STATISTICS IN MEDICINE, VOL. 11, 1705–1717 (1992)

ANALYSIS OF A CLINICAL TRIAL INVOLVING A COMBINED MORTALITY AND ADHERENCE DEPENDENT INTERVAL CENSORED ENDPOINT

LEMUEL A. MOYÉ, BARRY R. DAVIS AND C. MORTON HAWKINS

University of Texas Health Science Center, 1200 Herman Pressler, Houston, Texas 77025, U.S.A.

SUMMARY

Clinical trials often involve a variety of clinical and laboratory measures that are used as endpoints and sometimes two of these measures are combined in one endpoint. When the individual components of such a combined endpoint are 'time to event' measurements, the analysis is straightforward if each of the components is measured frequently and regularly over time. However, the analysis of the combined endpoint is more difficult when one component of the endpoint is right censored and the other is interval censored. This paper describes a statistic, based on a rank ordering of events for such a combined measure. The power of the test statistic is explored.

INTRODUCTION

It is important to develop sensitive statistical tests of treatment effects in clinical trials. Kaplan Meier life tables[1] used to analyse time to an endpoint event and the regression method of Cox[2] are examples of statistical procedures that use all available information. These 'time to event' procedures have increased the sensitivity of endpoint analysis, affording additional power to clinical trials in the detection of treatment differences. Combined endpoints are being used in clinical trials with increasing frequency.[3,4] The major advantage conferred by these more complex endpoints is that, while the infrequent occurrence of a single endpoint can leave the trial underpowered, a combination of endpoints may occur frequently enough to confer the required additional statistical power. Time to event measures may be used straightforwardly in the analysis of combined endpoints when their components are measured frequently and regularly over time.

However, different censoring characteristics for each of the individual components poses a problem for analysis not specifically addressed in the statistical literature. Although such an endpoint may be computed easily by counting the total number of events in each group and testing the null hypothesis of no treatment effect by performing a binomial test of proportions, this crude analysis does not take into account the time of each endpoint occurrence. In addition, if any of the individual components of the combined endpoint are dependent on patient co-operation, the statistical power is related to patient compliance.

This paper describes a statistical analysis for a combined endpoint whose individual components have different censoring mechanisms. Our approach requires a rank ordering of the times to event for all trial participants who experience the endpoint and is based on a U statistic. The work was motivated by an ongoing clinical trial evaluating the effect of angiotensin converting enzyme

0277–6715/92/131705–13$11.50

Received July 1991
Revised December 1991

(ACE) inhibition on postinfarction ventricular dilatation and death. The statistical power of our new test statistic is explored within the context of this trial.

METHODOLOGY

The procedure described here allows the frequency of endpoint events to be compared across the two therapy groups while taking into account the time until these events. We first introduce some notation to describe the calculation of this statistic.

Consider a clinical trial in which n patients are randomized to each of two therapy groups (control and treatment). The clinical hypothesis has two components. The first is that the treatment improves total mortality. The second is that for those patients who survive, the treatment will prevent a large deterioration in a quantitative endpoint measure M. An example of a quantitative endpoint measure might be the patient's blood pressure. When randomized, patients have a first measurement of the quantitative endpoint (M_1). Patients who die during the course of the trial may have no additional measurements of the quantitative endpoint. Those patients who survive may return at the trial's conclusion for a last measurement of the quantitative endpoint (M_L). These patients are considered to have experienced the quantitative component of the combined endpoint when they have met the threshold $\Delta M = M_1 - M_L \geqslant c$, where c is specified before the beginning of the trial. Some patients may survive the trial but not return for the last measurement M_L.

In summary, patients may experience one of four mutually exclusive occurrences related to the combined endpoint of death or survival and a deterioration in the quantitative endpoint M. These occurrences are:

1. death during the trial;
2. survival, but exceeding the threshold (that is, $\Delta M \geqslant c$);
3. survival and not exceeding the threshold ($\Delta M < c$);
4. survival, but not having M_L measured.

Let X and Y index the control group and treatment group experiences, respectively. We assume complete survival status information. The possible experiences of the ith control group patient may be described using the following notation:

1. $\delta_i = 1$ if the patient is dead and $\mathbf{T}(X_i, \delta_i = 1) = \mathbf{T}(X_i)$ time until death of this patient.
2. $\delta_i = 2$ if the patient survives and $\Delta M \geqslant c$ and $\mathbf{T}(X_i, \delta_i = 2) = \mathbf{T}(X_i^+) =$ time until M_L is measured.
3. $\delta_i = 3$ if the patient survives and $\Delta M < c$, and $\mathbf{T}(X_i, \delta_i = 3) = \mathbf{T}(X_i^*) =$ time until M_L is measured.
4. $\delta_i = 4$ if the patient survives but does not have M_L measured and $\mathbf{T}(X_i, \delta_i = 4) = \mathbf{T}(X_i^\sim) =$ time the patient spends in the trial.

Analogously, the jth patient in the treatment group has trial occurrences denoted as ε_j and a corresponding time to this occurrence denoted by $\mathbf{T}(Y_j)$, $\mathbf{T}(Y_j^+)$, $\mathbf{T}(Y_j^*)$, or $\mathbf{T}(Y_j^\sim)$, respectively.

