Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 126 (2004) 173–187

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Scheduling optimal examination times in a simple illness–death model

Atanu Biswas*, Anup Dewanji

Applied Statistics Unit, Indian Statistical Institute, 203, B.T. Road, Calcutta 700 035, India

Received 5 February 2002; accepted 19 July 2003

Abstract

Simple illness-death model arises in many medical and animal experiments as well as industrial applications. In a typical simple illness–death model, two types of events are of interest: the event of occurrence of disease or illness (D) which is assumed to be unobservable, and the event of failure or death (F) which is observable with or without disease. With the development of methodologies for making inference on the distribution of D, the design issue has also attracted some attention although not so greatly. In this work, the objective is to find an optimal design to safeguard the event of failure with illness before it is detected at an examination time. The standard likelihood-based criteria are difficult to apply since calculation of the expected information matrix is not straightforward. Here, we consider finding $K(\geqslant 1)$ optimal intermediate examination times during the span of a study involving simple illness–death model, using some new criteria. The performance of these criteria are investigated through a number of characteristics and comparisons are made. Some of the criteria can be suitably adjusted to make them adaptive in the sense that the subsequent examination time depends on the past and current data.

Keywords: Illness–death model; Optimal designs; Cancer screening studies; Performance characteristics; Adaptive designs

1. Introduction

In a typical illness–death model, one is often interested in the event of occurrence of disease or illness (D) which is assumed to be unobservable, but detectable with probability one through a diagnostic test. Primary interest in such a study is to detect the occurrence of D or to infer about the distribution of D. This event is followed

* Corresponding author. Tel.: +91-33-25772088; fax: +91-33-25773104.
E-mail addresses: atanu@isical.ac.in (A. Biswas), dewanjia@isical.ac.in (A. Dewanji).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.07.008

by the event of failure or death (F) which is observable in addition to the presence of disease. Failure (F) can also be observed before the disease occurs in which case the absence of disease is also recorded. Let us, by D or F, also mean, for notational convenience, the corresponding time of event whenever the context is such. Examples of simple illness–death models can be found in animal carcinogenicity experiments (e.g., Kodell and Nelson, 1980; Turnbull and Mitchell, 1984), where occurrence of a particular type of cancer (D) and death (F) is considered. Kalbfleisch and Lawless (1989) gave example of medical experiments involving infection with HIV (D) and detection and diagnosis of the disease AIDS (F). Industrial applications consist of the appearance of a fault (D) and breakdown of the machine (F). With the development of methodologies for making inference on the distribution of D, the design issue has also attracted some attention although not so greatly. Bergman and Turnbull (1983), Berry (1975) and Borgan et al. (1984) did some work in this direction. For a detailed discussion, see Dewanji and Biswas (2001), referred to as DB from now on, who considered finding optimal termination time or one intermediate inspection time in simple illness–death model.

In this work, we consider designing $K(\geqslant 1)$ examination times during the span of a study involving simple illness–death model in order to make inference on the distribution of D and/or make early diagnosis of disease D. The number of actual examinations may be less than K because, once an examination detects presence of disease, further examinations are not necessary for gathering information on D. In cancer screening studies, it is important to schedule the visits of susceptible patients for routine examinations (see Shapiro et al. (1982) and Tabar et al. (1985) in the context of breast cancer screening. See also Day and Walter, 1984). Early detection of cancer helps in taking a corrective step towards cure. There are some ad hoc approach to schedule the visits usually once in every year or so, although a more objective criterion to choose one or more time points for intermediate observation will be of interest. See Baker and Chu (1990) and Hu and Zelen (1997) for some discussion. Zelen (1993) addressed the same problem of designing K fixed examination times for early detection of disease, but in a somewhat different context. Instead of a simple illness–death model, he considered a simple compartment model with three irreversible compartments in a line representing healthy, pre-clinical disease and clinically detectable disease states, respectively. He introduced a utility function which is effectively the sum of weighted differences, at different examination times, between the probability of finding the pre-clinical disease and the probability of finding the clinically detectable disease. Optimal examination times are obtained by maximizing this utility function with respect to the K examination times simultaneously. Parmigiani (1993a) developed a decision theoretic approach based on cost consideration for choosing optimal examination times. The implementation, however, does not seem to be as simple. See also Parmigiani (1993b, 1997) in this context. Lee and Zelen (1998) illustrated how a properly framed screening schedule, essentially the same as that of Zelen (1993), as an alternative to annual screening, can lead to the savings of millions of dollars for breast cancer screening among women in the United States aged 50 and older.

In contrast with Zelen's optimal scheduling, the approach introduced in this paper is sequential in nature in the sense that the optimal examination times are chosen one at a

time, the later choices depending on the earlier examination times and also the outcome of those examinations. The optimality criterion is, as in Zelen (1993), probability-based and the objective function changes at different steps incorporating the outcome of the earlier examinations. As commented in Zelen (1993) and also in DB, this approach can also be applied in industrial context for scheduling intermediate inspection times for early detection of faults before complete break-down of machines.

In Section 2, we introduce a number of optimality criteria for different purposes and a model for the joint distribution of D and F. The optimal design clearly depends on the model parameters which need to be estimated based on data from some other sources or from previous stages in a multistage framework. Estimation of the model parameters based on relevant data has been discussed elsewhere (see, for example, DB). Section 3 considers the optimal scheduling of examination times satisfying the different criteria. In Section 4, we discuss some performance characteristics of the designs, namely the expected number of examinations (out of K), a risk function associated with the optimal scheduling, and the probability of detecting disease before failure occurs. We also consider sum of the absolute differences between the K quantiles of the distribution of D and the K examination times as a measure of the amount of information collected on the distribution of D. Section 5 presents applicability of the optimal criteria for more general models. In Section 6, we discuss the adaptive version of the problem in which the estimates of the model parameters are updated at every sequential step and the updated estimates are used to find the successive optimal inspection times. Section 7 ends with a discussion.

