA[a] models) of cumulative incidence of nosocomial infection obtained after removing the deterministic training component, with intervention effect estimates. Hospital General de Guadalajara. March 1982–December 1990. Spain

Parameters		Estimates	P-value
Strike	ω_1	3.69	<0.001
Strike	δ	0.49	0.006
Change in level	ω_2	0.65	<0.001
Missing datum	ϕ	-5.01	<0.001
MA(1)[b]	θ	–0.22	0.031
MA(12)[c]	Θ_1	0.22	0.029
MA(24)[d]	Θ_2	–0.28	0.002

[a] ARIMA, autoregressive, integrated, and moving average.
[b] MA(1), moving average parameter, first order.
[c] MA(12), seasonal moving average model using 12-month lag
[d] MA(24), seasonal moving average model using 24-month lag.

The variables used for the evaluation of the calendar effect were not significant with the exception of the number of days in each month ($DMTH_t$) ($P = 0.01$). No effect of midweek public holidays was detected. A negative value was obtained with respect to the effect of the Easter break showing a 0.50% decrease in incidence of infection during this period.

Table 3 Time series analysis (ARIMA[a] models) of cumulative incidence of nosocomial infection obtained after removing deterministic training and including other components. Hospital General de Guadalajara. March 1982–December 1990. Spain

Parameters		Estimates	P-value
Strike	ω_1	4.08	<0.01
Strike	δ	0.47	<0.01
Change in level	ω_2	0 43	<0.01
Missing datum	ϕ	–4.47	<0.01
Easter effect	τ	–0.50	0.10
New employment	υ	0.18	<0.01
Outlier	ρ	3.84	<0.01
September	λ	0.63	0.01
AR(12)[b]	Φ	–0.51	<0.01

[a] ARIMA, autoregressive, integrated, and moving average.
[b] AR(12) seasonal autoregressive model using 12-month lag.

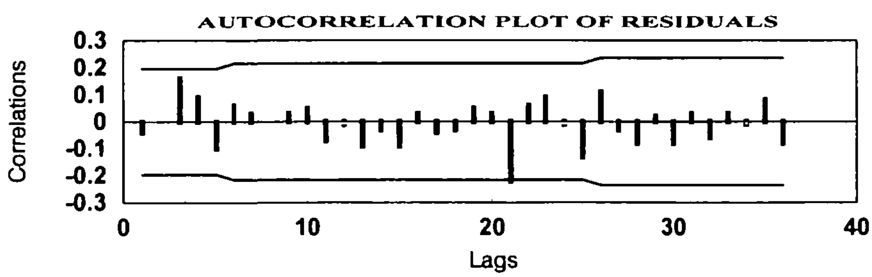
A significant increase (0.63%) was observed in September corresponding to a one-week local holiday.

The last variable incorporated into the model was the number of new nurses hired without specific training during the month. A positive relationship between the number of new employees and the occurrence of nosocomial infection (0.18% increase per new contract) was detected.

Finally, an outlier with respect to the fitted model in observation number 22 (December 1983) was automatically detected by the software (SCA).

Discussion

Epidemiological surveillance, particularly of transmissible diseases, is one of the more traditional public health activities. A recent contribution to its efficiency is the use of multivariate time series analysis to determine a relationship between incidence of



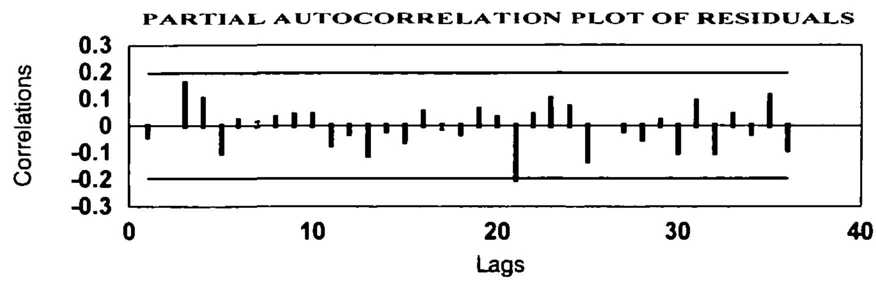


Figure 3 Diagnostic checking. Simple autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals from the seasonal autoregressive, integrated and moving average (ARIMA) fitting model. Hospital General de Guadalajara, Spain

infection and independent variables. This method permits the development of hypotheses to explain and anticipate the dynamics of the observed phenomenon and permit the establishment of a quality control system.[8,10]