We use a test statistic that compares each of the n patients in the control group with each of the n patients in the treatment group and scores the results of these comparisons. The assigned score when comparing control patient i with experimental treatment patient j uses both the occurrence of the event and the time to occurrence of these events and is denoted as ϕ_{ij}. In this situation, Table I describes the appropriate scoring function. These scores are accumulated over all n^2 comparisons.

Table I. Comparisons of SAVE endpoints across therapy groups

Score assigned to ϕ_{ij}	Condition	Comparison
1	$\delta_i = 1$	$\mathbf{T}(X_i) < \mathbf{T}(Y_j)$ $\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j^+)$ $\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j^*)$ $\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j^\sim)$
-1	$\varepsilon_j = 1$	$\mathbf{T}(X_i) > \mathbf{T}(Y_j)$ $\mathbf{T}(X_i^+) \geqslant \mathbf{T}(Y_j)$ $\mathbf{T}(X_i^*) \geqslant \mathbf{T}(Y_j)$ $\mathbf{T}(X_i^\sim) \geqslant \mathbf{T}(Y_j)$
a	$\delta_i = 2;\ \varepsilon_j = 3$	$\mathbf{T}(X_i^+) \leqslant \mathbf{T}(Y_j^*)$
$-a$	$\delta_i = 3;\ \varepsilon_j = 2$	$\mathbf{T}(X_i^*) \geqslant \mathbf{T}(Y_j^+)$
0	otherwise	

The function used to score the comparison of the ith control group patient's experience to the experience of the jth patient in the treatment group ($j = 1, \ldots, n$) is based on a hierarchy of the four possible trial experiences. If the result of the comparison favours the treatment group, ϕ_{ij} is positive. If the result favours the control group, ϕ_{ij} is negative. For example, if the ith control group patient has died during the follow-up period of the trial and the jth treatment group patient has survived (regardless of the status of M_L), then ϕ_{ij} is positive ($\phi_{ij} = 1$). If the jth treatment group patient died during the course of the trial and the ith control group patient survived (again, regardless of the status of M_L), then ϕ_{ij} is negative ($\phi_{ij} = -1$). This scoring system allows a death to take precedence over survival. If both the ith control group patient and the jth treatment group patient die, than the value of the score depends on the time to death of each of these patients. If the death occurred earlier in the follow-up experience for the control group patient, then $\phi_{ij} = 1$. If the earlier death was for a treatment group patient, $\phi_{ij} = -1$. Thus, in the circumstance where a death has occurred, the non-zero score only applies if the time of death of one patient is prior to the time of censorship of the other. This scoring function is that used in the Gehan statistic.[5]

The scoring function is expanded to include comparisons among patients who survived the trial. These additional comparisons are dependent on the occurrence of the quantitative endpoint measure, ΔM. At the trial's conclusion, if the ith control group patient survived and experienced $\Delta M \geqslant c$ and the jth treatment group patient survived and had $\Delta M < c$, then $\phi_{ij} = a$, where $a > 0$, reflecting the fact that the comparison in favour of the treatment. Analogously, if the jth treatment group patient experienced $\Delta M \geqslant c$ and the ith control group patient survived and $\Delta M < c$, then $\phi_{ij} = -a$. Only comparisons involving deaths or those when the pair of patients in the control and treatment groups survive and have M_L measured can lead to non-zero sources. All other comparisons result in zero scores. The proposed test statistic is

$$W_e = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_{ij}.$$

W_e is a U-statistic, and, suitably standardized, has an asymptotic normal distribution.[5] The null hypothesis H_0 is that there is no treatment effect on either the mortal or the morbid component of the endpoint. The alternative hypothesis H_1 is that the treatment has an impact on either the

mortality or on the morbidity component of the endpoint or both. Under the null hypothesis, the expected value of W_e, $E[W_e|H_0] = 0$ so that

$$\frac{W_e}{\sqrt{(\text{var}[W_e|H_0])}} \sim N(0, 1).$$

This two-sided test statistic with significance level α is therefore

$$\text{Reject } H_0 \text{ if } \frac{W_e}{\sqrt{\text{var}(W_e)}} \geqslant Z_{1-\alpha/2}.$$

We find the asymptotic power for this two-tailed test statistic at the α level from Meinert[6] as

$$\begin{aligned} &= P[W_e \geqslant Z_{1-\alpha/2}\sqrt{(\text{var}[W_e|H_0])}|H_1 \text{ true}] \\ &= 1 - \Phi_Z\left[\frac{Z_{1-\alpha/2}\sqrt{(\text{var}[W_e|H_0])} - E[W_e|H_1]}{\sqrt{(\text{var}[W_e|H_1])}}\right] \end{aligned} \tag{1}$$

where $Z_{1-\alpha/2}$ is the $1 - \alpha/2$ percentile value of the standard normal distribution and $\Phi_z(z)$ is the cumulative distribution function of the standard normal distribution evaluated at the point z. The task is to identify the moments of W_e under both the null and alternative hypotheses in such a way that the power of the test statistic can be identified and described in terms of the parameters of the trial's design. Proposals will be made for the probability distributions of $\mathbf{T}(X_i)$, $\mathbf{T}(X_i^+)$, $\mathbf{T}(X_i^*)$, and $\mathbf{T}(X_i^\sim)$ and the corresponding times to events in the treatment group. These distributions can then be used to calculate the required probabilities.