2. Preliminaries

The design problem discussed in the last section has not been addressed much in the literature mainly because of the difficulties involved in calculating and dealing with the expected information matrix in order to obtain the optimal designs satisfying the standard optimality criteria (e.g., A, $\mathscr{D}$, c-optimality). Computation of the expected Fisher's information is numerically a difficult task. For single examination time, this was attempted in our earlier paper (Dewanji and Biswas, 2001) in the absence of censored and missing data. In the presence of censoring or missing data, this computation requires knowledge about the corresponding censoring/missing mechanism, which is usually unknown in practice, making this computation even theoretically intractable. For this reason, we introduce below some simple optimality criteria which are easy to deal with and have compelling intuitive appeal, extending the idea introduced in DB. In this work, the objective is to make efficient inference on the distribution of D and/or to safeguard the event of failure with illness before it is detected at an examination time. Since D is the event of interest, we like to have as much information on D as possible at the examination times. As a natural generalization of the argument used in DB, either D should occur by the first examination time, t_1, say; if not, then D should occur by the second examination time, t_2 say; and so on. One would, therefore, be tempted to choose the first examination time t_1 by maximizing $P[D < t_1 < F]$, as in DB, and, if disease is not found by time t_1, then to choose the second examination time t_2 by maximizing

$P[D < t_2 < F | t_1 < D, F]$, and so on. This is criterion $\mathscr{C}_1$. In order to provide some protection against disease and failure due to disease taking place before the intermediate observation times, we can maximize $P[D < t_1 < F] - P[D < F < t_1]$ to find t_1, and subsequently maximize $P[D < t_j < F | t_{j-1} < D, F] - P[D < F < t_j | t_{j-1} < D, F]$ to find t_j, $j = 2, 3, \ldots, K$. Call this criterion as $\mathscr{C}_2$. This design should be able to detect disease before failure with higher probability. Note that both $\mathscr{C}_1$ and $\mathscr{C}_2$ reduce to those in DB for $K = 1$. However, by this natural generalization of those in DB, t_1 is same as the optimal time point when only one intermediate examination is allowed (see DB) resulting in a larger t_1 (as earlier time points may be chosen since K examinations are allowed by the design). The subsequent examination times, in this case, may not be necessary at all. Therefore, instead of risking the information on possible early occurrence of D, it is reasonable to spread the K examination times evenly, in some sense, over the range of D, that is $(0, \infty)$. For this purpose, we choose K time points $d_1 < \cdots < d_K$, say, in the range of D. We first choose d_1 satisfying $P[D < d_1] = 1/(K+1)$. Then, after finding t_{j-1}, the $(j-1)$th optimum examination time, d_j is chosen satisfying $P[t_{j-1} < D < d_j | D > t_{j-1}] = 1/(K - j + 2)$. That is, d_j is taken as the first of the $(K - j + 1)$ quantiles of the residual life time beyond t_{j-1} (with $t_0 = 0$). As a principle, we like the jth examination time not to exceed d_j, for $j = 1, \ldots, K-1$.

Then, we may choose the optimal examination times as

$$t_j = \min\{d_j, \arg\max_t P[D < t < F | t_{j-1} < D, F]\}, \quad 1 \leqslant j \leqslant K-1,$$

$$t_K = \arg\max_t P[D < t < F | t_{K-1} < D, F].$$

Choosing t_K is equivalent to choosing one optimal inspection time from the residual life time after t_{K-1} conditional on $[t_{K-1} < D, F]$. Let us denote this criterion by $\mathscr{C}_3$. Unlike the design by $\mathscr{C}_1$ (also $\mathscr{C}_2$), this criterion $\mathscr{C}_3$ will make more use of the K allowable examinations. As a result, it is likely to have higher probability of detecting disease before failure occurs. Also this will have more information on the distribution of D.

The above criterion $\mathscr{C}_3$ has been designed to collect as much information on D as possible so as to be able to estimate the distribution of D somewhat efficiently (see DB). There is also the other purpose of detecting the disease as early as possible for which an alternative criterion, as in DB, can be suggested by having a protection against the event of disease and failure taking place before the subsequent scheduled examination time. We achieve this, as in $\mathscr{C}_2$, by maximizing the difference between two conditional probabilities, $P[D < t < F | t_{j-1} < D, F] - P[D < F < t | t_{j-1} < D, F]$, with respect to t, for the optimal choice of the jth examination time, for $j = 1, \ldots, K$. However, after imposing the condition that the jth examination time should not exceed d_j, we have the optimal examination times as

$$t_j = \min\{d_i, \arg\max_t P[D < t < F | t_{j-1} < D, F] - P[D < F < t | t_{j-1} < D, F]\},$$

$$1 \leqslant j \leqslant K-1,$$

$$t_K = \arg\max_t P[D < t < F | t_{K-1} < D, F] - P[D < F < t | t_{K-1} < D, F].$$

Let us denote this criterion by $\mathscr{C}_4$. This criterion is particularly useful in studies in which early diagnosis of disease is important to start some corrective treatment (e.g., cancer screening studies), in addition to be able to collect information on the distribution of D. However, as noted in DB, the second term in the difference pulls down the optimal value of t_j's from those obtained by maximizing only the first term (i.e., the criterion $\mathscr{C}_3$).

Also, as in Zelen (1993), we can introduce a utility function given by

$$U = \sum_{j=1}^{K} \{P[t_{j-1} < D < t_j < F] - P[t_{j-1} < D < F < t_j]\}, \tag{2.1}$$

which is the sum of differences, at different examination times, between the probability of detecting disease D and the probability of death due to disease taking place before the examination. The optimal examination times can be obtained by maximizing U with respect to $t_1, \ldots, t_K$ simultaneously. Call this criterion $\mathscr{C}_5$.

