Proceedings of the 2005 International Conference on Simulation and Modeling
V. Kachitvichyanukul, U. Purintrapiban, P. Utayopas, eds.

THERAPEUTIC EQUIVALENCE IN SURVIVAL DATA

Karan P. Singh[1] , Kenneth J. Peschell[2] , Sejong Bae[3] ,
Fernando Treviño[4] , Ram Shanmugam[5] and Alfred A. Bartolucci[6]

[1,3,4]School of Public Health
University of North Texas Health Science Center
Fort Worth, Texas 76107, USA

[2,6]Department of Biostatistics
School of Public Health,
University of Alabama at Birmingham
Birmingham, AL 35294-0022, USA

[5]Department of Health Services Research
Southwest Texas State University
San Marcos, TX 78666, USA

ABSTRACT

Active control equivalence trials (ACETs), conducted with the goal of demonstrating therapeutic equivalence, are of growing importance to the pharmaceutical industry, clinical medicine, government, and academia. In this paper, therapeutic equivalence is defined in terms of equivalent clinical outcome (e.g., survival) without regard to assessment of bioequivalence. The likelihood-ratio-based asymptotic fiducial and Bayesian methods for therapeutic equivalence assessment (TEA) are developed in the context of survival analysis, using the Weibull distribution. The methods are illustrated using leukemia remission data.

1 INTRODUCTION

Active control equivalence trials (ACETs) have received much attention in the statistical literature. The reader is referred to Dunnett and Gent (1977), Blackwelder (1982), Patel and Gupta (1984), Hauck and Anderson (1986), Mau (1988); Makuch, Stephens, and Escobar (1989); Durrieman and Simon (1990); Fleming (1990); Dunnett and Gent (1996), Ebbutt and Frith (1998), Robins (1998), Weins and Iglewicz (2000). Therapeutic equivalence refers to a new treatment being as efficacious as a standard treatment. Although the FDA has a very strict definition of therapeutic equivalence (FDA, 1979, p. 2937) for regulatory purposes, therapeutic equivalence may be considered in a less restrictive context emphasizing "equivalent" efficacy in clinical outcome and therapeutic benefit to the patients. It is in this context, for the purpose of individual clinical decision making, that our methodology is most beneficial.

Some examples of the therapeutic equivalence problem include "reduction in dose or duration of chemotherapy, less extensive surgery, or substitution of an invasive technique by an external procedure" (Durrieman and Simon, 1990, p. 329). As a particular example, in surgery, therapeutic equivalence has been investigated between new replacement heart valves versus the standard of practice replacement heart valves (personal communication, Naftel, 1995, for other examples, Makuch and Johnson, 1989, for examples in AIDS research, Cooper, 1990). Therapeutic equivalence is important because a new treatment that is therapeutically equivalent to an existing one can be superior in other contexts. For instance, the new treatment may have fewer side effects or otherwise improve quality of life, may be easier to administer or have a preferred dosage schedule, or may be less expensive to produce. Also, a generic drug company may wish to demonstrate therapeutic equivalence to the FDA so that the company may introduce a competing product.

Therapeutic equivalence and bioequivalence are two different concepts, despite its controversy. Foremost, the endpoints in therapeutic equivalence and bioequivalence are very different; the endpoint in the former is some measure of clinical outcome, the endpoint in the latter is typically related to blood/plasma levels (personal communication, Makuch, 1995). Also, assessment of therapeutic equivalence generally requires conduct of a clinical trial(s), whereas bioequivalence is often assessed in a much smaller setting, e.g., bioassays. Durrleman and Simon (1990) write, "compared to the biological problem of bioavailability, the therapeutic equivalence of two treatments is a more pragmatic concept. Rather than assessing a theoretical equivalence, which makes little sense when the two treatments can be very different "in principio" (e.g., surgery

versus lithotripsy), one is interested in therapeutic decision making." Indeed, in the view of the FDA, bioequivalence is a prerequisite for therapeutic equivalence, clearly establishing a difference between the two concepts.

As a particular example, in the context of breast cancer research, in estrogen-receptive women, consider two drugs, A and B, each for prevention of spread of the cancer. It is hypothesized that spread of the cancer follows from either (1) increased peptide growth factor (GF) synthesis and subsequent binding of GFs with specific membrane-bound receptors which in turn signal the initiation of a sequence of intracellular actions resulting in cell division, or (2) binding of estrogen to those limited number of cells containing intracellular estrogen receptors; those cells, through a sequence of steps, secrete some of the GFs which then signal the initiation of (1) in nearby cells which otherwise are unresponsive to estrogen. Drug A could be targeted to block the binding of the GFs to their membrane receptors or to block the post-receptor signal transduction. Drug B could be targeted to block the binding of the estrogen to its receptor or to block an event in the estrogen post-receptor signal transduction mechanism. Success with both Drug A and Drug B would prevent spread of the cancer; therefore the drugs would yield equivalent clinical outcome. However, the active ingredients in the drugs could certainly be different, thereby making assessment of bioequivalence not feasible, despite the therapeutic equivalence. Other examples where bioequivalence is not a relevant concern may be found in hypertension research.