The construction of ϕ_{ij} from experimental data follows from Table I. The variance of the U statistic may be computed as follows. Let $N = 2n$. Then $N^{1/2}W_e$ is asymptotically normal. Under the null hypothesis $E[N^{1/2}W_e] = 0$. Using this condition, we may estimate the variance of $N^{1/2}W_e$ as

$$\hat{\sigma}^2 = \tfrac{1}{2}(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$$

where

$$\hat{\sigma}_1^2 = \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq j'}^{n} \phi_{ij}\phi_{ij'}$$

$$\hat{\sigma}_2^2 = \frac{1}{n^2(n-1)} \sum_{i=1, i \neq i'}^{n} \sum_{j=1}^{n} \phi_{ij}\phi_{i'j}.$$

EXAMPLE: THE SAVE CLINICAL TRIAL

Ventricular enlargement is a major cause of mortality after myocardial infarction, an observation suggesting that measures used to prevent ventricular enlargement may improve the survival of a patient who has sustained a heart attack. The Survival and Ventricular Enlargement (SAVE) trial is a randomized, double-blind, placebo-controlled clinical trial with the purpose of evaluating the effect of angiotensin converting enzyme (ACE) inhibition on postinfarction ventricular dilatation and death. Between 1987 and 1990 this multicentre trial entered 2231 patients who sustained a myocardial infarction within 16 days prior to randomization, were between 21 and 79 years of age, and had an ejection fraction (EF) determined by radionuclide ventriculogram of less than or equal to 40 per cent. Patients were randomized to either ACE inhibitor therapy or

placebo. The study ended in January 1992, with an average treatment and follow-up period of 3·5 years.[7]

The primary endpoint of SAVE is the occurrence of either death or a greater than or equal to 9 unit reduction in ejection fraction. Each patient has a radionuclide ventriculogram during their initial examination (baseline EF). Survivors have another determination at the study's end (terminal EF). The critical magnitude of the change in EF (9 unit or greater reduction) was determined from a review of data describing EF variability.[8] Thus two types of occurrence lead to a primary endpoint event, either patient death during the course of the trial, or for a patient who survives the trial, a reduction of $\Delta EF \geqslant 9$. Note that the observation of a $\Delta EF \geqslant 9$ is an interval censored measurement since the patient may have experienced the deterioration in ejection fraction at any time after randomization but before the terminal EF measurement. Although patients are required to return at the end of the trial for a terminal EF measurement, some patients will not comply. The absence of terminal EF's for some patients can be expected to affect the power of the trial.

The properties of the test statistic for the analysis of the combined endpoint to SAVE should be easily understood as functions of the design parameters of the trial. Examples of these parameters are the control group mortality and the anticipated per cent of control deaths prevented in patients on ACE inhibitor therapy, the per cent of patients in the control group expected to experience a reduction $\Delta EF \geqslant 9$ and the anticipated per cent of EF events prevented in patients on ACE inhibitor therapy in the treatment group.

Modelling the time to death

The hazard function for the time of death is assumed to be constant, resulting in a exponential probability density function for the times until death $\mathbf{T}(X_i)$ and $\mathbf{T}(Y_j)$, with hazard rate λ_1 for the control group, and λ_2 that for the treatment group.

Modelling the time to terminal EF determination

There are four random variables which must be considered in mathematically defining the probability of the event that a follow-up EF is measured and demonstrates a deterioration by nine or more units. Define them as

Q = time until the patient's EF has fallen by nine or more units *given the patient survives the trial*
W = time until the patient's terminal EF has been measured *given the patient survives the trial*
R = time until the patient's terminal EF has been measured and that $\Delta EF \geqslant 9$ *given the patient survives the trial*
S = time until the patient's terminal EF has been measured and that $\Delta EF < 9$ *given the patient survives the trial.*

The actual value of Q can be known only by continuously measuring the patient's ejection fraction, and such data are not available in the trial. We assume Q follows an exponential distribution with parameter α_1 for the control group, and parameter α_2 for the treatment group. The terminal EF's are scheduled during the last months of the trial without regard to clinical status or treatment group assignment. Let T_1 be the time at which terminal EF's can first begin to be measured, for example, T_1 equal two years post-randomization. We assume the time until the EF is obtained is uniformly distributed from time T_1 to time T, the end of the study with probability f. With probability $1 - f$ no terminal EF is obtained and the RVG information is considered

missing on these patients. Then W will have the following mixture density

$$f_W(w) = \frac{f}{T - T_1} I_{T_1 \leq W \leq T} + (1 - f) I_{\text{no EF}}$$

where I denotes the indicator function.