The optimal examination times by any of the above criteria act on individual basis as follows. The first examination for an individual is scheduled at time t_1. If disease D without failure is found at t_1, or failure F with or without disease occurs before time t_1 (that is, $D < t_1 < F$ or $D < F < t_1$ or $F < D, t_1$), then no further examination is scheduled. Note that an examination is performed only in the first of the above three possibilities for the purpose of obtaining average number of examinations (see Section 4), although examination is also performed in the other two cases of failures for the diagnostic purpose of ascertaining presence or absence of disease. However, if disease is not found at time t_1 (that is, $t_1 < D, F$), then another examination is scheduled at time t_2, and so on.

We start with the simple parametric assumption of exponential distribution for D with parameter α having density $f(x) = \alpha e^{-\alpha x}$, $\alpha > 0$, $x > 0$. The assumed conditional distribution of F, given $D = x$, is described in terms of its hazard rate which is a constant β up to x and then another constant γ (see Freund, 1961). Corresponding conditional density can be written as

$$g(y|x) = \begin{cases} \beta e^{-\beta y} & \text{if } y < x, \\ \gamma e^{-\beta x - \gamma(y-x)} & \text{if } y \geqslant x. \end{cases} \tag{2.2}$$

This model is used to find optimal scheduling of examinations and obtain their performances in Sections 3 and 4, respectively. The approach is simple and flexible enough for other general models and this is noted in Section 5.

3. Optimal scheduling

Note that the parameters α, β, γ are assumed to be known either as estimates based on data from a pilot survey or as guesses from past experience. Simple probability calculation gives, for model (2.2),

$$P[D < t < F | t_{j-1} < D, F] = \frac{\alpha}{\alpha + \beta - \gamma} [e^{-\gamma(t - t_{j-1})} - e^{-(\alpha+\beta)(t - t_{j-1})}].$$

Maximizing this with respect to t, it is easy to show that the optimal examination times, for the optimality criterion $\mathscr{C}_1$, are given by

$$t_j = t_{j-1} + \frac{\log((\alpha+\beta)/\gamma)}{\alpha+\beta-\gamma}, \quad j = 1,\ldots,K. \tag{3.1}$$

Similarly, for $\mathscr{C}_2$, one can calculate the difference $P[D < t < F | t_{j-1} < D, F] - P[D < F < t | t_{j-1} < D, F]$ as

$$\frac{2\alpha}{\alpha+\beta-\gamma}[e^{-\gamma(t-t_{j-1})} - e^{-(\alpha+\beta)(t-t_{j-1})}] - \frac{\alpha}{\alpha+\beta}[1 - e^{-(\alpha+\beta)(t-t_{j-1})}]$$

and, maximizing this with respect to t, the optimal examination times come out to be

$$t_j = t_{j-1} + \frac{1}{\alpha+\beta-\gamma}\log\left(\frac{\alpha+\beta+\gamma}{2\gamma}\right), \quad j = 1,\ldots,K. \tag{3.2}$$

Note that, for both the criteria, the optimal examination times are equispaced, as observed also by Zelen (1993) in a simpler problem but with a different criterion.

Next, using (3.1), the optimal examination times by the optimality criterion $\mathscr{C}_3$, are

$$t_j = \min\left\{d_j, t_{j-1} + \frac{\log((\alpha+\beta)/\gamma)}{\alpha+\beta-\gamma}\right\}, \quad j = 1,\ldots,K-1$$

and

$$t_K = t_{K-1} + \frac{\log((\alpha+\beta)/\gamma)}{\alpha+\beta-\gamma}$$

and, by the optimality criterion $\mathscr{C}_4$, using (3.2), the optimal examination times are

$$t_j = \min\left\{d_j, t_{j-1} + \frac{1}{\alpha+\beta-\gamma}\log\left(\frac{\alpha+\beta+\gamma}{2\gamma}\right)\right\}, \quad j = 1,\ldots,K-1$$

and

$$t_K = t_{K-1} + \frac{1}{\alpha+\beta-\gamma}\log\left(\frac{\alpha+\beta+\gamma}{2\gamma}\right).$$

Note that, for the exponential distribution of D as in Section 2, d_j turns out to be $-(1/\alpha)\log(1-1/(K-j+2)) + t_{j-1}$.

In order to use the criterion $\mathscr{C}_5$, model (2.2) gives the expression for the utility function (2.1) as

$$U = \sum_{j=1}^{K}\left\{\frac{2\alpha e^{-\gamma t_j}}{\alpha+\beta-\gamma}[e^{-(\alpha+\beta-\gamma)t_{j-1}} - e^{-(\alpha+\beta-\gamma)t_j}]\right.$$

$$\left. - \frac{\alpha}{\alpha+\beta}[e^{-(\alpha+\beta)t_{j-1}} - e^{-(\alpha+\beta)t_j}]\right\}.$$