There are a number of classical (non-Bayesian) methods for TEA. Dunnett and Gent (1977) present a method of significance testing to compare two binomial samples with data summarized in 2x2 tables. Blackwelder (1982) presents a method of hypothesis testing with a dichotomous outcome variable and sample sizes large enough for use of the normal approximation to the binomial. Patel and Gupta (1984) present a method of hypothesis testing with a normally distributed response variable. Hauck and Anderson (1986) use a confidence interval approach. Mau (1988) presents a method of Cox's "confidence distributions" using the normal approximation to the binomial. Fleming (1990) presents a method for time to event data, using Cox's proportional hazards regression to estimate the hazard ratio (or relative risk of failure) of the two treatments. The method uses confidence intervals for the hazard ratio to assess either superiority of one treatment or equivalence (for application in "non-inferiority" trials). Dunnett and Gent (1996) present a procedure with union-intersection and intersection-union hypothesis testing approaches to test "simultaneously for a positive difference and for equivalence" (Dunnett and Gent, 1996, p. 1729).

It is important to note that each of these methods pertains to either normally distributed means or binomially distributed proportions, with the exception of Fleming's (1990) method. Recall that his method is based on Cox regression, and therefore it is semiparametric. Our methodology (as well as the Bayesian methods addressed immediately below) employs parametric modeling in a progressively censored survival analysis setting. Alternatively, some Bayesian methods are available for TEA. Bartolucci and Singh (1993) and Singh (1996) present a method that is similar to the confidence interval approach above, instead using a Bayesian posterior credibility interval for the ratio of functions of parameters of survival distributions. Their method is developed using the translated exponential in the role of the survival distribution, along with inverted gamma and vague priors.

There are also numerous (1) classical and (2) Bayesian approaches to bioequivalence assessment. (1) classical: Westlake (1972, 1974, 1976, 1979, 1981), Metzler (1974), Kirkwood (1981), Hauck and Anderson (1984), Rocke, (1984), Schuirmann (1987), Chow and Shao (1990), Liu and Chow (1993), (2) Bayesian: Rodda and Davis (1980), Mandallaz and Mau (1981), Selwyn, Dempster, and Hall (1981), Flühier, Grieve, Mandallaz, Mau, and Moser (1983), Selwyn and Hall (1984), Selwyn, Hall, and Dempster (1985), Racine-Poon, Grieve, Flühier, and Smith (1987), Liu and Chow (1997), some of which are precursors of some of the above-mentioned methods for therapeutic equivalence.

Indeed, some of Westlake's discussions infer therapeutic equivalence from bioequivalence, and such inferences have been questioned and debated (Kirkwood, 1981; Westlake, 1981). Recall also that therapeutic equivalence can also be considered in a context emphasizing "equivalent" efficacy in clinical outcome and therapeutic benefit to the subjects, where bioequivalence is not a relevant concern. We wish to emphasize a focus on clinical outcome, survival patterns in particular, therefore we consider TEA based on clinical outcome to be independent of (or, in addition to) any assessments of bioequivalence. Additionally, all of the above-mentioned methods for bioequivalence assessment are based on either means of a normal distribution (the data are usually log transformed first – usually base 10, sometimes base e) or on binomial proportions; i.e., none of the methods make use of survival distributions.

2 METHODOLOGY

We present a likelihood ratio based asymptotic fiducial method for therapeutic equivalence assessment (TEA) for progressively censored survival data. In this paper, we develop the method for the Weibull distributed data. Our method requires an elicitation from a clinical expert to establish the maximum clinically insignificant difference between two treatments. Unlike some other methods, the method does not require a priori specification of the standard and experimental treatments. Our method provides a considerable amount of information for individual clinical

decision making, as opposed to merely a hypothesis testing of "reject" or "do not reject" type, or a p-value. The numerical value of our method is interpretable for individual clinical decision making. Also, for the Weibull distributed data, the method is easily implemented in SAS.