We now identify the probability density functions of $\mathbf{T}(X_i)$, $\mathbf{T}(X_i^+)$, $\mathbf{T}(X_i^*)$, and $\mathbf{T}(X_i^-)$. Beginning with the first of these, note that $\mathbf{T}(X_i)$ is the time to death for a patient in the control group, a time which has been postulated to follow an exponential distribution with parameter λ_1. Then the probability density function of $x = \mathbf{T}(X_i)$ is given by

$$f_X(x) = \lambda_1 e^{-\lambda_1 x}.$$

To identify the distribution of the random variable $\mathbf{T}(X_i^+)$ we note first that the distribution of $\mathbf{T}(X_i^+)$, the time at which the terminal follow-up EF is obtained and $\Delta\text{EF} \geq 9$ given the patient has survived the trial is the probability density of R multiplied by the probability of surviving the trial, that is,

$$f_{T(X_i^+)} = f_R(r) e^{-\lambda_1 T}.$$

R is equal to W whenever $0 \leq Q \leq W \leq T$, where Q is the time the patient's ejection fraction has deteriorated by at least nine units and W is the time the terminal EF is measured. We may write the density of W when $0 \leq Q \leq W \leq T$ as

$$\left[\frac{f}{T - T_1} I_{T_1 \leq W \leq T}\right] \int_0^w \alpha_1 e^{-\alpha_1 u} du.$$

Therefore, the density of R may therefore be written as

$$f_R(r) = \frac{f}{T - T_1} (1 - e^{-\alpha_1 r}) I_{T_1 \leq r \leq T}.$$

Note that this density is conditional on the patient surviving until the trial's end. The density of $\mathbf{T}(X_i^+)$ is the joint density of the random variable R (time until the terminal EF given the patient has survived until the end of the trial) and the event that the patient survives until the end of the trial (required for $\mathbf{T}(X_i^+)$ to occur) $= f_R(r) \exp(-\lambda_1 T)$.

Analogously, let the random variable S be the time of the terminal EF when $\Delta\text{EF} < 9$. Then $S = W$ when $0 \leq W \leq \min(Q, T)$. The probability distribution of $\mathbf{T}(X_i^*)$ is then the product of the conditional probability distribution of S (that is, conditioned on survival) multiplied by the probability the patient survives until the end of the trial. The computations for $\mathbf{T}(X_i^*)$ follow as

$$f_S(s) = \left[\frac{f}{T - T_1} I_{T_1 \leq S \leq T}\right] \int_s^x \alpha_1 e^{-\alpha_1 u} du.$$

The density of S may therefore be written as

$$f_S(s) = \frac{f}{T - T_1} e^{-\alpha_1 s} I_{T_1 \leq S \leq T}$$

Note that this density is also conditional on the patient surviving until the trial's end. The density of $\mathbf{T}(X_i^*)$ is the joint density of the conditional random variable S (conditioned on survival) and the event that the patient survives until the end of the trial (required for $\mathbf{T}(X_i^*)$ to occur) $= f_S(s) \exp(-\lambda_1 T)$. The computation for $\mathbf{T}(X_i^-)$ proceeds analogously.

The power of the test statistic W_e is a function of the mean and variance of W_e under the null and alternative hypotheses. Appendixes 1 and 2 demonstrate the probabilities which involve the time to the events X_i, X_i^+, X_i^*, X_i^- and the time to events Y_i, Y_j^+, Y_j^*, Y_j^-. From the distributions derived in this section the mean and variance of ϕ_{ij} may be obtained and hence the mean and variance of W_e.

In order to link this model to the actual parameters under which the SAVE clinical trial was designed, the parameters of this model must be calibrated to those of SAVE. Let n = total number of patients in each of the therapy groups, d = cumulative death rate in the control group, g = the RVG endpoint event rate ($\Delta EF \geqslant 9$) for survivors in the control group, e_m = efficacy of captopril in preventing mortality ($0 \leqslant e_m \leqslant 1$), and e_g = efficacy of captopril in preventing the occurrence of the RVG endpoint in survivors ($0 \leqslant e_r \leqslant 1$). In SAVE, $n = 1115$, $d = 0{\cdot}2$, $g = 0{\cdot}09$, $e_m = 0{\cdot}16$, and $e_g = 0{\cdot}40$ over the 3·5 year follow-up duration of the trial.

Power computations

We wished to examine the relationship of the power of our statistic to two of the design parameters in SAVE. To compute the power, the variances of ϕ_{ij} must be calculated under both the null and alternative hypotheses and are shown to be functions of joint probabilities simultaneously involving the times to three events. The expressions are given in Appendix II. These results, when utilized in equation (1), allow for the power of the trial to be computed as a function of the design parameters λ_1, λ_2, α_1, and α_2. Figure 1 examines the relationship between the power of the W_e statistic as a function of the weighting parameter a. Increasing the weight parameter increases the power of the W_e statistic. A weight of zero denotes no EF component of the endpoint, and the power of the trial is 51 per cent. However, as the parameter a increases, the statistic incorporates an increasing portion of information from the EF component. Since it was anticipated that the treatment would prevent EF deterioration, adding the EF component to the statistic increases its power.