Table 1
Optimal scheduling for different $\mathscr{C}_j$'s with $K = 3$ and 5

Parameters			$K = 3$			$K = 5$				
α	β	γ	t_1	t_2	t_3	t_1	t_2	t_3	t_4	t_5
			4.62	9.24	13.86	4.62	9.24	13.86	18.48	23.10
			2.70	5.41	8.11	2.70	5.41	8.11	10.81	13.52
0.2	0.1	0.15	1.44	3.47	6.93	0.91	2.03	3.47	5.49	8.96
			1.44	3.47	6.17	0.91	2.03	3.47	5.49	8.20
			1.49	3.41	6.11	1.03	2.24	3.73	5.65	8.35
			5.75	11.51	17.26	5.75	11.51	17.26	23.01	28.77
			3.08	6.17	9.25	3.08	6.17	9.25	12.33	15.42
0.1	0.1	0.15	2.88	6.93	12.69	1.82	4.05	6.93	10.99	16.74
			2.88	5.96	9.04	1.82	4.05	6.93	10.01	13.10
			1.87	4.20	7.28	1.35	2.92	4.80	7.12	10.21
			21.07	42.14	63.22	21.07	42.14	63.22	84.29	105.36
			10.81	21.63	32.44	10.81	21.63	32.44	43.25	54.07
0.02	0.03	0.045	14.38	34.66	55.73	9.12	20.27	34.66	54.93	76.00
			10.81	21.63	32.44	9.12	19.93	30.74	41.56	52.37
			6.86	15.25	26.06	5.06	10.88	17.74	26.12	36.94

Maximizing U with respect to the t_j's simultaneously, we obtain the following recursive relation for the optimal Δ_i's, where $\Delta_i = t_i - t_{i-1}$:

$$e^{-\gamma\Delta_{j+1}} = \frac{\alpha+\beta-\gamma e^{(\alpha+\beta-\gamma)\Delta_j}}{\alpha+\beta-\gamma}, \quad j = 1,\ldots,K-1,$$

$$\Delta_K = \frac{1}{\alpha+\beta-\gamma}\log\left[\frac{\alpha+\beta+\gamma}{2\gamma}\right]. \tag{3.3}$$

The optimal examination times $t_1,\ldots,t_K$, can now be obtained from the Δ_j's in (3.5). Unlike Zelen (1993), they are not equispaced. Tables 1 and 2 give the optimal t_j's for different criteria for some combinations of parameters. The five entries in each cell correspond to the five criteria $\mathscr{C}_1$–$\mathscr{C}_5$, given by (3.1)–(3.5), respectively.

In order to save space in the following tables, we denote the set of parameters (α, β, γ) by θ and the three parameter combinations of Table 1 by θ_1, θ_2 and θ_3, respectively.

We see in both Tables 1 and 2, the optimal examinations times by $\mathscr{C}_1$ extend far beyond the expected value of D, given by $1/\alpha$, specially for large K, whereas those by $\mathscr{C}_3$ are somewhat evenly located on either sides of $1/\alpha$. Those by $\mathscr{C}_2$ and $\mathscr{C}_4$ are less than the corresponding ones by $\mathscr{C}_1$ and $\mathscr{C}_3$, respectively, and the ones by $\mathscr{C}_5$ are generally lower. Hence, the criterion $\mathscr{C}_3$ seems a reasonable one and $\mathscr{C}_5$ seems to be too conservative.

Table 2
Optimal scheduling for different $\mathscr{C}_j$'s with $K = 10$

θ	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
	4.62	9.24	13.86	18.48	23.10	27.73	32.35	36.97	41.59	46.21
	2.70	5.41	8.11	10.81	13.52	16.22	18.92	21.62	24.33	27.03
θ_1	0.48	1.00	1.59	2.26	3.03	3.94	5.06	6.50	8.52	11.99
	0.48	1.00	1.59	2.26	3.03	3.94	5.06	6.50	8.52	11.23
	0.58	1.22	1.92	2.70	3.59	4.62	5.84	7.32	9.24	11.95
	5.75	11.51	17.26	23.01	28.77	34.52	40.28	46.03	51.78	57.54
	3.08	6.17	9.25	12.33	15.42	18.50	21.58	24.66	27.75	30.83
θ_2	0.95	2.01	3.18	4.52	6.06	7.88	10.12	12.99	17.05	22.80
	0.95	2.01	3.18	4.52	6.06	7.88	10.12	12.99	16.08	19.16
	0.80	1.67	2.63	3.69	4.88	6.23	7.80	9.67	12.00	15.08
	21.07	42.14	63.22	84.29	105.36	126.43	147.50	168.58	189.65	210.72
	10.81	21.63	32.44	43.25	54.07	64.88	75.69	86.51	97.32	108.13
θ_3	4.77	10.03	15.92	22.60	30.31	39.42	50.58	64.96	85.24	106.31
	4.77	10.03	15.92	22.60	30.31	39.42	50.23	61.05	71.86	82.68
	3.07	6.39	10.03	14.04	18.51	23.56	29.38	36.25	44.63	55.45

4. Performance characteristics

In the previous two sections, we introduced five optimality criteria for different purposes. In this section, we introduce four different quantities as performance characteristics associated with different purposes and study the performance of the five optimal designs with respect to these characteristics using model (2.2).

The first characteristic we consider is the average number of examinations (out of K), denoted by $E[N]$, where N is the number of examinations actually used. Since each examination has a cost associated with it, the average number of examinations may seem to be a reasonable characteristic to consider. However, if cost is not of much concern, and noting that more number of examinations at optimally chosen time points lead to more information on the distribution of D and proper utilization of the K allowable examinations, larger values of $E[N]$ is desirable. The distribution of N can be obtained as

$$P(N \geqslant j) = P[t_{j-1} < D < t_j < F] + P[t_j < D, F]$$

for $j = 1, \ldots, K$. Hence, $E[N]$ can be derived as

$$E[N] = \sum_{j=1}^{K} \{P(t_{j-1} < D < t_j < F) + P(t_j < D, F)\}$$

$$= \sum_{j=1}^{K} \left\{ \frac{\alpha}{\alpha+\beta-\gamma} e^{-\gamma(t_j - t_{j-1})} + \frac{\beta-\gamma}{\alpha+\beta-\gamma} e^{-(\alpha+\beta)(t_j - t_{j-1})} \right\} e^{-(\alpha+\beta)t_{j-1}}. \tag{4.1}$$

Of the five criteria proposed in Section 2, $\mathscr{C}_1$ and $\mathscr{C}_2$ are likely to have less values of $E[N]$, with $\mathscr{C}_1$ having lesser.

