ENVIRONMETRICS
Environmetrics, **10**, 439–455 (1999)

SEQUENTIAL MONITORING OF LOW EVENT RATES: AN APPLICATION IN ENVIRONMENTAL EPIDEMIOLOGY

ARMAND MAUL[1]* AND JEAN-FRANCOIS VIEL[2]

[1]*Department of Statistics, University of Metz, 57045 Metz, France*
[2]*Department of Public Health, Faculty of Medicine, 25030 Besançon, France*

SUMMARY

A sequential testing procedure for monitoring epidemiological data is considered. The approach is based on a discrete time process of interim analyses using the likelihood ratio as a test statistic. Sampling continues until either some predetermined practical time limit is reached or a decision can be made about the hazard rates characterizing the populations to be compared. The stopping boundaries of the sequential testing procedure are determined so as to control type I error at a given level during the whole study. Thus, the critical threshold is derived from the exact distribution of the maximum value of the test statistic which is observed across all the monitoring times. To this end, the appropriate quantiles in the probability distributions of interest are obtained empirically by Monte Carlo simulation, assuming both the null hypothesis and given experimental conditions. The method is illustrated by an example concerning the incidence of leukaemia in young people living in the vicinity of the French La Hague nuclear reprocessing plant. On the whole the previous analyses do not provide evidence for a spatial–temporal clustering of leukaemia, at least within the space–time window which was examined around the nuclear reprocessing plant (35 km radius, time period 1978–1992). Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS risk assessment; survival analysis; sequential tests; censored data; likelihood ratio; Monte Carlo simulation; leukaemia incidence

1. INTRODUCTION

The assessment of the impact of sources of pollution on the health status of communities is of considerable importance. In recent years, there has been growing interest in the development of statistical methods which are appropriate for the assessment of possible environmental pollution sources. A large body of literature has therefore developed concerning the analysis of disease incidence to detect the presence of clusters of disease around putative sources of hazard (Hills and Alexander 1989; Lawson and Waller 1996; Tango 1984; Viel *et al*. 1995; Waller *et al*. 1994; Whittemore *et al*. 1987).

Regarding the use of hypothesis tests for the assessment of putative sources (Bithell 1990; Cook-Mozzafari *et al*. 1989; Hills and Alexander 1989; Lawson 1993), the original test specifically designed for count data was proposed by Stone (1988). Nevertheless, in addition to its limitations, which have been discussed by Lawson and Waller (1996), it must be emphasized that Stone's test allows for a point-of-time assessment of the hazard only. In other words, this means that, unlike other authors (Chen *et al*. 1984; Ederer *et al*. 1966; Knox 1964; Lawson and

* Correspondence to: A. Maul, Department of Statistics, University of Metz, F-57045 Metz Cedex 01, France.

CCC 1180–4009/99/040439–17$17·50

Received 15 April 1998
Accepted 26 January 1999

Viel 1995; Mantel 1967), who focused the interest on space–time clustering, Stone's test is not appropriate to the detection of space–time interactions.

One enhancement to all the procedures designed to account for the space–time cluster structure would be to consider time intervals and to include jointly the recorded times of the cases, as suggested by Lawson and Viel (1995). This could lead both to greater sensitivity and the possibility of a stepwise assessment of the putative source while monitoring the data by following a prospective design. Furthermore, the use of a prospective design may avoid going through *a posteriori* inference problems relevant to prior knowledge of an apparent effect or reported disease incidence near a putative source. In particular, both hypothesis tests and study region definition can be biased by making *a posteriori* inference (Gardner 1989; Hills and Alexander 1989; Lawson and Waller 1996).

However, as a general statement applicable to prospective designs, inferences about the parameters of interest or the power of the tests of comparison obviously tend to improve as the available information which is accumulated over time increases. On the other hand, considering the possible implications for public health intervention, prevention policy or ethical aspects, it is desirable that a change from normal patterns be detected as soon as possible. In this regard, sequential monitoring offers a compromise between the two previous antagonistic time constraints in the sense that sequential analysis of the data is to be continued until either a decision can be made about the null hypothesis or some predetermined practical time limit for the study is reached. Thus, reviewing the situation in a sequential way has become a topic of major interest in risk assessment methodology since interim analyses enable investigators to make more efficient use of limited research resources.

Sequential analysis of survival data has therefore received considerable attention in the literature (Wetherill 1986) and, more particularly, in a medical context (Lan and DeMets 1983; Whitehead 1992). Several methods to derive exact stopping boundaries for group sequential clinical trials were developed recently (Mehta *et al.* 1994; Lin *et al.* 1991; Pawitan and Hallstrom 1990). These methods are based on the exact joint permutation distribution of rank statistics observed across all the monitoring times. Lan and Zucker (1993) present a unified conceptual framework for sequential monitoring covering a wide variety of clinical settings. Reviews of some currently used sequential methods are given by Fleming and DeMets (1993) and Lee (1994).

Nevertheless, the credibility of environmental health investigations can be compromised if inappropriate procedures are used. In particular, interim monitoring of the data may generate a high number of censored values and/or tied failure times which is known to be a source of computational difficulty in the field of survival data analysis (Kalbfleisch and Prentice 1980). But the main problem when testing the null hypothesis in a sequential way is to adjust the critical threshold so as to maintain type I error rate under a prespecified level of significance during the entire study. Another concern arises when the phenomenon under study is relatively rare, since this may lead to less reliable model fitting and also invalid asymptotic distribution of standard test statistics.

The purpose of this paper is to outline a sequential process to compare disease incidence rates in several groups of individuals which is appropriate to deal with sparse count data and/or the previous undesirable consequences of interim monitoring. The method is based on a discrete time expression of the hazard while examining the study subjects individually. The critical threshold and the stopping rule are derived directly from the probability distribution of the likelihood ratio, which is used as a test statistic, assuming the null hypothesis is true. This appears to be a preferable approach rather than using the unnecessarily conservative Bonferonni's inequality

(Miller 1981) to adjust the critical region for the multiple comparison problem since, in the present context, the consecutive comparisons are highly dependent.

Thus, our prior objective will consist of assessing the effects of: the sample size, the level of type I error, the predetermined practical time limit, and the hazard rates on the critical threshold of the test. The appropriate threshold, which is determined empirically by Monte Carlo simulations, will help us to state the decision rule in future comparative studies so as to maintain type I error rate to some prespecified level of significance during the whole experiment.

The procedures of this work are illustrated by a numerical example involving childhood leukaemia incidence around the La Hague nuclear waste reprocessing plant. Interim monitoring of the data is performed to: (i) compare the incidence rates observed with other reference incidence rates in France; and (ii) detect a possible spatial pattern regarding the risk of leukaemia in the vicinity of the putative source. Such analyses may be useful for determining problematic areas deserving further investigations and/or remedial actions which then could be undertaken sooner.

2. STATISTICAL METHODS

2.1. Model

Although the scope and the numerical example of this paper will be focused on the comparison of two groups of individuals only, the results presented in this work can be easily generalized to any number of groups by following the approach which is outlined hereafter.