Figure 2 demonstrates the relationship between the power of the W_e statistic and the terminal EF compliance at the end of the study. Patients return for the terminal EF according to rate f, which is allowed to increase to 80 per cent. If the endpoint of the trial were mortality alone, the power of the trial designed to only examine mortality would be low. Note increasing compliance with the terminal EF is associated with increased power of the W_e statistic. Thus, Figure 2 demonstrates an increase from this lower bound of 51 per cent in the power which is achieved by explicitly modelling the quantitative, morbid component of the combined endpoint. The power of the study increases from this lower bound as patient adherence with the terminal EF increases.

DISCUSSION

The goal of clinical trials is to measure the impact of the therapeutic intervention. To that end, endpoints are chosen carefully, and the most sensitive statistical tools are utilized to analyse the endpoint data from these trials. The Gehan statistic[5] was among the first statistical procedures developed which allowed the incorporation of time to event in the data analysis by comparing the survival times of each patient across therapy groups and ranking these times. Since the development of the Gehan statistic, alternative powerful statistical tools of greater power have been discovered, for example, the logrank statistic. More recently, combined endpoints have also become useful tools to clinical trial designers. Careful choice of the individual components of the combined endpoints allows the development of a well focused clinical trial that requires a smaller number of patients to satisfactorily address the main hypothesis.

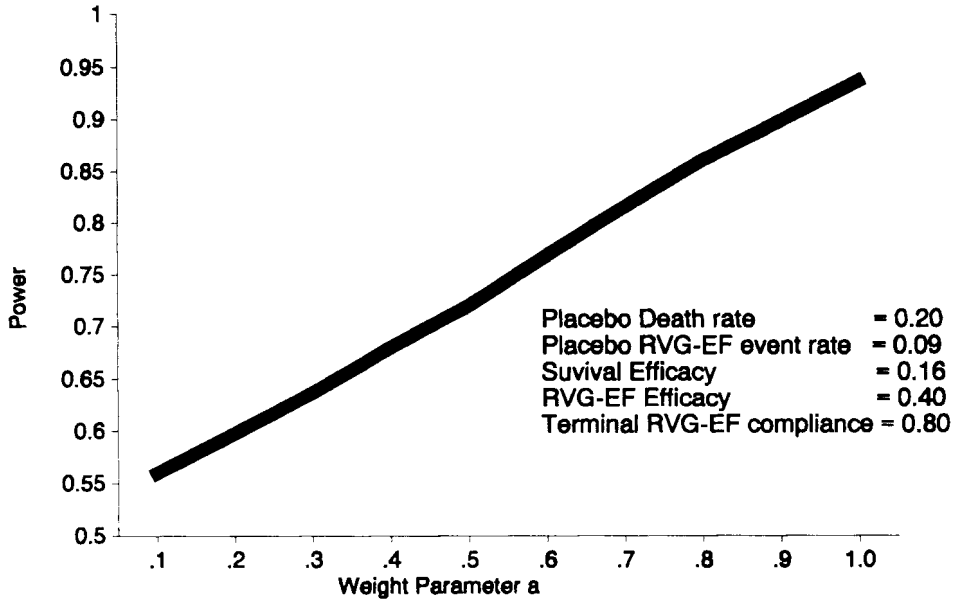

Figure 1. Power of the W_c statistic in SAVE as a function of weight parameter a

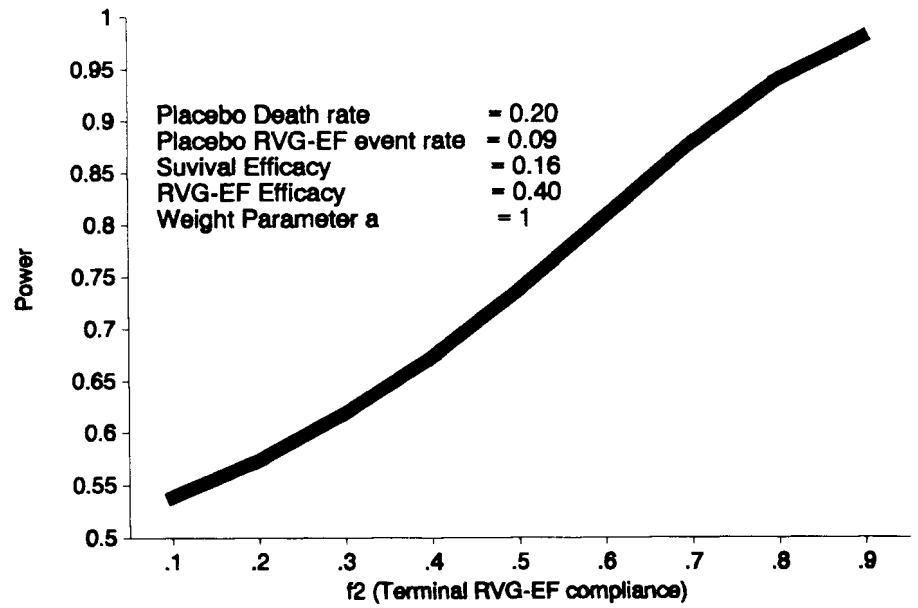

Figure 2. Power of the W_c statistic in SAVE as a function of terminal RVG-EF compliance

The inclusion of the EF determination in the combined endpoint led to an increase in the power of the SAVE trial. However, EF's measured by radionuclide ventriculograms are invasive, expensive procedures not without risk to the patient. They cannot be obtained on a regularly scheduled basis and thus the precise timing of when the patient's ejection fraction has changed is

not available. The exact time to event for this quantitative component of the endpoint is not known because after the baseline measurement, the EF is measured only once in the majority of survivors during the course of the 3·5 years of follow-up of the trial. However, the exact date of death is known.