The method requires data structure of an ordered pair for each of two treatment groups, (x_{iA}, c_{iA}) for i = 1 to n_A and (x_{iB}, c_{iB}) for i = 1 to n_B; note that n_A need not equal n_B. Each x_i is either a time of death (t_i ; survival time; uncensored observation) or a time of loss-to-follow-up (t_I^+ ; a progressively censored observation). Each c_i is either a 0 if the observation is uncensored or a 1 if the observation is censored. Therefore, the number of deaths for either treatment group, r_k, is given by

$$r_k = n_k - \sum_{i=1}^{n_k} c_{ik}, \qquad k = A, B \tag{1}$$

The two survival distributions are considered (clinically) equivalent if the distance between them is less than some clinically pre-specified value (note that this discussion is actually relevant for practical consideration). A question then arises: what is a suitable measure of the distance? In the setting of likelihood ratio test of

$$H_0: f(t;\ \psi_1) = f(t;\ \psi_2) \text{ versus}$$
$$H_a: f(t;\ \psi_1) \neq f(t;\ \psi_2), \tag{2}$$

where ψ, ψ_1 and ψ_2 are parameter vectors, the likelihood ratio may be transformed into a measure of the distance

$$D = 2[L(\psi_1, \psi_2) - L(\psi)] \tag{3}$$

Although D is an unknown with respect to the sample space, it is a random variable in parameter space. We now apply the fiducial arguments: $F_1(\Delta|D) = F(D|\Delta)$ (Fisher, 1973; Quenouille, 1958). Δ represents a distance measure between two distributions. $F_1(\Delta|D)$ is the cumulative fiducial probability function and $F(D|\Delta)$ is the cumulative distribution function of the sufficient statistic D given Δ. The distribution $f(D|\Delta)$ is not easy but an asymptotic result provides an alternate approach.

Under H_0, D has an asymptotic chi-squared distribution with v degrees of freedom equal to the reduction from the total number of parameters in ψ_1 and ψ_2 to the number of parameters in ψ. Under Ha, D has an asymptotic non-central chi-squared distribution,

$f(\chi^2; v, \lambda)$ with v degrees of freedom and non-centrality parameter

$$\lambda = (\theta_r - \theta_{r0})' V_{r0}^{-1} (\theta_r - \theta_{r0}) \tag{4}$$

where θ_r and θ_{ro} correspond to a reparameterization of the hypotheses as

$$H_0: \theta_r = \theta_{ro} \text{ versus } \quad Ha: \theta_r \neq \theta_{ro} \tag{5}$$

and where V_{r0}^{-1} is the inverse of the dispersion matrix, with elements

$$\left(V_{ij}^{-1}\right) = -E\left(\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right) = E\left(\frac{\partial \ln L}{\partial\theta_i}\frac{\partial \ln L}{\partial\theta_j}\right) \tag{6}$$

(Kendall and Stuart, 1973). Therefore, although $f(D|\Delta)$ is not easily obtained, $f(D|\lambda)$ is easily obtained, as $f(\chi^2; v, \lambda)$. We therefore consider the fiducial distribution of λ given D, $f_f(\lambda|D)$, employing the fiducial argument $F_f(\lambda|D) = F(D|\lambda)$ (Fisher, 1973; Quenouille, 1958), estimating $F(D|\lambda)$ by $F(D|\hat{\lambda})$, now subsequent to the demonstration that $\hat{\lambda}$ is a sufficient statistic. The question arises: Is λ as a suitable measure of the distance between the two distributions as is Δ?

Both Δ and λ are random variables ranging from 0 to ∞. If $\Delta=0$, H_0 is true, as that is the only possible way for the numerator and denominator of the likelihood ratio to be equal. If $\lambda=0$, then H_0 is true, by the definition of λ. As each Δ and λ increases, evidence builds in support of Ha; therefore each increases as the distributions differ. The relationship between Δ and λ is implicit, but monotonic. Therefore, asymptotically, λ is as suitable a measure of the distance as is Δ.

To perform the TEA, information must be elicited from a clinical expert in order to establish the maximum clinically insignificant difference between the two distributions. This information, which we will base on median survival times, must then be used to determine a clinically specified value, λ_{expert}. Once λ_{expert} has been established, one simply calculates $P_f(\lambda \leq \lambda_{expert}|D) = F_f(\lambda_{expert}|D) \approx F(\lambda_{expert}|\hat{\lambda})$ to obtain the fiducial probability that the two distributions are as close as λ_{expert} or closer. Note that the same value λ_{expert} is used for both the fiducial and Bayesian methods. $\hat{\lambda}$ is asymptotically sufficient. Notably, there is no need to set a significance level a priori; in fact, there is considerable flexibility for individual clinical decision making at this point.

Remark: As pointed out by the referee, the proposed approach can be translated into a confidence interval approach. The non-centrality parameter, λ, is a transformation of the unknown parameter, θ_r, . In fact, it is the Mahalanobis distance of θ_r from the value specified by the null hypothesis. By using standard asymptotic arguments one

could derive a confidence interval for λ, for example, through the delta method or the likelihood ratio test. Confidence interval could then be compared with λ_{expert} . An alternate way is to build, instead of a confidence interval, an upper limit for λ that can be computed with λ_{expert} .