Thus, let us consider g groups of individuals and let $\pi_{ij}(t)$ be the hazard rate corresponding to the jth individual ($j = 1, \ldots, n_i$) of the ith group ($i = 1, \ldots, g$) at time t. The hazard is given as a discrete expression of the time, i.e. the hazard function is supported on the integers ($t = 1, 2, \ldots$). Clearly, $\pi_{ij}(t)$ is the probability for individual j in group i to fail (e.g. onset of clinical diagnosed leukaemia) at time t provided that the individual was still at risk at time $t-1$. The random variable associated with the duration corresponding to this individual is denoted Y_{ij}. If Y_{ij} is censored on one side (e.g. the right) the observed survival time will be denoted y_{ij}^c. This means that the survival (i.e. event-free) time of the individual was at least y_{ij}^c. Both right or left single censoring may be considered hereafter.

The further development emphasizes the special case which arises when the hazard rate characterizing the individuals is time-fixed. In other words, we assume that the survival times are exponentially distributed, i.e. the hazard is a constant for all the individuals belonging to the same group, that is $\pi_{ij}(t) = \pi_i$ ($t = 1, 2, \ldots$; $j = 1, \ldots, n_i$).

We have

$$\begin{cases} \Pr(Y_{ij} = y_{ij}) = (1-\pi_i)^{y_{ij}-1}\pi_i \\ \Pr(Y_{ij} = y_{ij}^c) = (1-\pi_i)^{y_{ij}^c} \end{cases} \qquad (i = 1, \ldots, g;\ j = 1, \ldots, n_i;\ y_{ij} \text{ or } y_{ij}^c = 1, 2, \ldots) \quad (1)$$

2.2. Test statistic

The observed survival times which are available for group i ($i = 1, \ldots, g$) at stage t ($t = 1, 2, \ldots$) in the sequential procedure are:

$$\underline{y}_i^t = (y_{i1}, \ldots, y_{i,n_i-k_i^t}, y_{i,n_i-k_i^t+1}^t, \ldots, y_{i,n_i}^t)$$

where the last k_i^t individuals are still at risk at time t. Note that t is then taken as the censoring value.

The likelihood function of the sample $\underline{y}^t = (\underline{y}_1^t, \ldots, \underline{y}_g^t)$ at stage t is given by

$$L(\underline{y}^t|\underline{\pi}) = \prod_{i=1}^{g} \left\{ \prod_{j=1}^{n_i-k_i^t} [(1-\pi_i)^{y_{ij}-1}\pi_i] \prod_{j=n_i-k_i^t+1}^{n_i} (1-\pi_i)^{y_{ij}^t} \right\} \tag{2}$$

The test statistic calculated at stage t of the sequential process is expressed in terms of the ln likelihood ratio. In particular, when comparing the incidence rates of g populations with a reference value, i.e. H_0: $\pi_1 = \cdots = \pi_g = \pi_0$, the test statistic can be written as

$$-2 \ln \Lambda^t(\underline{y}^t|\pi_0) = 2\sum_{i=1}^{g} \left\{ \ln\left(\frac{1-\hat{\pi}_i^t}{1-\pi_0}\right) \sum_{j=1}^{n_i} y_{ij}^* + d_i^t \ln\left(\frac{\hat{\pi}_i^t(1-\pi_0)}{\pi_0(1-\hat{\pi}_i^t)}\right) \right\} \tag{3a}$$

where the number of events observed in group i at stage t of the process, $n_i - k_i^t$, is denoted d_i^t, whereas y_{ij}^* is for y_{ij}^t or y_{ij} according to whether Y_{ij} has been censored at y_{ij}^t or not, respectively. Equation (3a) is based on substitution of the maximum-likelihood estimate of π_i which is calculated at stage t, namely:

$$\hat{\pi}_i^t = d_i^t / \sum_{j=1}^{n_i} y_{ij}^*.$$

It is interesting to point out that if the estimates of the π_is are small then equation (3a) simplifies to one of the two following equivalent forms

$$-2 \ln \Lambda^t(\underline{y}^t|\pi_0) = 2\sum_{i=1}^{g} \left\{ d_i^t \left(-1 + \frac{\pi_0}{\hat{\pi}_i^t} + \ln\frac{\hat{\pi}_i^t}{\pi_0} \right) \right\} \tag{3b}$$

or

$$-2 \ln \Lambda^t(\underline{y}^t|\pi_0) = 2\sum_{i=1}^{g} \left\{ \pi_0 \sum_{j=1}^{n_i} y_{ij}^* - d_i^t - d_i^t \ln\frac{\pi_0 \sum_{j=1}^{n_i} y_{ij}^*}{d_i^t} \right\} \tag{3c}$$

Note that if $d_i^t = 0$ (i.e. $\hat{\pi}_i^t = 0$) then the contribution of group i to $-2 \ln \Lambda^t(\underline{y}^t|\pi_0)$ is

$$2\pi_0 \sum_{j=1}^{n_i} y_{ij}^*.$$

Similarly, if we want to test the equality of the different hazard rates in the g groups examined, the corresponding equations are obtained straightforwardly by just substituting $\hat{\pi}^t$, that is the MLE of the common hazard rate, for π_0 in equations (3a–c).

2.3. Sequential stopping rule and decision criteria

We address situations in which the data are monitored at interim time points that are regularly spaced. At each look, t is increased by one up to some predetermined practical time limit $t_{\max}$. The choice of $t_{\max}$, which must be made *a priori*, requires a specific answer. For example, a unique exposure episode, potentially inducing a pathology with a short latency period, usually warrants a short length of surveillance. The opposite applies for repeated exposures and cancer. Besides, political or financial limitations can preclude the setting of a permanent surveillance system (e.g. cancer registry), and allow only a temporary survey. The stopping boundaries, when performing the sequential procedure, are derived from the exact distribution of the maximum value of the ln likelihood ratio statistic, $-2 \ln \Lambda_{\max}$, which is observed across all the monitoring times, given n_i $(i = 1, \ldots, g)$, α, $t_{\max}$ and the level of the common hazard rate as stated in H_0.

Hence.

$$-2 \ln \Lambda_{\max} = \max\{-2 \ln \Lambda^t | t \in [1, \ldots, t_{\max}]\}$$

The critical threshold will be denoted by $s(n_1, \ldots, n_g, \alpha, t_{\max}, H_0)$ since it depends on the sample size, the level of type I error, the practical time limit and the null hypothesis. Its value is determined so as to fulfil the following constraint

$$\alpha = \Pr(-2 \ln \Lambda_{\max} > s(n_1, \ldots, n_g, \alpha, t_{\max}, H_0) | H_0) \tag{4}$$

Therefore, α may be considered the probability of observing falsely a significant outcome before the practical time limit has been reached and provided H_0 is true. Practically, the value of the critical threshold as obtained from equation (4) allows to control type I error rate under α.