It is important to determine an ordering of the possible outcomes a patient may have in such a trial. Death was considered the most serious outcome. EF information was used only for patients surviving the trial. This ordering is different from the use of a multivariate survival vector, where information on multiple endpoints are evaluated with no priority given based on the nature of those events.

Using this ordering allowed us to generalize the approach of Gehan in the construction of a statistic to incorporate the limited information available concerning the EF component of the combined endpoint. However, a new scoring function was required for the test statistic so that deaths and changes in EF could be combined in a clinically suitable way. This restricted the number of possible ways to assign a score for ϕ_{ij}. However, it also allowed a general ranking of the times until events for all participants in SAVE.

As defined, the test statistic W_e treats patients who do not return for the terminal EF as missing with respect to the EF component of the endpoint. This seems appropriate for statistics that use ordering information. The distribution theory requires the assumption that the censoring variables have the same distribution within each treatment group, another assumption which appears to be appropriate for a randomized clinical trial when the entry patterns are the same in the two groups.

The probability framework required for the expected value and variance of the test statistic distinguishes between the actual time the patient's ejection fraction deteriorates and the time the terminal EF is obtained. Assuming that the time when the EF is reduced by the critical amount follows an exponential distribution allows the incorporation of cumulative rates of EF changes anticipated by the study designers. However, the timing of the measurement of the terminal EF is also a random variable, and in this case, follows a uniform distribution. The timing of the terminal EF is independent of the therapy group assignment.

There has been an important focus in SAVE on measuring as many terminal EF's as possible. However, in a small number of patients, the terminal EF cannot be obtained. Several of these patients with missing EF information at the end of the study have had an EF measured during the follow-up period of the study due to the development of severe congestive heart failure. One possible extension of the W_e statistic would be to substitute the information from this determination of EF for patients who have no terminal EF. The inclusion of this EF information may further increase the power of the W_e statistic. However, important limitations surround this proposed extension. The incorporation of this EF information into the modelling for the test statistic is problematic since it is not obtained on all patients, and the timing of its occurrence depends on the patient's clinical circumstance and is different from the timing of the terminal EF which is scheduled at the end of the trial for all surviving participants. In addition, the use of this determination in patients who do not have a terminal EF holds the potential for bias in the study. Patients who do not have a terminal EF may be more inclined to poorer health than those who can return for the terminal EF. If these patients who do not return for a terminal EF also have an earlier EF these earlier EF's will have a different probability of demonstrating a deterioration in EF. These important issues must be more completely analysed before any such extension of W_e can be undertaken.

A generalization of this problem is to consider a design where the EF is obtained at regularly spaced intervals over the follow-up period of the trial on all patients. Since the EF is a continuous measure, it is possible that it can increase as well as decrease, causing important changes in the

ΔEF when measured from baseline and complicating the interpretation of the serial ΔEF obtained over time. Thus, one might circumvent this difficulty by choosing as the endpoint the time until the smallest EF occurs, but additional work is required on the consequences of this endpoint definition.

The use of the parameter a in the construction of the scoring function allows the contribution of a score based solely on the endpoint comparisons among survivors to be different from that based on mortality comparisons. For the example chosen in this manuscript, $a = 1$, assigning equal weight to all scores generated in the construction of W_e. However, allowing the statistic to use one weight for the mortality component of the endpoint and another weight for the morbid component affords the trial designer some flexibility in determining the extent of influence each of these two components has on the behaviour of the W_e statistic.

The performance of W_e is evaluated readily because it is explicitly modelled under the null and alternative hypothesis and is therefore a function of the design parameters of the trial. This evaluation demonstrates the relationship between the power of the trial, the survival efficacy and EF adherence. The application of this statistic to SAVE demonstrates the importance of adherence and the minimum adherence that will result in acceptable power.

APPENDIX I: IDENTIFYING THE MEAN OF THE SCORE FUNCTION

The groundwork has been laid in terms of an event space for the endpoint events of interest in SAVE. The purpose of identifying these probability density functions has been to ultimately use the occurrence of events involving the joint outcomes of more than one patient to compute the mean and variance of the scoring function ϕ_{ij} and to assemble the mean value of ϕ_{ij}. The expected value of the score function under either of the null or alternative hypotheses is identified by computing the relevant probabilities of Table I and finding

$$\begin{aligned} E[\phi_{ij}] = {} & P[\mathbf{T}(X_i) < \mathbf{T}(Y_j)] + P[\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j^+)] + P[\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j^*)] + P[\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j^-)] \\ & - P[\mathbf{T}(X_i) > \mathbf{T}(Y_j)] - P[\mathbf{T}(X_i^+) \geqslant \mathbf{T}(Y_j)] - P[\mathbf{T}(X_i^*) \geqslant \mathbf{T}(Y_j)] - P[\mathbf{T}(X_i^-) \geqslant \mathbf{T}(Y_j)] \\ & + aP[\mathbf{T}(X_i^+) \leqslant \mathbf{T}(Y_j^*)] - aP[\mathbf{T}(X_i^*) \geqslant \mathbf{T}(Y_j^+)]. \end{aligned}$$

A summary of how the scoring function is constructed as in Table I. An example of one of the computations for this expectation follows.