3 WEIBULL DISTRIBUTION DATA

The likelihood function of *r* survival times and *n* - *r* times of loss-to-follow-up is

$$e^{L(\beta,\gamma)} = \frac{\gamma^r}{\beta^{r\gamma}}\left(\prod_{i=1}^{r} t_i^{\gamma-1}\right)e^{-\frac{1}{\beta^\gamma}\sum_{i=1}^{n} o_i^\gamma} \tag{7}$$

where t_i denotes a survival time until death; uncensored time, t_i^+ denotes a censored time, $o_i = t_i$ for uncensored observations and $o_i = t_i^+$ for censored observations. In terms of testing

$$H_0: f(t; \beta_A, \gamma_A) = f(t; \beta_A, \gamma_A), \text{ versus}$$
$$H_a: f(t; \beta_A, \gamma_A) \neq f(t; \beta_B, \gamma_B) \tag{8}$$

the estimated likelihood ratio is given by

$$\Lambda = \frac{e^{L(\hat{\beta},\hat{\gamma})}}{e^{L(\hat{\beta}_A,\hat{\gamma}_A,\hat{\beta}_B,\hat{\gamma}_B)}},\ 0 < \Lambda < 1 \tag{9}$$

The MLEs $\hat{\beta}$ and $\hat{\gamma}$ must be obtained by an iterative procedure. Estimates of the asymptotic variances of $\hat{\beta}$ and $\hat{\gamma}$ and of the asymptotic covariances of $\hat{\beta}$ and $\hat{\gamma}$ are, respectively, by

$$\hat{Var}(\hat{\beta}) = \frac{1}{-\left.\frac{\partial^2 L(\beta,\gamma)}{\partial\beta^2}\right|_{\beta=\hat{\beta}}} = \frac{r\hat{\gamma}^2}{\hat{\beta}^2} \tag{10}$$

$$\hat{Var}(\hat{\gamma}) = \frac{1}{-\left.\frac{\partial^2 L(\beta,\gamma)}{\partial\gamma^2}\right|_{\gamma=\hat{\gamma}}}$$
$$= r\hat{\gamma}^{-2} + \hat{\beta}^{-\hat{\gamma}}\sum_{i=1}^{n} o_i^{\hat{\gamma}}\left(\ln o_i - \ln\hat{\beta}\right)^2 \tag{11}$$

$$\hat{Cov}(\hat{\beta},\hat{\gamma}) = \frac{1}{-\left.\frac{\partial^2 L(\beta,\gamma)}{\partial\beta\partial\gamma}\right|_{\beta=\hat{\beta},\gamma=\hat{\gamma}}}$$
$$= -\hat{\gamma}\hat{\beta}^{-(\hat{\gamma}+1)}\sum_{i=1}^{n} o_i^{\hat{\gamma}}\left(\ln o_i - \ln\hat{\beta}\right) \tag{12}$$

Determination of the Non-centrality Parameter ($\hat{\lambda}$) and an Asymptotic Fiducial Distribution for TEA

We reparameterize the hypotheses as follows:

$$H_0: \begin{bmatrix}\beta_A - \beta_B\\ \gamma_A - \gamma_B\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}, \text{ versus}$$
$$H_a: \begin{bmatrix}\beta_A - \beta_B\\ \gamma_A - \gamma_B\end{bmatrix} \neq \begin{bmatrix}0\\0\end{bmatrix} \tag{13}$$

The non-centrality parameter is given by

$$\lambda = (\theta_r - \theta_{r0})' V_{r0}^{-1}(\theta_r - \theta_{r0}) \tag{14}$$

whose estimate is given as follows:

$$\hat{\lambda} = \begin{bmatrix}\hat{\beta}_A - \hat{\beta}_B\\ \hat{\gamma}_A - \hat{\gamma}_B\end{bmatrix}' \hat{V}_{r0}^{-1}\begin{bmatrix}\hat{\beta}_A - \hat{\beta}_B\\ \hat{\gamma}_A - \hat{\gamma}_B\end{bmatrix} \tag{15}$$