The decision rule at stage t $(t = 1, 2, \ldots, t_{\max})$ is as follows:

- if $-2 \ln \Lambda^t \leqslant s(n_1, \ldots, n_g, \alpha, t_{\max}, H_0)$ and $t < t_{\max}$, then continue;
- if $-2 \ln \Lambda^t > s(n_1, \ldots, n_g, \alpha, t_{\max}, H_0)$ and $t \leqslant t_{\max}$, stop and reject H_0, i.e. declare the hazard rates are different either from each other or the reference value, according to the statement made in the null hypothesis;
- if $-2 \ln \Lambda^t \leqslant s(n_1, \ldots, n_g, \alpha, t_{\max}, H_0)$ and $t = t_{\max}$, then the sequential test is declared inconclusive.

Note that the probability of observing a significant result under H_0 at any stage t, such as $t \leqslant t_{\max}$, is in fact lower than α. The value of α therefore provides an over-estimate of the actual type I error rate in case a decision is made and the sequential process is stopped before $t_{\max}$. In practical terms, this is an error in the right direction since our approach leads to more conservative tests.

2.3.1. Comparison of two populations

A special case arises when comparing two populations only. In this case sequential sampling allows a decision to be made about the hazard rates on the basis of the cumulative numbers of events which are observed within the groups to be compared at each interim analysis of the monitoring process. This approach may be useful, for example, to demonstrate the apparent existence of a distinct space–time clustering in the disease incidence.

In the following, we assume again that the number of 'person-years' corresponding to group i ($i = 1, 2$) is equal to $n_i t$

$$\left(\text{i.e. } \sum_{j=1}^{n_i} y_{ij}^* \cong n_i t\right).$$

Then, if the number of events observed in group i at any stage of the process is d_i, the test statistic can be formed as

$$-2 \ln \Lambda \cong 2\left[d_1 \ln \frac{d_1}{n_1} + d_2 \ln \frac{d_2}{n_2} - (d_1 + d_2)\ln\left(\frac{d_1 + d_2}{n_1 + n_2}\right)\right] \tag{5a}$$

or

$$-2 \ln \Lambda \cong 2 \ln\left[d_1^{d_1} d_2^{d_2} r^{d_2}\left(\frac{1 + 1/r}{d_1 + d_2}\right)^{d_1 + d_2}\right] \tag{5b}$$

Note that $-2 \ln \Lambda$ depends on d_1, d_2 and the sample ratio, $r = n_1/n_2$ only, as it appears in equation (5b). Define

$$A = \{(d_1, d_2) \in N \times N | -2 \ln \Lambda \leqslant s(n_1, n_2, \alpha, t_{\max}, H_0)\} \tag{6}$$

where the critical threshold $s(n_1, n_2, \alpha, t_{\max}, H_0)$ is determined so as to satisfy the following equation

$$1 - \alpha = \Pr(-2 \ln \Lambda_{\max} \leqslant s(n_1, n_2, \alpha, t_{\max}, H_0) | H_0)$$

(see equation (4)).

This allows to control type I error at level α during the whole procedure as mentioned before. Accordingly, the joint probability that (D_1, D_2) belong to A or $\bar{A}$, given H_0 is true, is given as

$$\Pr((D_1, D_2) \in A | H_0) = 1 - \alpha \tag{7a}$$

and

$$\Pr((D_1, D_2) \in \bar{A} | H_0) = \alpha \tag{7b}$$

respectively. Hence, the decision criteria in terms of (d_1, d_2) can be stated as follows: if the combination (d_1, d_2), which is observed at any step of the sequential process, belongs to $\bar{A}$ then H_0 is rejected, otherwise it is not.

2.3.2. Choosing between two opposed hypotheses

Another worthwhile problem arises if we are interested in choosing between two opposing hypotheses about the hazard rate characterizing a single population, namely:

$$H_1 : \pi \leqslant \pi_1$$

and

$$H_2 : \pi \geqslant \pi_2$$

(it is assumed that $\pi_1 < \pi_2$).

When choosing between the previous opposing hypotheses about the hazard rate, it is convenient to make a decision directly on the basis of the cumulative numbers of events which are observed at each interim analysis of the sequential process. The decision regarding compliance with one of these two hypotheses implies a certain risk of reaching an erroneous conclusion. The process has to be continued until the path of the total number of events (d^t) observed at stage t falls outside an uncertainty region. This region is delimited by the curves obtained from formula (8) where the following notation has been used for convenience: $s(n, \alpha_i, t_{\max}, H_i) = s_i$ $(i = 1,2)$ (see Appendix).

$$d^t < \frac{(\pi_2 - \pi_1)\sum_{j=1}^{n} y_j^*}{\ln(\pi_2/\pi_1)} - \frac{s_2/2}{\ln(\pi_2/\pi_1)} \tag{8a}$$

$$d^t > \frac{(\pi_2 - \pi_1)\sum_{j=1}^{n} y_j^*}{\ln(\pi_2/\pi_1)} + \frac{s_1/2}{\ln(\pi_2/\pi_1)} \tag{8b}$$

Clearly, in the situation corresponding to equation (8a), we conclude that the hazard rate is at most as large as π_1, and the risk of accepting H_1, when actually H_2 is true, is less than α_2. Similarly, if equation (8b) is satisfied, we conclude that the actual hazard rate is at least as large as π_2; however, the risk of accepting H_2, when in fact H_1 is true, is less than α_1. Finally, the situation of a cumulative number of events remaining between the previous curves during the whole experiment leads to an inconclusive test, i.e. the true hazard rate probably lies somewhere between π_1 and π_2. Furthermore, if the number of 'person-years' is equal to the product nt, which is valid in the case of a stable population and/or, as an approximation, in the case of low event rates, then equations (8a, b) can be written as

$$d^t < \frac{(\lambda_2 - \lambda_1)t}{\ln(\lambda_2/\lambda_1)} - \frac{s_2/2}{\ln(\lambda_2/\lambda_1)} \tag{9a}$$

$$d^t > \frac{(\lambda_2 - \lambda_1)t}{\ln(\lambda_2/\lambda_1)} + \frac{s_1/2}{\ln(\lambda_2/\lambda_1)} \tag{9b}$$

In fact, using equations (9a, b) instead of equations (8a, b) amounts to considering a Poisson model with a constant parameter (i.e. $\lambda_i = \pi_i n$; $i = 1, 2$) for describing the disease incidence in the population examined.

2.4. Determining the critical threshold by Monte Carlo simulation

The critical threshold $s(n_1, \ldots, n_g, \alpha, t_{\max}, H_0)$ corresponding to any test is determined by Monte Carlo simulation as a function of n_i $(i = 1, \ldots, g)$, α, $t_{\max}$ and the hazard rate as stated in H_0.

Each point estimation in the graphs of this paper results from 10,000 distinct simulation runs. Moreover, the survival of the individuals at risk within each simulated test is examined at each step of the sequential procedure by following a Bernoulli random process. Thus, the parameter of the Bernoulli distribution for group i is equal to the hazard or incidence rate characterizing this group.