Since, in most practical situations, the aim is to reduce the time between disease occurrence and diagnosis of disease so that a treatment, if needed, can start at the earliest, let us define 'loss' as the time duration when the patient was disease affected but not diagnosed (and hence no treatment is given). Therefore, the risk (R) is the expected time duration when there is disease but not diagnosed. Thus the risk (R) is

$$\begin{aligned} R = \sum_{j=1}^{K} \{ & E(t_j - D | t_{j-1} < D < t_j < F) P(t_{j-1} < D < t_j < F) \\ & + E(F - D | t_{j-1} < D < F < t_j) P(t_{j-1} < D < F < t_j) \} \\ & + E(F - D | t_K < D < F) P(t_K < D < F). \end{aligned}$$

It can be seen that, for model (2.2),

$$R = \sum_{j=1}^{k} (t_j A_{2j} - A_{1j} + A_{3j} - A_{4j}) + (A_5 - A_6), \tag{4.2}$$

where

$$\begin{aligned} A_{1j} &= \mathrm{e}^{-\gamma t_j} \left\{ \frac{\alpha}{\alpha+\beta-\gamma} [t_{j-1} \mathrm{e}^{-(\alpha+\beta-\gamma)t_{j-1}} - t_j \mathrm{e}^{-(\alpha+\beta-\gamma)t_j}] \right. \\ &\quad \left. + \frac{\alpha}{(\alpha+\beta-\gamma)^2} [\mathrm{e}^{-(\alpha+\beta-\gamma)t_{j-1}} - \mathrm{e}^{-(\alpha+\beta-\gamma)t_j}] \right\}, \\ A_{2j} &= \mathrm{e}^{-\gamma t_j} \frac{\alpha}{\alpha+\beta-\gamma} [\mathrm{e}^{-(\alpha+\beta-\gamma)t_{j-1}} - \mathrm{e}^{-(\alpha+\beta-\gamma)t_j}], \\ A_{3j} &= \frac{\alpha}{\alpha+\beta} [t_{j-1} \mathrm{e}^{-(\alpha+\beta)t_{j-1}} - t_j \mathrm{e}^{-(\alpha+\beta)t_j}] \\ &\quad + \left\{ \frac{\alpha}{(\alpha+\beta)^2} + \frac{\alpha}{\gamma(\alpha+\beta)} \right\} [\mathrm{e}^{-(\alpha+\beta)t_{j-1}} - \mathrm{e}^{-(\alpha+\beta)t_j}] \\ &\quad - (t_j + \gamma^{-1}) \mathrm{e}^{-\gamma t_j} \frac{\alpha}{\alpha+\beta-\gamma} [\mathrm{e}^{-(\alpha+\beta-\gamma)t_{j-1}} - \mathrm{e}^{-(\alpha+\beta-\gamma)t_j}], \\ A_{4j} &= \frac{\alpha}{\alpha+\beta} [t_{j-1} \mathrm{e}^{-(\alpha+\beta)t_{j-1}} - t_j \mathrm{e}^{-(\alpha+\beta)t_j}] + \frac{\alpha}{(\alpha+\beta)^2} [\mathrm{e}^{-(\alpha+\beta)t_{j-1}} - \mathrm{e}^{-(\alpha+\beta)t_j}] \\ &\quad - \mathrm{e}^{-\gamma t_j} \frac{\alpha}{(\alpha+\beta-\gamma)^2} [\mathrm{e}^{-(\alpha+\beta-\gamma)t_{j-1}} - \mathrm{e}^{-(\alpha+\beta-\gamma)t_j}] \\ &\quad - \mathrm{e}^{-\gamma t_j} \frac{\alpha}{\alpha+\beta-\gamma} [t_{j-1} \mathrm{e}^{-(\alpha+\beta-\gamma)t_{j-1}} - t_j \mathrm{e}^{-(\alpha+\beta-\gamma)t_j}] \end{aligned}$$

and $A_5 = A_{3j}$ at $t_{j-1} = t_K$ and $t_j = \infty$, and $A_6 = A_{4j}$ at $t_{j-1} = t_K$ and $t_j = \infty$. It is expected that, for large K, $\mathscr{C}_1$ will have more risk than at least $\mathscr{C}_3$.

Another practical aim is to be able to detect disease before failure occurs with high probability. Therefore, a third characteristic is

$$P = \sum_{j=1}^{K} P[t_{j-1} < D < t_j < F]$$

$$= \sum_{j=1}^{K} \left\{ \frac{\alpha}{\alpha+\beta}[e^{-(\alpha+\beta)t_{j-1}} - e^{-(\alpha+\beta)t_j}] \right.$$

$$\left. - \frac{\alpha e^{-\gamma t_j}}{\alpha+\beta-\gamma}[e^{-(\alpha+\beta-\gamma)t_{i-1}} - e^{-(\alpha+\beta-\gamma)t_i}] \right\}, \tag{4.3}$$

which is likely to be lower for the criterion $\mathscr{C}_1$.

When the purpose is to estimate the distribution of D, the criterion $\mathscr{C}_3$ is likely to be better. In order to have as much information as possible on the distribution of D, without having to calculate the related information matrix, it may be desirable to choose the K examination times as close as possible to the K corresponding quantiles. Therefore, we introduce the following characteristic given by the sum of absolute differences between the optimal examination times and the corresponding quantiles of the distribution of D. Formally, this characteristic Q is defined as

$$Q = \sum_{j=1}^{K} |t_j - q_j|$$

$$= \sum_{j=1}^{K} \left| t_j + \frac{1}{\alpha}\log\left(1 - \frac{j}{K+1}\right) \right|, \tag{4.4}$$

where $q_j, j = 1, \ldots, K$, are the K quantiles of the distribution of D given by $P[q_{j-1} < D < q_j] = 1/(K+1)$, for $j = 1, \ldots, K$, with $q_0 = 0$. For the exponential distribution of D as in Section 2, $q_j = -(1/\alpha)\log(1 - j/(K+1))$. Note that, as long as $t_{j-1} = d_{j-1}$, d_j is same as the q_j for the exponential model of D.