As the two groups receive different treatments, the two populations are independent, and $Var(\beta_A - \beta_B) = Var(\beta_A) + Var(\beta_B)$, $Var(\gamma_A - \gamma_B) = Var(\gamma_A) + Var(\gamma_B)$, and $Cov(\beta_A - \beta_B, \gamma_A - \gamma_B) = Cov(\beta_A, \gamma_A) + Cov(\beta_B, \gamma_B)$ are used. We estimate the individual information components by the inverses of their separately determined variance estimates. We therefore estimate $\hat{V}_{r0}^{-1}$ by

$$\hat{V}_{r0}^{-1} = \begin{bmatrix}\left\{\hat{Var}\left(\hat{\beta}_A - \hat{\beta}_B\right)\right\}^{-1} & \left\{\hat{Cov}\left(\hat{\beta}_A - \hat{\beta}_B, \hat{\gamma}_A - \hat{\gamma}_B\right)\right\}^{-1}\\ \left\{\hat{Cov}\left(\hat{\beta}_A - \hat{\beta}_B, \hat{\gamma}_A - \hat{\gamma}_B\right)\right\}^{-1} & \left\{\hat{Var}\left(\hat{\gamma}_A - \hat{\gamma}_B\right)\right\}^{-1}\end{bmatrix} \tag{16}$$

and then obtain $\hat{\lambda}$ as

$$\hat{\lambda} = \begin{bmatrix}\hat{\beta}_A - \hat{\beta}_B\\ \hat{\gamma}_A - \hat{\gamma}_B\end{bmatrix}'$$
$$\begin{bmatrix}\left\{\hat{Var}\left(\hat{\beta}_A - \hat{\beta}_B\right)\right\}^{-1} & \left\{\hat{Cov}\left(\hat{\beta}_A - \hat{\beta}_B, \hat{\gamma}_A - \hat{\gamma}_B\right)\right\}^{-1}\\ \left\{\hat{Cov}\left(\hat{\beta}_A - \hat{\beta}_B, \hat{\gamma}_A - \hat{\gamma}_B\right)\right\}^{-1} & \left\{\hat{Var}\left(\hat{\gamma}_A - \hat{\gamma}_B\right)\right\}^{-1}\end{bmatrix}$$
$$\begin{bmatrix}\hat{\beta}_A - \hat{\beta}_B\\ \hat{\gamma}_A - \hat{\gamma}_B\end{bmatrix} \tag{17}$$

Having obtained $\hat{\lambda}$, an asymptotic fiducial distribution, $f_f(\lambda|D)$, may now be defined as discussed in Section 2.0.

Determination of λ_{expert}

We propose the required elicitation from an expert to be simply an estimate of the maximum clinically insignificant difference in median survival times for the two treatment groups; we shall call the estimate dsme. We first compute $\hat{\beta}_{\text{mid}}$, the midpoint of $\hat{\beta}_{\text{A}}$ and $\hat{\beta}_{\text{B}}$, and $\hat{\gamma}_{\text{mid}}$, the midpoint of $\hat{\gamma}_{\text{A}}$ and $\hat{\gamma}_{\text{B}}$. Simultaneously, we "spread" the $\hat{\beta}$s and $\hat{\gamma}$s, around their respective midpoints, decrementing and incrementing by 1% of each respective midpoint. We search until the implied median difference reaches d_{sme} (the median of the Weibull distribution is given by $\hat{\text{med}} = \hat{\beta} \times 0.69314718^{1/\hat{\gamma}}$), obtaining $\hat{\beta}_{\text{Asme}}$, $\hat{\beta}_{\text{Bsme}}$, $\hat{\gamma}_{\text{Asme}}$, and $\hat{\gamma}_{\text{Bsme}}$. We compute the implied value of λ, naming it λ_{expert}, as

$$\lambda_{\text{expert}} = \begin{bmatrix} \hat{\beta}_{Asme} - \hat{\beta}_{Bsme} \\ \hat{\gamma}_{Asme} - \hat{\gamma}_{Bsme} \end{bmatrix}' \hat{\mathbf{V}}_{\mathbf{rsme}}^{-1} \begin{bmatrix} \hat{\beta}_{Asme} - \hat{\beta}_{Bsme} \\ \hat{\gamma}_{Asme} - \hat{\gamma}_{Bsme} \end{bmatrix} \tag{18}$$

We estimate $\mathbf{V}_{\mathbf{rsme}}^{-1}$ by

$$\hat{\mathbf{V}}_{\mathbf{rsme}}^{-1} = \begin{bmatrix} V_{rsme11}^{-1} & V_{rsme21}^{-1} \\ V_{rsme21}^{-1} & V_{rsme22}^{-1} \end{bmatrix}, \tag{19}$$

with components given by

$$V_{rsme11}^{-1} = \left\{ \hat{V}ar\left(\hat{\beta}_A - \hat{\beta}_B \right) \right\}^{-1}, \tag{20}$$

$$= \left(\frac{r_{Asme} \hat{\gamma}_{Asme}^2}{\hat{\beta}_{Asme}^2} + \frac{r_{Bsme} \hat{\gamma}_{Bsme}^2}{\hat{\beta}_{Bsme}^2} \right)^{-1}$$

$$V_{rsme12}^{-1} = \left\{ \hat{C}ov\left(\hat{\beta}_A - \hat{\beta}_B, \hat{\gamma}_A - \hat{\gamma}_B \right) \right\}^{-1},$$