Furthermore, setting the critical threshold at a level which is appropriate to ensure a type I error rate equal to α, amounts to determining the quantile at level $1-\alpha$ in the distribution of $-2\ln\Lambda_{\max}$. This can be done by simulation for any combination of the parameters governing the experimental design, assuming the null hypothesis is true. Nevertheless, the scope of this paper is focused on the comparison of two populations only. Figure 1 gives, for example, the mean and the quantiles at level 0·90, 0·95 and 0·99 in the distribution of $-2\ln\Lambda_{\max}$ as a function of the common hazard rate as stated in H_0, assuming $t_{\max}$ has been set at two different levels and there is a total of 100 individuals equally split into the two groups to be compared. It is interesting to note that any given quantile may be considered a constant as π becomes large enough. Moreover, it can be shown from other simulations assuming varying values of the total sample size, n, that the distribution of $-2\ln\Lambda_{\max}$ depends mainly on $t_{\max}$ and H_0. This means that under H_0 and for a given value of $t_{\max}$, the distribution of $-2\ln\Lambda_{\max}$ and consequently the quantiles may be considered approximately independent of n, provided $n\pi$ is large enough (say $n\pi \geqslant 1$). As to the incidence of the practical time limit of the testing procedure on the quantiles, the curves in Figure 1 indicate that the higher the value of $t_{\max}$, the higher the quantiles.

Figure 2 displays the variation of the critical threshold in a three-dimensional plot. In particular, the surface in Figure 2 shows the variation of the quantile at level 0·95 in the empirical distribution of $-2\ln\Lambda_{\max}$ under H_0 as a function of $t_{\max}$ and π. The visual information on the 3-D scatterplot shown in Figure 2 has been enhanced by using a multivariate smoothing procedure,

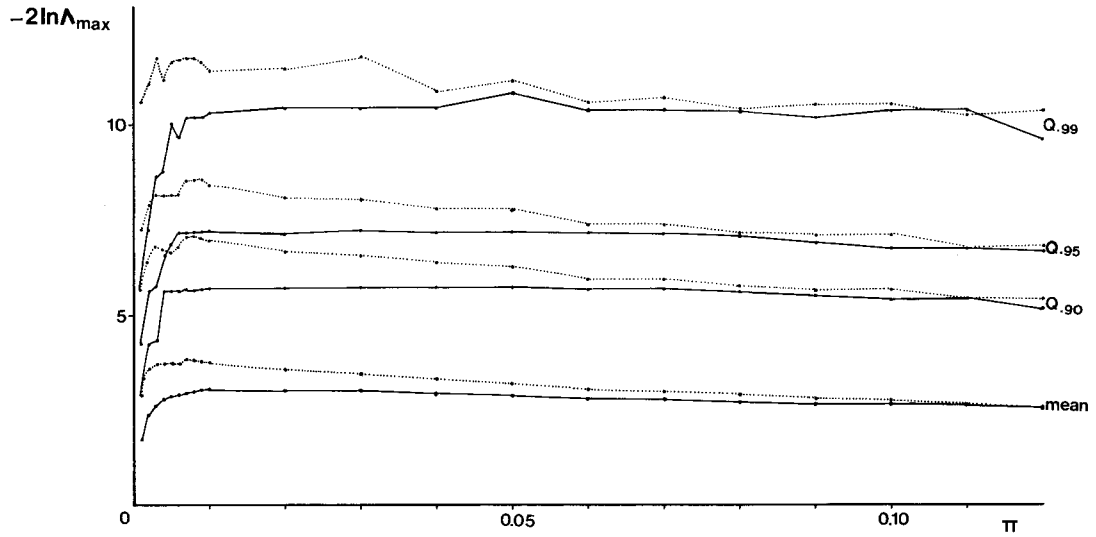

Figure 1. Quantiles in the distribution of $-2\ln\Lambda_{\max}$ under H_0 ($\pi_1 = \pi_2 = \pi$) as a function of the common hazard rate, assuming $n = 100$, $r = 10{,}000$ and two different levels for the practical time limit: $t_{\max} = 20$ (solid curve) and $t_{\max} = 200$ (dotted curve). Each point estimation in the different curves was obtained from $r = 10{,}000$ distinct Monte Carlo simulation runs

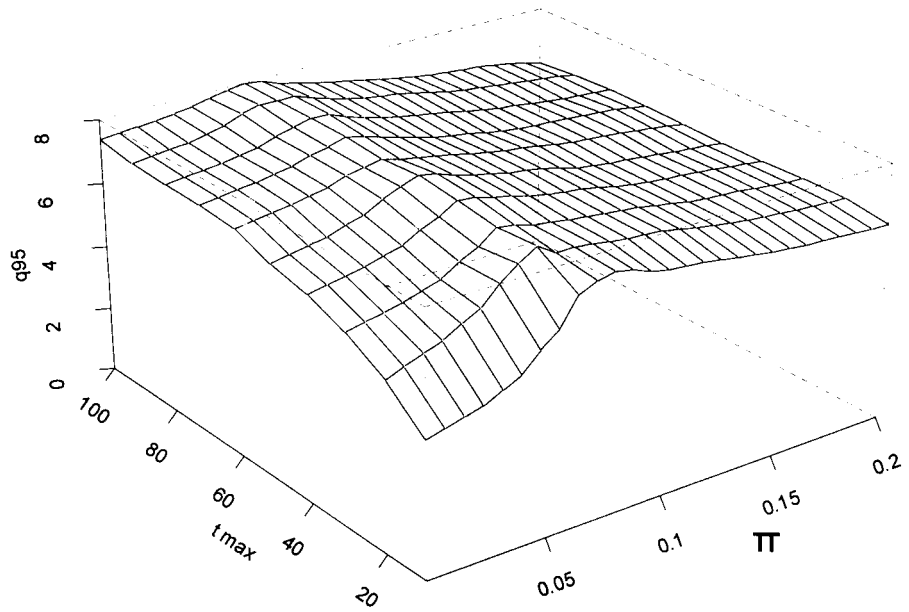

Figure 2. Estimated quantiles at level 0·95 in the distribution of $-2 \ln \Lambda_{\max}$ under H_0 ($\pi = \pi_1 = \pi_2$) as a function of the common hazard rate and $t_{\max}$; $n = 100$ and $r = 10{,}000$

namely locally weighted regression (Cleveland and Devlin 1988). The surface shown in Figure 2 is relatively flat provided that $t_{\max}$ and π are not too close to zero. This property concerning the general shape of the surface corresponding to any other quantile enables one to choose the value of the critical threshold in order to maintain type I error under any prespecified level.

Following another approach, namely when comparing the hazard rate with a reference value, the variation of the quantiles in the distribution of $-2 \ln \Lambda_{\max}$ under H_0 as a function of the reference value for the hazard rate is shown in Figure 3. From this figure, it appears that the different quantiles may be considered a linear function of π, at least if π is not too close to zero. Moreover, it is worthwhile to point out that, if the values of $t_{\max}$ and α are fixed and provided π remains small enough, the quantiles depend in fact on the product $n\pi$, whatever the values of n and π as taken individually.