The present study endeavours to explain some aspects of the dynamics of nosocomial infection. To this end, a time series of nosocomial infection incidence in a medium-size hospital was analysed. The method chosen is of ecologic type where data is accumulated and the analysis unit is a time fraction. This kind of study is applicable when it is not possible to conduct individual case studies. There is much controversy in the literature over the limited possibility of inference provided by ecologic studies although it is well known that this type of study is a good source of hypotheses.[31]

Two factors should be taken into consideration when associating the results of accumulated incidence of nosocomial infection with the different aspects of the quality of hospital care: firstly, risk factors intrinsic to the patient (age, severity of illness, ...) and secondly, a criterion which reflects the validity of the technique used to measure the indicator. In order to verify the absence of biases produced by a poor classification system, the surveillance system was evaluated periodically through an external audit of an 'Infection Committee' to ensure data quality. The gold standard was the information provided by cross-sectional studies and according to this criterion the quality of the information collected using the system was classified as excellent (sensitivity 80% and positive predictive value 95%).[32,33]

In this study the nosocomial infection problem was viewed as an inverse production problem or an adverse result of the hospitalization. In the classic industrial model, the absence of strikes, the large number of working days, personnel training etc. indicate an increase in productivity, similarly, in a hospital, these same factors can be associated with a reduction in the incidence of nosocomial infection. The analysis performed supports some aspects of this parallelism.

Donabedian[34] has argued that the motivation of a medical professional is directly proportional to the quality of the care provided. Whereas, Vuori indicated that lack of quality was due partly to reduced capacity, bad attitude and lack of continuous training of medical staff.[35] Hospital quality control generally evaluates the assistance provided by professionals. It is in this area where greatest variations take place and where the different correcting actions are most effective.

Both the Joint Comission and the SENIC study give considerable weight to continuous training programmes in the evaluation of surveillance and infection control systems.[36,37]

Indicators are the basis of the Joint Commission's performance monitoring system. An indicator is not a direct measure of quality but rather a quantitative measure of an aspect of patient care that can be used as a guide to monitor and evaluate the quality of health care.[22]

A drastic decrease in the initial levels of nosocomial infection in response to the continuous surveillance and personnel training programme was initially observed. Following this, a stable level of occurrence was attained. This may be considered as the irreducible minimum associated with the hospital and surveillance system itself. This minimum level of infection may provide a hospital standard for an *a posteriori* quality control system. Any further reduction of this level may be expensive and involve additional efforts and changes in the surveillance system.

There was a marked increase in nosocomial infection corresponding to a prolonged medical strike. The ultimate causes of this observed increase could form the object of further study. A preliminary study designed to quantify the effects of the medical strike on care indicators has been recently conducted. A case-control design of two different periods (the strike period and a reference period) detected errors by hospital personnel such as an increase in duration of urinary catheterization, erroneous antimicrobial dosage and length of antimicrobial course and lack of antibiotic prophylaxis prior to non-programmed surgical interventions.[33] The high level of infection recorded suggests that all the levels of a health care institution should be aware of the possible consequences in the case of a similar event.

The estimation of the missing datum by time series techniques appears to be the only possible way taking into account the temporal dependence of the data.

The different composition in week days of each month showed no effect on infection production. In particular, it may be concluded that weekends have no effect on the incidence of infection.

The opposite effects of two holidays periods, Easter and the September local holiday, could be explained by their very different nature. While in Easter the trend is to leave the city, in September local festivities affect all citizens for one week.

One of the most interesting conclusions of this study is the association observed between infection incidence and hiring of non-trained personnel. Detailed research is needed to establish the causes of this association and to impose restrictions on the activities of this personnel group.

The finding of an outlier was unexpected. Historical investigation of the circumstances present during the relevant month could give insight into new characteristics of nosocomial infection.

Finally, it may be said that, although there was no deterministic variation, stochastic seasonality suggested moderate variations due to different seasons of the year.

The above results reinforce Vuori and Donabedian's theoretical hypotheses from a practical point of view and provide several objectives for further individual case studies.[35,36]

Finally, the direct association between the accumulated incidence of nosocomial infection and aspects positively related with the quality of health care (urinary tract infection prevention programme and continuous personnel training) or negatively related (strike of medical staff and unmotivated personnel)[34,35] appear to confirm the utility of this indicator as a quantitative measure which can be used to monitor and assess the health care services given to patients. Temporal series analysis shows a potential in the examination of sources of variation in variables of interest, based on sound scientific data.

Acknowledgement

This project was partially supported by Plan Regional de Investigación de la Comunidad Autónoma de Madrid (grant 00036/92).

References

[1] Garner JS, Jarvis WR, Emori TG *et al.* CDC definitions for nosocomial infections. *Am J Infect Control* 1988;**16**:128–40.

[2] Langmuir AD. The surveillance of communicable diseases of national importance. *N Engl J Med* 1963;**253**:182–92.