$$= \begin{Bmatrix} -\hat{\gamma}_{Asme} \hat{\beta}_{Asme}^{-(\hat{\gamma}_{Asme}+1)} \sum_{i=1}^{n} o_i^{\hat{\gamma}} \left(\ln o_i - \ln \hat{\beta} \right) \\ + -\hat{\gamma}_{Bsme} \hat{\beta}_{Bsme}^{-(\hat{\gamma}_{Bsme}+1)} \sum_{i=1}^{n} o_i^{\hat{\gamma}} \left(\ln o_i - \ln \hat{\beta} \right) \end{Bmatrix}^{-1}, \tag{21}$$

$$V_{rsme21}^{-1} = V_{rsme12}^{-1}, \tag{22}$$

and

$$V_{rsme22}^{-1} = \left\{ \hat{V}ar\left(\hat{\gamma}_{Asme} - \hat{\gamma}_{Bsme} \right) \right\}^{-1},$$

$$= \begin{Bmatrix} r_{Asme} \hat{\gamma}_{Asme}^{-2} + \hat{\beta}_{Asme}^{-\hat{\gamma}_{Asme}} \sum_{i=1}^{n} o_i^{\hat{\gamma}} \left(\ln o_i - \ln \hat{\beta} \right)^2 \\ + r_{Bsme} \hat{\gamma}_{Bsme}^{-2} + \hat{\beta}_{Bsme}^{-\hat{\gamma}_{Bsme}} \sum_{i=1}^{n} o_i^{\hat{\gamma}} \left(\ln o_i - \ln \hat{\beta} \right)^2 \end{Bmatrix}^{-1} \tag{23}$$

Note that a disadvantage of the estimation of $V_{rsme12}^{-1} = V_{rsme21}^{-1}$ and V_{rsme22}^{-1} is that the sums that they contain are functions of the raw data and the original MLEs; unfortunately, these approximations are arguably among the best available choices. Also, note that although the subject matter expert only provides information to estimate implied medians with a difference of d_{sme}, we are actually able to estimate the parameters of the distribution, resulting in the ability to make inferences regarding therapeutic equivalence of the distributions, and not merely therapeutic equivalence of the medians.

4 ACUTE MYELOGENOUS LEUKEMIA DATA

A phase III trial was performed to compare remission induction in two groups of acute myelogenous leukemia patients (Vogler *et al.*, 1992). The goal of the trial was to demonstrate that the experimental treatment, the anthracycline idarubicin (IDR) in combination with cytarabine (CA), was superior, in remission induction, to a standard treatment, the anthracycline daunorubicin (DNR) in combination with CA. The trial demonstrated the hypothesized superiority of IDR in remission induction. The only inference made with respect to survival was that no statistically significant difference existed between the two treatment groups, assessed by the log-rank and generalized Wilcoxon rank sum tests. In this section, we demonstrate the likelihood ratio based asymptotic fiducial method for TEA by applying it to this leukemia data, with the goal of demonstrating therapeutically equivalent survival in the two treatment groups. We begin with description of the data structure as well as preliminary descriptive analyses of the data.

Two hundred thirty patients were randomized, 111 to IDR, and 119 to DNR. We use 109 and 115 patients, respectively (the exclusions are for incorrect diagnosis, randomized but not treated, or death prior to treatment). There were 104 deaths in the IDR group and 103 deaths in the DNR group, implying 5 and 12 censored observations, re-

spectively. The data are measured in months. To examine the Weibull distributions to model the survival patterns, we perform the likelihood ratio test comparing, one Weibull population with two Weibull populations.

The maximum likelihood estimates (MLEs) of the Weibull distribution for the IDR group and the DNR group are $\hat{\beta}_e$ = 16.371 and $\hat{\gamma}_e$ = 0.910, and $\hat{\beta}_s$ = 14.595 and $\hat{\gamma}_s$ = 0.814, respectively. The MLEs for the two groups combined are $\hat{\beta}_c$ = 15.447 and $\hat{\gamma}_c$ = 0.858. For one Weibull population versus two Weibull populations the likelihood ratio test has a p-value of *0.552*, failing to demonstrate evidence for two populations. This result is the desired one as a preliminary for TEA, and we may later examine clinically equivalent survival using the Weibull distribution. We examine goodness of fit using the method of Hollander and Proschan (1979) and Lee (1992, p. 191). It is noteworthy that this method does lack power, as the test is against a universal alternative. However, the strength of this test is its ability to handle progressive censored data. For IDR and DNR, the test has p-values of 0.8734 and 0.8695, respectively, yielding conclusions that the Weibull distribution is not inappropriate.