Computation of $P[\mathbf{T}(X_i) < \mathbf{T}(Y_j)]$

This is the joint probability of the following events:

(i) the ith patient in the control group dies in $[0, T]$;
(ii) the jth patient in the active group dies in $[0, T]$;
(iii) the time to death of the patient in the control group is less than the time to death in the treatment group.

Let $x = \mathbf{T}(X_i)$ and $y = \mathbf{T}(Y_j)$. This probability becomes

$$P[\mathbf{T}(X_i) \leqslant \mathbf{T}(Y_j)] = \int_0^T \left[\int_0^y \lambda_1 e^{-\lambda_1 x} dx \right] \lambda_2 e^{-\lambda_2 y} dy = 1 - e^{-\lambda_2 T} - \frac{\lambda_2}{\lambda_1 + \lambda_2} (1 - e^{-(\lambda_1 + \lambda_2) T}).$$

The remaining computations for the $E[\phi_{ij}]$ proceed analogously.

APPENDIX II: COMPUTATION OF THE VARIANCE

It remains to examine the var(W_e) under both the null and alternative hypothesis. Notice that

$$\text{var}(W_e) = E[W_e - E[W_e]]^2 = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij} - E(W_e)\right]^2$$

$$= \frac{1}{n^4}E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}\right)^2 - \frac{1}{n^4}E^2[\phi_{ij}].$$

Since $E[\phi_{ij}]$ has been identified in Appendix I we may concentrate here on the evaluation of the double sum. It has been demonstrated by Gehan[5] that

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{j\neq j'}\phi_{ij}\phi_{ij'}$$

$$+ \sum_{i=1,\, i\neq i'}^{n}\sum_{j=1}^{n}\phi_{ij}\phi_{i'j} + \sum_{i=1,\, i\neq i'}^{n}\sum_{j=1}^{n}\sum_{j\neq j'}\phi_{ij}\phi_{i'j'}$$

and we now work to identify the expectation of each of these four double sums. The expectation of the first term on the right is

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{ij}^2\right] = n^2E[\phi_{ij}^2]$$

as there are n^2 terms all with the same expectation. The last term is seen to be

$$E\left[\sum_{i=1,\, i\neq i'}^{n}\sum_{j=1,\, j\neq j'}^{n}\phi_{ij}\phi_{i'j'}\right] = \sum_{i=1,\, i\neq i'}^{n}\sum_{j=1,\, j\neq j'}^{n}E[\phi_{ij}]E[\phi_{i'j'}].$$

Under H_0, the $E[\phi_{ij}] = 0$ so this last term makes no contribution to the variance of W under the null hypothesis. However, under the alternative hypothesis each of the $n^2(n-1)^2$ terms does make a non-zero contribution and must be considered in the expression for the var($W_e|H_1$).

The contribution of the cross product expressions with the terms $\phi_{ij}\phi_{ij'}$ and $\phi_{ij}\phi_{i'j}$ must now be examined. Since the construction of W_e allows an interchange of the roles of i and j, it follows that

$$E\left[\sum_{i=1}^{n}\sum_{j=1,\, j\neq j'}^{n}\phi_{ij}\phi_{ij'}\right] = E\left[\sum_{i=1,\, i\neq i'}^{n}\sum_{j=1}^{n}\phi_{ij}\phi_{i'j}\right] = n^2(n-1)E[\phi_{ij}\phi_{ij'}]$$

and writing the var(W_e) as a function of the ϕ_{ij}

$$\text{var}[W_e|H_0] = \frac{1}{n^4}(n^2E[\phi_{ij}^2] + 2n^2(n-1)E[\phi_{ij}\phi_{ij'}])$$

$$\text{var}[W_e|H_1] = \frac{1}{n^4}(n^2E[\phi_{ij}^2] + 2n^2(n-1)E[\phi_{ij}\phi_{ij'}] + n^2(n-1)^2E[\phi_{ij}] - n^4E^2[\phi_{ij}]).$$

The terms involving $E[\phi_{ij}]$ do not appear in the expression for the variance under the null hypothesis since under this assumption they are equal to zero. In addition, since our interest lies in asymptotic results, we may ignore terms on the order of $o(n^{-2})$ or less. Thus the variance under H_0 reduces to

$$\text{var}\left[\frac{1}{n^2}W_e|H_0\right] \sim \frac{2}{n}E[\phi_{ij}\phi_{ij'}]. \tag{3}$$

Table II. Values and events involved in the construction of $\phi_{ij}\phi_{ij'}$