In Table 3, we present the values of different performance characteristics for model (2.2) and all the five design criteria (presented as the five entries in each cell) calculated for the same three combinations of parameters as in Tables 1 and 2. We have worked with many other parameter combinations, the results of which could not be presented here due to space constraint; however, the qualitative features are the same.

As expected, $E[N]$ is the lowest for $\mathscr{C}_1$. We also note that this is highest for $\mathscr{C}_5$ conforming with its conservative nature. In terms of R, although $\mathscr{C}_1$ performs better for small K, it becomes worse for $K = 10$, where $\mathscr{C}_3$ gets better; $\mathscr{C}_5$ does not seem to perform well in this regard. With respect to P, $\mathscr{C}_1$ performs the worst and $\mathscr{C}_3$ the best with others being close to $\mathscr{C}_3$. However, in terms of Q, $\mathscr{C}_3$ stands out ahead of others and $\mathscr{C}_1$ and $\mathscr{C}_5$ seem to perform badly. Therefore, $\mathscr{C}_3$ seems to be the best criterion in general, although, for specific purposes, others may perform better.

Table 3
Performance characteristics for different $\mathscr{C}_j$'s

θ	$K=3$				$K=5$				$K=10$			
	$E(N)$	R	P	Q	$E(N)$	R	P	Q	$E(N)$	R	P	Q
	0.77	1.89	0.44	15.89	0.78	1.52	0.44	48.46	0.78	1.48	0.44	209.78
	1.22	2.54	0.49	4.38	1.31	1.34	0.52	19.69	1.33	0.90	0.53	104.30
θ_1	1.62	2.89	0.49	0.00	2.47	1.91	0.55	0.00	4.51	1.00	0.60	0.76
	1.64	3.23	0.48	0.76	2.48	2.14	0.54	0.76	4.52	1.12	0.60	0.76
	1.64	3.26	0.48	0.93	2.38	2.09	0.54	1.36	4.11	1.00	0.60	4.69
	0.75	2.05	0.30	10.85	0.77	1.39	0.31	44.59	0.77	1.28	0.31	227.70
	1.32	3.38	0.33	5.59	1.49	1.78	0.37	9.54	1.56	0.81	0.39	80.82
θ_2	1.23	2.47	0.34	1.18	1.92	1.51	0.39	1.18	3.62	0.71	0.44	1.18
	1.36	3.44	0.33	5.79	1.98	2.09	0.38	5.79	3.64	0.96	0.44	5.79
	1.67	4.07	0.32	10.31	2.45	2.80	0.38	15.31	4.23	1.44	0.43	24.29
	0.74	6.65	0.23	20.27	0.77	4.15	0.24	107.52	0.77	3.61	0.24	715.24
	1.37	11.65	0.25	53.48	1.59	6.43	0.29	52.47	1.70	2.49	0.31	174.53
θ_3	0.97	7.11	0.25	13.59	1.55	4.12	0.29	13.59	3.00	1.90	0.34	13.59
	1.37	11.65	0.25	53.48	1.71	6.62	0.29	54.85	3.04	2.71	0.34	54.85
	1.69	13.58	0.24	70.18	2.48	9.69	0.29	111.83	4.29	5.20	0.34	202.42

5. Optimal scheduling in general

Although we have demonstrated the results in Section 3 based on the simple exponential model for D and model (2.2) for F given D, the criteria $\mathscr{C}_j$'s can be easily employed for more general models. Suppose, in general, $F_1(x;\mu)$ represents the distribution of D and $F_2(y|x;\nu)$, the conditional distribution of F given $D=x$ (parameters μ and ν may be vectors). The distribution $F_2(y|x;\nu)$ needs a careful explanation. As in (2.2), this distribution can be expressed in two cases: (1) for $y<x$, this is for the failure before disease occurs and may be written as $F_2(y|x;\nu)=F_2(y;\nu)$, and (2) for $y>x$, this is for the failure after disease occurs. Then, in order to employ the criteria $\mathscr{C}_j, j=1,\ldots,4$, the following probability calculations need to be considered:

$$P[D<t<F]=\int_0^t \bar{F}_2(t|x;\nu)\,\mathrm{d}F_1(x;\mu), \tag{5.1}$$

$$P[D<F<t]=\int_0^t [F_2(t|x;\nu)-F_2(x|x;\nu)]\,\mathrm{d}F_1(x;\mu) \tag{5.2}$$

for $t'<t$,

$$P[D<t<F|t'<D,F]=\frac{P[t'<D<t<F]}{P[t'<D,t'<F]}$$

$$=\frac{\int_{t'}^t \bar{F}_2(t|x;\nu)\,\mathrm{d}F_1(x;\mu)}{\bar{F}_1(t';\mu)\bar{F}_2(t';\nu)} \tag{5.3}$$

and

$$P[D < F < t | t' < D, F] = \frac{P[t' < D < F < t]}{P[t' < D, t' < F]} = \frac{\int_{t'}^{t} [F_2(t|x;\nu) - F_2(x|x;\nu)]\, \mathrm{d}F_1(x;\mu)}{\bar{F}_1(t';\mu)\bar{F}_2(t';\nu)}, \tag{5.4}$$

where $\bar{F}_i(\cdot)$ denotes the corresponding $1-F_i(\cdot)$ for $i=1,2$. For $\mathscr{C}_5$, we need calculation of the two probabilities $P[t' < D < t < F]$ and $P[t' < D < F < t]$, as used in (2.1), which are as in the numerators of (5.3) and (5.4), respectively.