Now, we consider Weibull TEA. The maximum clinically insignificant difference in median survival times for the two treatment groups, dsme, has been specified by a subject matter expert to be 3 months. The fiducial probability of therapeutic equivalence with dsme = 3 months is 0.949. The first consideration for a patient is whether or not an approximately 95% "chance" of "equivalent" survival is acceptable. The subject matter expert has also specified 3 < dsme < 6 to be a clinical "gray area". Therefore, further flexibility for individual clinical decision making is derived by examining the fiducial probability through this "gray area". The fiducial probabilities of therapeutic equivalence with dsme 3.5, 4, 4.5, 5, and 5.5= months are 0.98981, 0.99864, 0.99988, 0.99999, and 0.99999, respectively. For instance, if an individual is willing to consider a difference of 4 months in median survival times to be clinically insignificant, then the "chance" of "equivalent" survival is better than 99%. All of this information, in conjunction with all other information regarding the treatments (e.g., quality of life), may be used by an individual patient with his or her physician to make a treatment decision.

The authors of this manuscript have illustrated a very nice application of fiducial approach in therapeutic studies. This contribution is significant in dealing with the complex issue of understanding non-observed parameter with information in observed data.

ACKNOWLEDGEMENTS

The authors very much appreciate the referee's well thoughtful and constructive comments and suggestions in strengthening the paper.

REFERENCES

Bartolucci, A.A., and Singh, K.P. 1993. Therapeutic Equivalence: A Bayesian approach Proceedings of the Biopharmaceutical Section of the American Statistical Association, 349-357.

Blackwelder, W.C. 1982. "Proving the Null Hypothesis" in Clinical Trials. Controlled Clinical Trials 3: 345-353.

Chow, S.-C., and Shao, J. 1990. An Alternative Approach for the Assessment of Bioequivalence between Two Formulations of a Drug. Biometrical Journal 32: 969-976.

Cooper, E. C. 1990. Active Control Equivalence Trials. Journal of Acquired Immune Deficiency Syndromes 3(Suppl. 2): S77-S81.

Dunnett, C.W. and Gent, M. 1977. Significance Testing to Establish Equivalence Between Treatments, With Special Reference to Data in the Form of 2 x 2 Tables. Biometrics 33: 593-602.

Dunnett, C.W. and Gent, M. 1996. An Alternative to the Use of Two-Sided Tests in Clinical Trials. Statistics in Medicine 15: 1729-1738.

Durrelman, S and Simon, R 1990 Planning and monitoring of equivalence studies. Biometrics 46: 329-336

Ebbutt, A.F. and Frith, L. 1998. Practical issues in equivalence trials. Statistics in Medicine 17 (15-16): 1691-1701.

Fisher, R.A. 1973. Statistical Methods and Scientific Inference, 3rd edition. Hafner Press, New York.

Fleming, T.R. 1990. Evaluation of Active Control Trials in AIDS. Journal of Acquired Immune Deficiency Syndromes 3(Suppl. 2): S82-S87.

Flühier, H., Grieve, A.P., Mandallaz, D., Mau, J., and Moser, H.A. 1983. Bayesian Approach to Bioequivalence Assessment: An Example. Journal of Pharmaceutical Sciences 72: 1178-118 1.

FDA Food and Drug Administration. 1979. Therapeutically Equivalent Drugs. Availability of List. Federal Register 44: 2932-2953.

Hauck, W.W. and Anderson, S. 1984. A New Statistical Procedure for Testing Equivalence in Two-Group Comparative Bioavailability Trials. Journal of Pharmacokinetics and Biopharmaceutics 12: 83-91.

Hauck, W.W. and Anderson, S. 1986. A Proposal for Interpreting and Reporting Negative Studies. Statistics in Medicine 5: 203-209.

Hollander, M. and Proschan, F. 1979. Testing to determine the underlying distribution using randomly censored data. Biometrics 35, 393-401.

Kendall, M.G., and Stuart, A. 1973. The Advanced Theory of Statistics, Volume 2, Inference and Relationships, 3rd edition. Hafner, New York.

Kirkwood, T.B.L. 1981. Bioequivalence Testing - A Need to Rethink. Biometrics 37: 589-591.

Lee, E.T. 1992. Statistical Methods for Survival Data Analysis, 2nd edition. Wiley, New York.

Liu, J. and Chow, S.C. 1997. A two one-sided tests procedure for assessment of individual bioequivalence. Journal of the biopharmaceutical in statistics 7 (1): 49-76.

Liu, J.P. and Chow, S.C. 1993. On Assessment of Bioequivalence for Drugs with Negligible Plasma Levels. Biometrical Journal 35: 109-123.

Makuch, R. 1995. Personal Communication, New Haven, Connecticut.

Makuch, R., and Johnson, M. 1989. Issues in Planning and Interpreting Active Control Equivalence Studies. Journal of Clinical Epidemiology 42: 503-511.