Values of ϕ_{ij} and corresponding events		Values of $\phi_{ij'}$ and the corresponding events									
		$\phi_{ij'} = 1$				$\phi_{ij'} = a$	$\phi_{ij'} = -a$	$\phi_{ij'} = -1$			
		$T(X_i) < T(Y_{j'})$	$T(X_i) \leq T(Y_{j'})^+$	$T(X_i) \leq T(Y_{j'}^*)$	$T(X_i) \leq T(Y_{j'}^-)$	$T(X_i^+) \leq T(Y_{j'}^*)$	$T(X_i^*) \geq T(Y_{j'}^+)$	$T(X_i) > T(Y_{j'})$	$T(X_i^+) \geq T(Y_{j'})$	$T(X_i^*) \geq T(Y_{j'})$	$T(X_i^-) \geq T(Y_{j'})$
$\phi_{ij} = 1$	$T(X_i) < T(Y_j)$	1	1	1	1			-1			
	$T(X_i) \leq T(Y_j^+)$	1	1	1	1			-1			
	$T(X_i) \leq T(Y_j^*)$	1	1	1	1			-1			
	$T(X_i) \leq T(Y_j^-)$	1	1	1	1			-1			
$\phi_{ij} = a$	$T(X_i^+) \leq T(Y_j^*)$					a^2			$-a$		
$\phi_{ij} = -a$	$T(X_i^*) \geq T(Y_j^+)$						a^2			a	
$\phi_{ij} = -1$	$T(X_i) > T(Y_j)$	-1	-1	-1	-1			1			
	$T(X_i^-) \geq T(Y_j)$					$-a$			1		
	$T(X_i^*) \geq T(Y_j)$						a			1	
	$T(X_i^-) \geq T(Y_j)$										1

The variance expression under the alternative hypothesis does not simplify so easily since it must contain terms involving $E[\phi_{ij}]$. Note

$$n^2(n-1)^2 E^2[\phi_{ij}] - n^4 E^2[\phi_{ij}] = -2n^3 E^2[\phi_{ij}] + n^2 E^2[\phi_{ij}].$$

The asymptotic result for the variance of the test statistic under the alternative hypothesis now becomes

$$\text{var}[W_e|H_1] \sim \frac{2}{n}[E[\phi_{ij}\phi_{ij'}] - E^2[\phi_{ij}]]. \tag{4}$$

It now remains to examine the expressions involving $\phi_{ij}\phi_{ij'}$. We note that the $E[\phi_{ij}]$ is readily found under the null hypothesis to be zero and under the alternative hypothesis by the results of Appendix II. However, the evaluation of the cross product term $E[\phi_{ij}\phi_{ij'}]$ also can be identified in a straightforward manner. Consider the different values the product $\phi_{ij}\phi_{ij'}$ can have, remembering that i indexes only one patient in the control group, and j and j' index two different patients in the treatment group. Such an evaluation leads to the comparisons presented in Table II. The shaded regions in the table are combinations of events that are impossible and may therefore be excluded from consideration in the construction of $\phi_{ij}\phi_{ij'}$. We see then that the $E[\phi_{ij}\phi_{ij'}]$ can be constructed by assembling the appropriate probabilities.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Dr. L. J. Wei, who provided the initial idea for this project, and Dr. James Ware, whose suggestions and guidance were instrumental in the development of this project. The authors also wish to thank the reviewers and editor of this manuscript for their helpful suggestions which led to its improved organization. The authors also thank Ms. Lynda R. Crouch for her help in the final preparation of this manuscript.

REFERENCES

1. Mantel, N. 'Evaluation of survival data and two new rank order statistics arising in its consideration', *Cancer Chemotherapy Reports*, **50**, 163–170 (1966).
2. Cox, D. R. 'Regression models and life tables (with discussion)', *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
3. Lipid Research Clinics Program. 'The Lipid Research Clinics Coronary Primary Prevention Trial Results. 1. Reduction in incidence of coronary heart disease', *Journal of the American Medical Association.* **251**, 351–364 (1984).
4. Echt, D. S., Leibson, P. R., Mitchell, L. B., Peters, R. W., Obaias-Manno, D., Barber, A. H., Arensberg, D., Baker, A., Friedman, L., Greene, H. L., Huther, M. L. and Richardson, D. W. 'Mortality and morbidity in patients receiving encainide, flecainide, or control. The Cardiac Arrthymia Suppression Trial', *New England Journal of Medicine*, **324**, 781–788 (1991).
5. Gehan, E. A. 'A generalized Wilcoxon test for comparing arbitrarily singly-censored samples', *Biometrika*, **52**, 203–223 (1965).
6. Meinert, C. L. *Clinical Trials Design, Conduct, and Analysis*, Oxford University Press, Oxford, 1986, pp. 84–85.
7. Moyé, L. A. Pfeffer, M. A., Braunwald, E., for the SAVE Cooperative Group. 'Rationale and design of a trial to assess patient survival and ventricular enlargement after myocardial infarction', *Journal of the American College of Cardiology*, **68**, 70D–79D (1991).
8. Mahamarian, J. L., Moyé, L., Verani, M. S., Easton, T., Francis, M. and Pratt, C. M. 'Criteria for the accurate interpretation of changes in left ventricular ejection fraction and cardiac volumes as assessed by rest and excercise gated radionuclide angiography', *Journal of the American College of Cardiology*, **18**, 112–119 (1991).