In order to sum up the results shown in the previous figures, and regarding practical purposes, it must be emphasized that, if the product $n\pi$ is large enough (say $n\pi$ is larger than 1 or 2) then (i) the critical threshold, $s(\alpha, t_{\max})$, depends on α and $t_{\max}$ only when performing a comparison test of the hazard rates in several populations, whereas (ii) its value, $s(\alpha, t_{\max}, \pi)$, is approximately an increasing linear function of π when a hazard rate is to be compared with a reference value.

Thus, in consideration of all these remarks it is always possible to set the value of the critical threshold so as to control type I error under a given level during the whole study. This is actually one of the major points of the procedure presented in the present work.

3. EXAMPLE

The different aspects of the method are illustrated by an example which is concerned with the incidence of childhood leukaemia around the La Hague nuclear waste reprocessing plant, for the

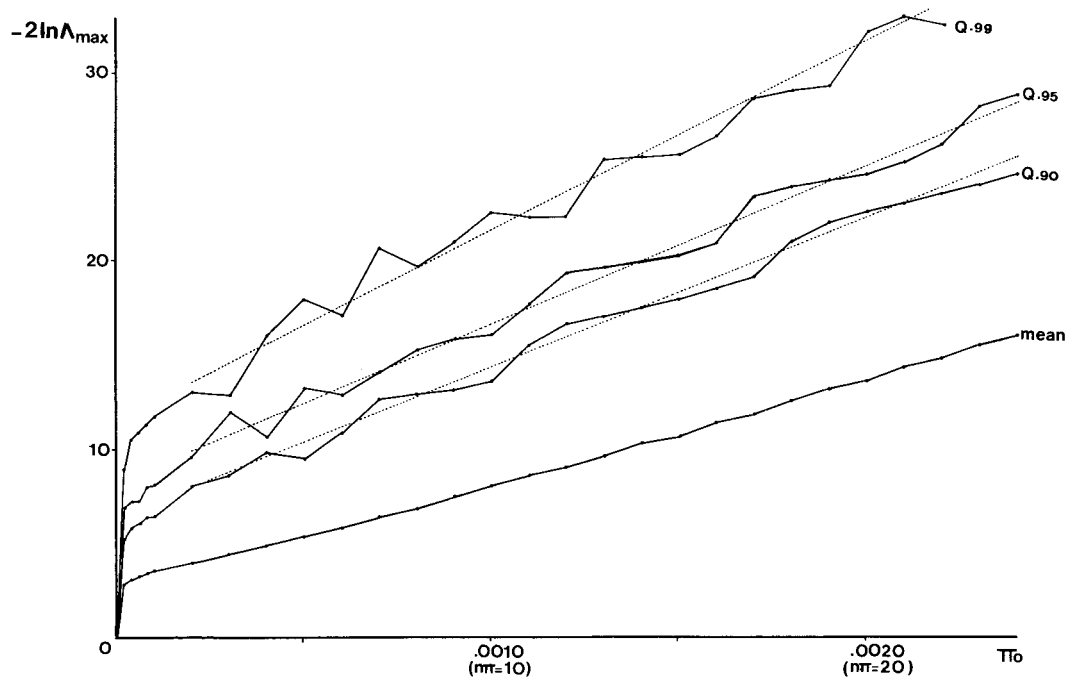

Figure 3. Quantiles in the distribution of $-2 \ln \Lambda_{\max}$ under H_0 ($\pi = \pi_0$) as a function of the reference value for the hazard rate with fitted least squares regression line (dotted curves); $n = 10{,}000$, $t_{\max} = 20$ and $r = 10{,}000$

period from 1 January 1978 to 31 December 1992 among people aged under 25 years (Lawson and Viel 1995; Viel and Richardson 1990; Viel *et al.* 1993, 1995). La Hague is located in Normandy on the Channel coast in France (see Figure 4). The study area is composed of 10 electoral wards. These wards have been divided into three distinct subregions according to their distance from the plant, namely 10 km, 20 km and 35 km around the plant, as can be seen in Figure 4. The data are presented in Table I which shows the person-years, the observed numbers of leukaemia cases and the estimated incidence rates corresponding to each subregion. For instance, the total person-years were estimated at 891,688, leading to a crude incidence rate equal to 2·80 per 100,000, whereas the incidence rate was estimated at 7·79 in the area closest to the plant.

The 25 cases of leukaemia diagnosed during the study period are examined prospectively in a sequential way within a 20 years' time window. The recording of spatial incidence of leukaemia cases with time of occurrence as an ordering label will be used to investigate some aspects of

Table I. Observed numbers of leukaemia cases and incidence rates around La Hague nuclear reprocessing plant according to distance from plant (adapted from Viel *et al.* 1995)

Distance from plant (km)	Person-years	Observed cases	Estimated incidence rate per 100,000
<10	51,332	4	7·79
10–20	493,390	11	2·23
20–35	346,971	10	2·88
Total	891,688	25	2·80

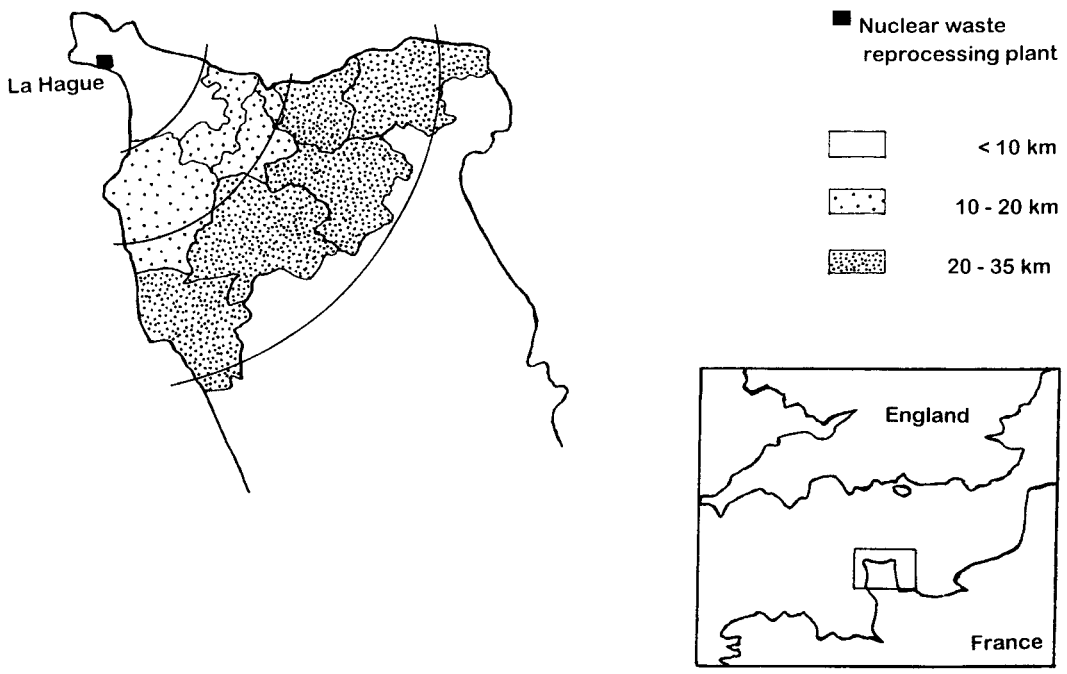

Figure 4. Subregions included in the study area according to their distance from the plant

childhood leukaemia incidence. The space–time interaction will be analysed at regularly spaced times, i.e. every year, in order to (i) perform comparison tests with reference rates or regarding the choice between two opposing hypotheses, and (ii) identify the possible existence of a distinct cluster of childhood leukaemia.