Since the integrals involved in the above probabilities may not be analytically obtained in general, the maximization required to obtain the optimal designs satisfying the $\mathscr{C}_j$'s may not have closed form solutions as in (3.1)–(3.5). For example, with model (2.2) for conditional distribution of F given D, if we assume a Weibull(λ, p) distribution for D, probabilities (5.1)–(5.4) are, respectively,

$$P[D < t < F] = \int_0^t \mathrm{e}^{-\beta x - \gamma(t-x)} \lambda^p p x^{p-1} \mathrm{e}^{-(\lambda x)^p}\, \mathrm{d}x, \tag{5.5}$$

$$P[D < F < t] = \int_0^t \mathrm{e}^{-\beta x} [1 - \mathrm{e}^{-\gamma(t-x)}] \lambda^p p x^{p-1} \mathrm{e}^{-(\lambda x)^p}\, \mathrm{d}x, \tag{5.6}$$

$$P[D < t < F | t' < D, F] = \frac{\int_{t'}^{t} \mathrm{e}^{-\beta x - \gamma(t-x)} \lambda^p p x^{p-1} \mathrm{e}^{-(\lambda x)^p}\, \mathrm{d}x}{\mathrm{e}^{-\beta t' - (\lambda t')^p}}, \tag{5.7}$$

$$P[D < F < t | t' < D, F] = \frac{\int_{t'}^{t} \mathrm{e}^{-\beta x} [1 - \mathrm{e}^{-\gamma(t-x)}] \lambda^p p x^{p-1} \mathrm{e}^{-(\lambda x)^p} \mathrm{d}x}{\mathrm{e}^{-\beta t' - (\lambda t')^p}}, \tag{5.8}$$

none of which can be analytically obtained. By calculating the integrals numerically for some different values of $(\lambda, p, \beta, \gamma)$ and for $K = 3$, we find the optimal choices of t_1, t_2 and t_3 as given in Table 4. The five entries in each cell correspond to the five criteria.

Note that the calculation of the optimal examination times by $\mathscr{C}_5$ is a relatively tedious job when we deviate from exponentiality for the distribution of D because of the simultaneous numerical maximization. The performance characteristics $E[N]$ and P (see (4.1) and (4.3)) can be obtained easily using probabilities (5.5)–(5.8); one can also easily obtain Q (see (4.4)) once the quantiles of the Weibull distribution for D are determined. But the expression for risk R (see (4.2)) is complicated and will involve difficult integrals. However, since the purpose of this section is just to demonstrate the derivation of optimal designs in general and not to compare the different criteria on the basis of the performance characteristics, which has been done in the previous section to some extent, we do not attempt here to obtain their values.

Table 4
Optimal design with Weibull distribution for D for $K = 3$

Parameters			$p = 1.2$			$p = 0.8$		
l	β	γ	t_1	t_2	t_3	t_1	t_2	t_3
			5.20	9.52	13.62	3.89	8.69	13.78
			3.19	5.90	8.50	2.18	4.81	7.58
0.2	0.1	0.15	1.77	3.70	6.60	1.05	3.17	7.52
			1.77	3.70	6.38	1.05	3.17	5.87
			1.75	3.75	6.25	1.10	2.90	5.55
			17.99	32.99	47.26	12.88	28.79	45.62
			10.46	19.38	27.97	6.95	15.35	24.13
0.05	0.03	0.05	7.10	14.80	26.35	4.21	12.60	28.50
			7.10	14.80	23.53	4.23	12.60	21.28
			6.40	13.16	21.96	3.60	9.20	17.20

6. Adaptive optimal scheduling

Although the K optimal examination times $t_1, \ldots, t_K$ satisfying any of the criteria $\mathscr{C}_1$–$\mathscr{C}_5$ can be fixed before the actual study begins, for a particular patient, in practice, not all of them may be necessary. By construction, they are in increasing order to maintain the chronological timing. However, as the study progresses and accumulates information on the disease process up to, say, the $(j-1)$th examination time, use of this information for choosing the optimal jth examination time becomes important. One can improve upon the initial optimal choice of t_j, by using the information accumulated up to t_{j-1} from all the individuals (assumed homogeneous) under study, thus making the optimal choice of t_j adaptive. We propose to use this information to update or improve the estimate of $\theta = (\alpha, \beta, \gamma)$ basing it on the current accumulated data. This can be done by using EM algorithm (Dempster et al., 1977) as described in DB. Then, the improved estimate can be used to choose the next optimal examination time.

Suppose $\hat{\theta}^{(j-1)} = (\hat{\alpha}^{(j-1)}, \hat{\beta}^{(j-1)}, \hat{\gamma}^{(j-1)})$ denotes the current estimate of θ obtained by using information up to time t_{j-1}, for $j = 1, \ldots, K$. Then, using $\mathscr{C}_1$ for example, the adaptive jth optimal examination time can be chosen as, from (3.1),

$$t_j = t_{j-1} + \frac{\log((\hat{\alpha}^{(j-1)} + \hat{\beta}^{(j-1)})/\hat{\gamma}^{(j-1)})}{\hat{\alpha}^{(j-1)} + \hat{\beta}^{(j-1)} - \hat{\gamma}^{(j-1)}}. \tag{6.1}$$