As a first test, we want to compare the incidence rate which is observed in the closest subregion to the plant with a reference value, i.e. 2·89 per 100,000. This figure may be considered a rough estimate of the mean incidence rate of childhood leukaemia in France (Viel *et al.* 1995). Figure 5 shows the path of the ln likelihood ratio calculated at each interim analysis. The value of $-2 \ln \Lambda$ observed by the end of each of the 15 years of the study did not fall beyond the horizontal lines indicating the critical threshold at the 5% and 1% error level. The values corresponding to these error levels: 4·49 and 7·73, respectively, were obtained by Monte Carlo simulation, assuming $n = 3422$ (average number of exposed people, i.e. 51,332/15), $t_{\max} = 20$ and $\pi_0 = 2{\cdot}89 \times 10^{-5}$. Notwithstanding that the study has in fact been planned to be continued for 20 years, at least in our example since the critical threshold was calculated accordingly, we may state that there is no real evidence for the incidence rate in subregion 1 to exceed the reference value after a 15-year period of observation.

The sequential method for choosing between two opposing hypotheses about the incidence rate is illustrated in Figure 6 using the whole data set, that is the data available for the three subregions around the plant during the 15-year period of time considered. Suppose we want to decide on the following two opposing hypotheses

$$H_1 : \pi \leqslant 2{\cdot}89 \times 10^{-5}$$

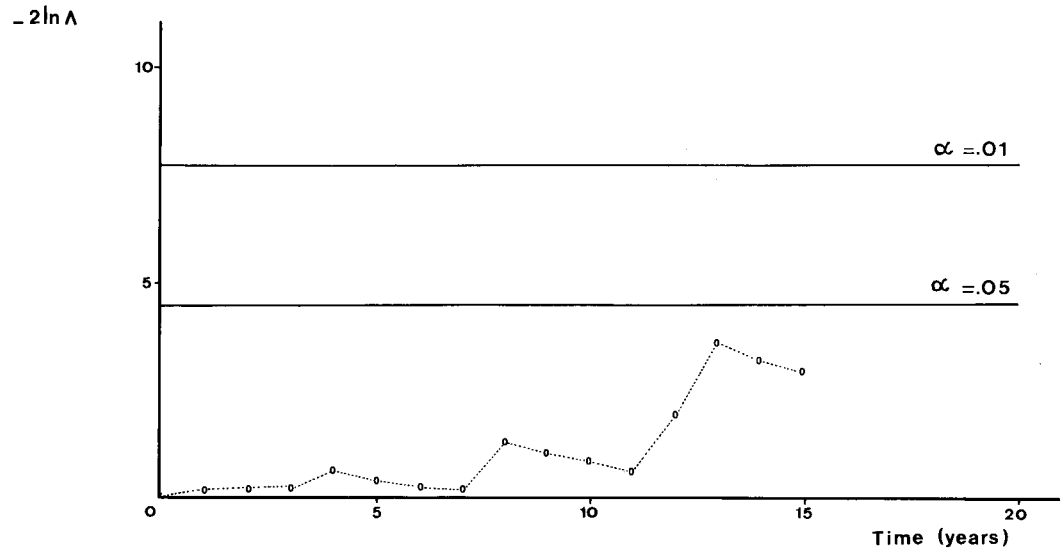

Figure 5. Sequential sampling for comparing the incidence rate with a reference value using the La Hague data for subregion 1 (<10 km), assuming $n = 3422$, $t_{max} = 20$ and the reference value $\pi_0 = 2{\cdot}89 \times 10^{-5}$

and

$$H_2 : \pi \geqslant 5{\cdot}78 \times 10^{-5}$$

Each of these hypotheses is subject to two types of sampling error, namely: α_1 (α_2) is the probability of accepting H_2 (H_1) when actually H_1 (H_2) is true. When both α_1 and α_2 are fixed, it is possible to draw two parallel lines obtained from formula (8a, b). These lines are given in Figure 6 for different levels of α_1 and α_2. The process of interim monitoring is continued until either the cumulative number of events falls outside the uncertainty region delimited by two associated curves or until the predesignated practical time limit has been reached. If the cumulative number of events remains in the uncertainty zone by the time the practical limit for the study has been reached then the true incidence rate is probably somewhere between π_1 and π_2. In the present case, H_1 could be accepted from the 14th or the 15th observation year onwards at the 5% or 1% error level, respectively.

Regarding the possible existence of a pattern of spatial heterogeneity in the incidence of childhood leukaemia, it may be useful to compare the different subregions two by two in terms of their corresponding numbers of leukaemia cases. The process outlined in Section 2.3.1 is illustrated in Figure 7 which gives, for example, the decision zones and the path of the numbers of leukaemia cases observed during the successive time intervals in subregions 2 and 3. The process can be performed in the same way by taking any two subregions among the three considered. Since none of the paths thus obtained fell outside the uncertainty zone at the 5% error level, there is no evidence for any structured heterogeneity pattern in childhood leukaemia incidence, at least within the space–time window examined.

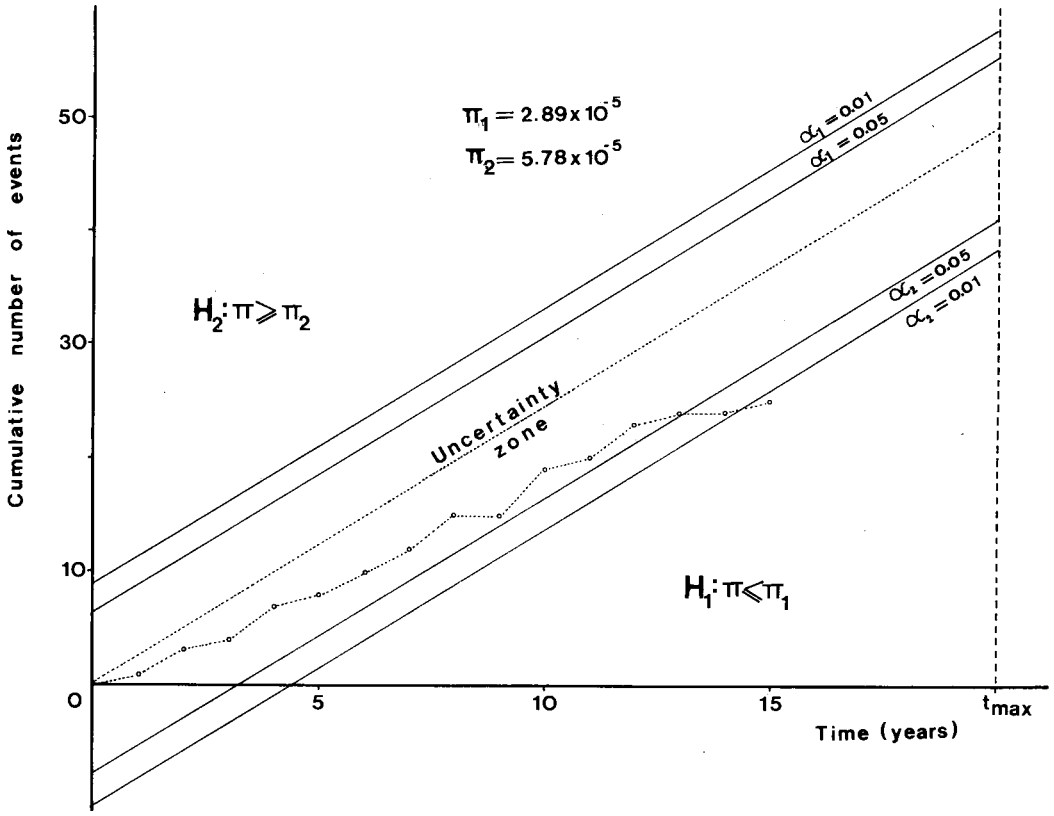

Figure 6. Sequential sampling for deciding between two hypotheses (i.e. $H_1: \pi \leqslant \pi_1$ vs $H_2: \pi \geqslant \pi_2$ with $\pi_1 = 2{\cdot}89 \times 10^{-5}$ and $\pi_2 = 5{\cdot}78 \times 10^{-5}$) using the La Hague data

4. CONCLUDING REMARKS

The sequential procedure suggested in this paper provides a particularly convenient way for comparing disease incidence rates in a wide class of biomedical and environmental health investigations. The method allows for investigating the incidence of space–time disease clustering around a putative environmental pollution source by following a stepwise assessment of the hazard.

It accommodates the possibility of handling data sets comprising sparse count data and/or high rates of censored failure times. The critical threshold of the test is determined so as to control Type I error at a prespecified level during the entire sequential testing procedure. This can be done empirically by Monte Carlo simulation as a function of the sample size, the level of type I error, the practical time limit and the hazard rates.

As to the numerical example dealing with leukaemia incidence, the statistical analyses used in this paper do not provide evidence for an excess of childhood leukaemia around La Hague nuclear reprocessing plant, at least within the space–time window (i.e. distance from plant lower than 35 km; period 1978–1992) of this study. This result seems to contradict the conclusion stated by Viel *et al.* (1995) who demonstrated the apparent existence of a distinct cluster of childhood

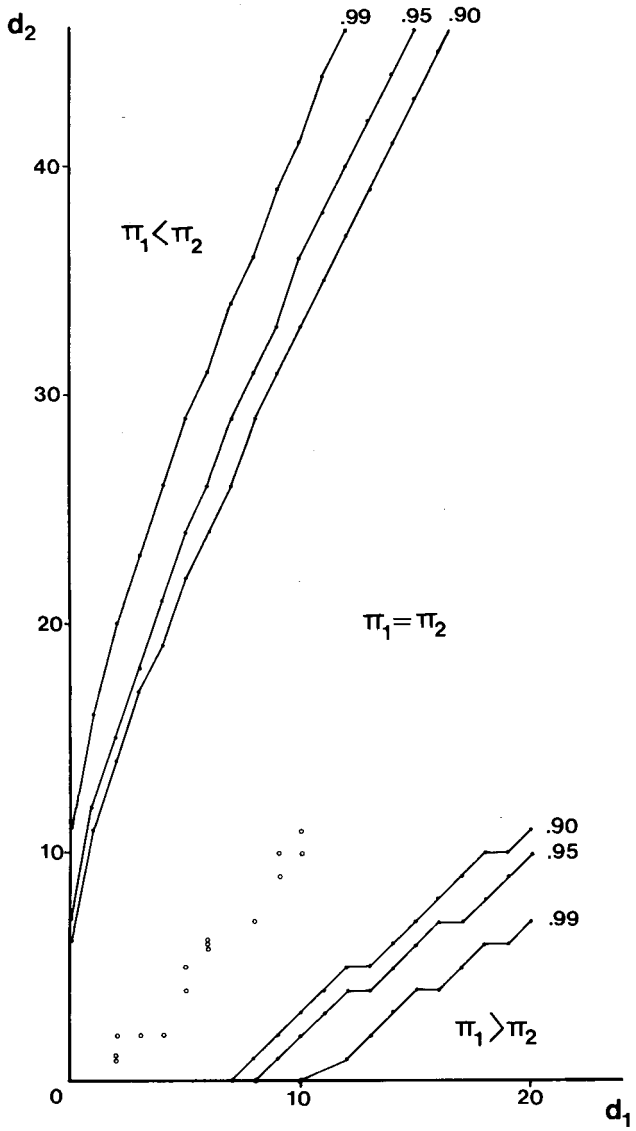

Figure 7. Decision zones and path of the observed numbers of leukaemia cases in subregions 2 (10–20 km) and 3 (20–35 km) around La Hague nuclear reprocessing plant (adapted from Viel *et al.* 1995). The average numbers of exposed people in the subregions examined were estimated by $n_1 = 23{,}131$ and $n_2 = 32{,}750$, respectively

leukaemia in the immediate vicinity of the plant. However, it should be pointed out that the statement of the latter study is the result of a unique point-of-time assessment of the hazard after a 15 year period of time and not of a prospective and continual observation of an incidence time series. As a matter of fact, if we compare the SIRs of the area 0–10 km and 10–20 km by testing the equality of two Poisson distributions, then we find a p-value of 0·05. Hence, the method suggested in this paper is likely to be somewhat more conservative in terms of testing hypotheses,

since it is carried out within the framework of a stepwise assessment which takes into account the space–time interaction. In this respect, the results of the sequential procedure are in line with the absence of directional space–time interaction reported by Lawson and Viel (1995) following another method.

APPENDIX

The following notations are used for convenience:

$$s(n, \alpha_i, t_{\max}, H_i) = s_i \ (i = 1, 2)$$

$$2\left\{\pi \sum_{j=1}^{n} y_j^* - d^t - d^t \ln \frac{\pi \sum_{j=1}^{n} y_j^*}{d^t}\right\} = -2 \ln \Lambda^t(\underline{y}^t|\pi)$$

(see equation (3c)).