As in (6.1), the adaptive optimal choice of the jth examination time can be easily written down from (3.2)–(3.5), using the criteria $\mathscr{C}_2 - \mathscr{C}_5$, respectively. Note that in each case, t_j has the expression as t_{j-1} *plus* a positive quantity, which is being evaluated at the current estimate $\hat{\theta}^{(j-1)}$. Thus, the adaptive optimal scheduling also leads to naturally (increasing) ordered examination times, at least for model (2.2). For general models (see Section 5) also, the criteria $\mathscr{C}_1$–$\mathscr{C}_5$ lead to naturally ordered examination times since, from the definitions in Section 2, the general form of t_j can be written as t_{j-1} *plus* a positive quantity. As, by the criteria $\mathscr{C}_1$–$\mathscr{C}_4$, the optimal

examination times are chosen sequentially, making the choice adaptive in general is straightforward. However, $\mathscr{C}_5$ involves simultaneous maximization of (2.1) with respect to $t_1,\ldots,t_K$, which, unlike in (3.5), does not have closed form solution in general, so that the technique of making the optimal scheduling adaptive is not applicable. Therefore, $\mathscr{C}_5$ is not a suitable criterion for adaptive optimal scheduling in general.

The calculation of the performance characteristics in this case becomes next to impossible since the examination times t_j's themselves are random depending on parameter estimates which also keep changing over different j. One may still be interested in comparing the adaptive optimal scheduling with the corresponding non-adaptive version. Firstly, if the parameters are completely known, then there is no need of estimating them and then using them for adaptive optimal scheduling; the designs of Section 3 are to be used in this case. Secondly, if the parameters are not known and have to be estimated, then the performance characteristics can only be estimated and it is only natural that the adaptive scheduling will do better since it is based on more and more information.

7. Discussion

Since all the criteria discussed in this paper optimize some probability terms only and do not depend on the likelihood, the corresponding designs, unlike those obtained from likelihood-based criteria, do not change due to minor changes in the secondary aspects of data (for example, possibility of censoring or not). See DB for more discussion on this. Also, as mentioned in DB, another advantage of the criteria $\mathscr{C}_j, j=1,\ldots,5$, is easy incorporation of one or more covariates, denoted by $Z=z$, in the optimal design. For example, if the distribution of D happens to depend on $Z=z$ via the exponential parameter $\alpha=\alpha(z)=\alpha_0 e^{\alpha_1 z}$ (say), the optimal design $t_j=t_j(z)$ by any criterion, can be readily obtained simply by replacing α in (3.1)–(3.5) by $\alpha(z)$. Therefore, the optimal design for individuals with different Z values will be different (making it more realistic) but can be obtained by using one formula. This is extremely difficult with likelihood-based criteria.

Note that the optimal examination times implicitly depends on the knowledge that a maximum of K examinations are allowed. However, if this knowledge is not available or there is no such restriction on the number of examinations (which is the likely scenario in many cases), the criteria $\mathscr{C}_3$–$\mathscr{C}_5$ fail. One possible solution is to start with a K, and as the examinations are held, vary the value of K, depending on the current budget and information, while choosing the successive d_j's of Section 2. This also makes the choice of t_j's adaptive in some sense.

Acknowledgements

We are thankful to the referees and the Executive Editor for their valuable comments which helped in writing the revision.

References

Baker, S.G., Chu, K.C., 1990. Evaluating screening for the early detection and treatment of cancer without using a randomized control group. J. Amer. Statist. Assoc. 85, 320–327.

Bergman, S.W., Turnbull, B.W., 1983. Efficient sequential designs for destructive life testing with application to animal serial sacrifice experiments. Biometrika 70, 305–314.

Berry, G., 1975. Design of carcinogenesis experiments using Weibull distribution. Biometrika 62, 321–328.

Borgan, Ø., Liestøl, K., Ebbesen, P., 1984. Efficiencies of experimental designs for an illness–death model. Biometrics 40, 627–638.

Day, N.E., Walter, S.D., 1984. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics 40, 1–13.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood estimation from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39, 1–22.

Dewanji, A., Biswas, A., 2001. An optimal design for simple illness–death model. J. Statist. Plann. Inference 96, 289–300.

Freund, J.E., 1961. A bivariate extension of the exponential distribution. J. Amer. Statist. Assoc. 56, 971–977.

Hu, P., Zelen, M., 1997. Planning clinical trials to evaluate early detection programmes. Biometrika 84, 817–829.

Kalbfleisch, J.D., Lawless, J.F., 1989. Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. J. Amer. Statist. Assoc. 84, 360–372.

Kodell, R.L., Nelson, C.J., 1980. An illness–death model for the study of the carcinogenic process using survival/sacrifice data. Biometrics 36, 267–277.

Lee, S.J., Zelen, M., 1998. Scheduling periodic examinations for the early detection of disease: applications to breast cancer. J. Amer. Statist. Assoc. 93, 1271–1281.

Parmigiani, G., 1993a. On optimal screening ages. J. Amer. Statist. Assoc. 88, 622–628.

Parmigiani, G., 1993b. Scheduling inspections in reliability. In: Basu, A. (Ed.), Advances in Reliability. North Holland, Amsterdam, pp. 303–319.

Parmigiani, G., 1997. Timing medical examinations via intensity functions. Biometrika 84, 803–816.

Shapiro, S., Venet, W., Strax, P.H., Venet, L., Roeser, R., 1982. Ten to fourteen year effect of screening on breast cancer mortality. J. Nat. Cancer Inst. 69, 349–355.

Tabar, L., Fagerberg, C.J., Gad, A., Baldetort, L., Holmberg, L.H., Grontoft, O., Ljungquist, U., Lundstrom, B., Mansom, J.C., Ekhund, G., 1985. Reduction in mortality from breast cancer after mass screening with mammography. Lancet 1, 829–832.

Turnbull, B.W., Mitchell, T.J., 1984. Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments. Biometrics 40, 41–50.

Zelen, M., 1993. Optimal scheduling of examinations for the early detection of disease. Biometrika 80, 279–293.