Consider the following systems which are expressed in terms of the likelihood ratio statistic with respect to s_1 and s_2. In particular, the systems identified as (I) and (II) lead to accepting hypothesis H_1 ($\pi \leqslant \pi_1$) and H_2 ($\pi \geqslant \pi_2$), respectively.

$$\text{(I)} \begin{cases} -2 \ln \Lambda^t(\underline{y}^t|\pi_2) > s_2 \\ -2 \ln \Lambda^t(\underline{y}^t|\pi_1) \leqslant s_1 \end{cases} \qquad \text{(II)} \begin{cases} -2 \ln \Lambda^t(\underline{y}^t|\pi_1) > s_1 \\ -2 \ln \Lambda^t(\underline{y}^t|\pi_2) \leqslant s_2 \end{cases}$$

In order to obtain explicit expressions to be used for making a decision, the previous systems are replaced by the following more stringent conditions

$$\text{(I)} \begin{cases} -2 \ln \Lambda^t(\underline{y}^t|\pi_2) > s_2 \\ -2 \ln \Lambda^t(\underline{y}^t|\pi_1) = 0 \end{cases} \qquad \text{(II)} \begin{cases} -2 \ln \Lambda^t(\underline{y}^t|\pi_1) > s_1 \\ -2 \ln \Lambda^t(\underline{y}^t|\pi_2) = 0 \end{cases}$$

Then, it is easy to show that the latter systems lead to the following equations, regarding the cumulative number of events observed at stage t.

$$\text{(I) } d^t < \frac{(\pi_2 - \pi_1) \sum_{j=1}^{n} y_j^*}{\ln(\pi_2/\pi_1)} - \frac{s_2/2}{\ln(\pi_2/\pi_1)} \text{ and (II) } d^t > \frac{(\pi_2 - \pi_1) \sum_{j=1}^{n} y_j^*}{\ln(\pi_2/\pi_1)} + \frac{s_1/2}{\ln(\pi_2/\pi_1)}$$

Conversely, if one of these equations is true, then the first equation in the corresponding system, that is (I) or (II), respectively, is automatically satisfied.

ACKNOWLEDGEMENT

One of the authors (J. F. Viel) was supported by the Biomed 2 project 'A European initiative in advanced methods for disease mapping and risk assessment'.

REFERENCES

Bithell, J. F. (1990). 'An application of density estimation to geographical epidemiology'. *Statistics in Medicine* **9**, 691–701.

Chen, R., Mantel, N. and Klingberg, M. A. (1984). 'A study of three techniques for time-space clustering in Hodgkin's disease'. *Statistics in Medicine* **3**, 173–184.

Cleveland, W. S. and Devlin, S. J. (1988). 'Locally weighted regression: an approach to regression analysis by local fitting'. *Journal of the American Statistical Association* **83**, 596–610.

Cook-Mozzafari, P. J., Darby, S. C., Doll, R., Forman, D., Hermon, C. and Pike, M. C. (1989). 'Geographical variation in mortality from leukaemia and other cancers in England and Wales in relation to proximity to nuclear installations, 1969–78'. *British Journal of Cancer* **59**, 476–485.

Ederer, F., Myers, M. H. and Mantel, N. (1966). 'A statistical problem in space and time: Do leukaemia cases come in clusters?'. *Biometrics* **20**, 626–638.

Fleming, T. R. and DeMets, D. L. (1993). 'Monitoring of clinical trials: issues and recommendations'. *Control. Clin. Trials* **14**, 183–197.

Gardner, M. J. (1989). 'Review of reported increases of childhood cancer rates in the vicinity of nuclear installations in the United Kingdom'. *Journal of the Royal Statistical Society Series A* **152**, 307–325.

Hills, M. and Alexander, F. (1989). 'Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review'. *Journal of the Royal Statistical Society Series A* **152**, 353–363.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Knox, E. G. (1964). 'The detection of space-time interactions'. *Applied Statistics* **13**, 25–29.

Lan, K. K. and DeMets, D. L. (1983). 'Discrete sequential boundaries for clinical trials'. *Biometrika* **70**, 659–663.

Lan, K. K. and Zucker, D. M. (1993). 'Sequential monitoring of clinical trials: the role of information and Brownian motion'. *Statistics in Medicine* **12**, 753–765.

Lawson, A. B. (1993). 'On the analysis of mortality events around a prespecified fixed point'. *Journal of the Royal Statistical Society Series A* **156**, 363–377.

Lawson, A. B. and Viel, J. F. (1995). 'Tests for directional space-time interaction in epidemiological data'. *Statistics in Medicine* **14**, 2383–2391.

Lawson, A. B. and Waller, L. (1996). 'A review of point pattern methods for spatial modelling of events around sources of pollution'. *Environmetrics* **7**, 471–487.

Lee, J. W. (1994). 'Group sequential testing in clinical trials with multivariate observations: a review'. *Statistics in Medicine* **13**, 101–111.

Lin, D. Y., Wei, L. J. and DeMets, D. L. (1991). 'Exact statistical inference for group sequential trials'. *Biometrics* **47**, 1399–1408.

Mantel, N. (1967). 'The detection of disease clustering and a generalized regression approach'. *Cancer Research* **27**, 209–220.

Mehta, C. R., Patel, N., Senchaudhuri, P. and Tsiatis, A. (1994). 'Exact permutational tests for group sequential clinical trials'. *Biometrics* **50**, 1042–1053.

Miller, R. G. (1981). *Simultaneous Statistical Inference*, 2nd Edition. New York: Springer.

Pawitan, Y. and Hallstrom, A. (1990). 'Statistical interim monitoring of the cardiac arrhythmia suppression trial'. *Statistics in Medicine* **9**, 1081–1090.

Stone, R. A. (1988). 'Investigations of excess environmental risks around putative sources: statistical problems and a proposed test'. *Statistics in Medicine* **7**, 649–660.

Tango, T. (1984). 'The detection of disease clustering in time'. *Biometrics* **40**, 15–20.

Viel, J. F. and Richardson, S. T. (1990). 'Childhood leukaemia around the La Hague nuclear waste reprocessing plant'. *British Medical Journal* **300**, 580–581.

Viel, J. F., Richardson, S. T., Danel, P., Boutard, P., Malet, M., Barrelier, P., Reman, O. and Carré, A. (1993). 'Childhood leukemia incidence in the vicinity of La Hague nuclear-waste reprocessing facility (France)'. *Cancer Causes and Control* **4**, 341–343.

Viel, J. F., Pobel, D. and Carre, A. (1995). 'Incidence of leukaemia in young people around the La Hague nuclear waste reprocessing plant: a sensitivity analysis'. *Statistics in Medicine* **14**, 2459–2472.

Waller, L., Turnbull, F., Clark, L. and Nasca, P. (1994). 'Spatial pattern analysis to detect rare disease clusters'. In *Case Studies in Biometry*, eds. N. Lang and L. Ryan. New York: Wiley.
Wetherill, G. B. (1986). *Sequential Methods in Statistics*, 3rd Edition. London: Chapman and Hall.
Whitehead, J. (1992). *The Design and Analysis of Sequential Clinical Trials*, 2nd Edition. New York: Ellis Horwood.
Whittemore, A. S., Friend, N., Brown, B. W. and Holly, E. A. (1987). 'A test to detect clusters of disease'. *Biometrika* **74**, 631–635.