Makuch, R.W., Stephens, M.A., and Escobar, M. 1989. Generalized Binomial Models to Examine the Historical Control Assumption in Active Control Equivalence Studies. The Statistician 38: 61-70.

Mandallaz, D., and Mau, J. 1981. Comparison of Different Methods for Decision-Making in Bioequivalence Assessment. Biometrics 37: 213-222.

Mau, J. 1988. A Statistical Assessment of Clinical Equivalence. Statistics in Medicine 7: 1267-1277.

Metzler, C.M. 1974. Bioavailability - A Problem in Equivalence. Biometrics 30: 309-317.

Naftel, D. C. 1995. Personal Communication, Birmingham, Alabama.

Patel, H. I. and Gupta, G.D. 1984. A Problem of Equivalence in Clinical Trials. Biometrical Journal 26: 417-474.

Quenouille, M.H. 1958. Fundamentals of Statistical Reasoning. Hafner, New York.

Racine-Poon, A., Grieve, A.P., Flühier, H., and Smith, A.F.M. 1987. A Two-Stage Procedure for Bioequivalence Studies. Biometrics 43: 847-856.

Racine-Poon A, Grieve AP, Flühier H, Smith A.F.M. Bayesian methods in practice: experience in the pharmaceutica industry (with discussion). Applied Statistics 1986;35:93 150.

Robins, J.M. 1998. Correction for non-compliance in equivalence trials. Statistics in Medicine 17 (3): 269-302.

Rocke, D.M. 1984. On Testing for Bioequivalence. Biometrics 40: 225-230.

Rodda, B.E., and Davis, R.L. 1980. Determining the Probability of an Important Difference in Bioavailability. Clin. Pharmacol. Ther 28: 247-252.

Schuirmann, D.J. 1987. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. Journal of Pharmacokinetics and Biopharmaceutics 15: 657-680.

Selwyn, M. R., Dempster, A. P., and Hall, N. R. 1981. A Bayesian Approach to Bioequivalence for the 2 x 2 Changeover Design. Biometrics 37: 11-21.

Selwyn, M.R., and Hall, N.R. 1984. On Bayesian Methods for Bioequivalence. Biometrics 40: 1103-1108.

Selwyn, M.R., Hall, N.R., and Dempster, A.P. 1985. Comment on Testing for Bioequivalence. Biometrics 41: 561.

Singh, K.P. 1996. Therapeutic Equivalence Using a Rich Family of Prior Distributions, in Lifetime Data: Models in Reliability and Survival Analysis, edited by Jewell, N.P., Kimber, A.C., Lee, M.-L. T., and Whitmore, G.A.. Kiuwer Academic Publishers, Boston.

Vogler WR, Velez-Garcia E, Weiner RS, et al.: A phase III trial comparing idarubicin and daunorubicin in combination with cytarabine in acute myelogenous leukemia: a Southeastern Cancer Study Group. Journal of Clinical Oncology 10(7): 1103-1111, 1992.

Wald, A. 1943. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. Trans. Amer. Math. Soc. 54: 426.

Weins, B.L. and Iglewicz, B. 2000. Design and analysis of three treatment equivalence trials. Control Clinical Trials 21 (2): 127-137.

Westlake, W.J. 1972. Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials. Journal of Pharmaceutical Sciences 61: 1340-1341.

Westlake, W.J. 1974. The Use of Balanced Incomplete Block Designs in Comparative Bioavailability Trials. Biometrics 30: 319-327.

Westlake, W.J. 1976. Symmetric Confidence Intervals for Bioequivalence Trials. Biometrics 32: 741-744.

Westlake, W.J. 1979. Statistical Aspects of Comparative Bioavailability Trials. Biometrics 35: 273-280.

Westlake, W.J. 1981. “Response to Bioequivalence Testing-A Need to Rethink.” Biometrics 37: 591-593.

AUTHOR BIOGRAPHIES

KARAN P. SINGH is Professor and Chairman, Department of Biostatistics, School of Public Health, University of North Texas Health Science Center at Fort Worth, Texas, USA. He is an elected member of International Statistical Institute (ISI) and a fellow of the Modeling and Simulation Society of Australia and New Zealand (MSSANZ). His email address is <ksingh@hsc.unt.edu>

KENNETH J. PESCHELL is a former advisee of Professors Alfred A. Bartolucci and Karan P. Singh. Presently he is working for a pharmaceutical company.

SEJONG BAE is an Assistant Professor, Department of Biostatistics, School of Public Health, University of North Texas Health Science Center at Fort Worth, Texas, USA. His email address is <sbae@hsc.unt.edu>

FERNANDO TREVINO is Professor and Dean, School of Public Health, University of North Texas Health Science Center at Fort Worth, Texas, USA. He is an elected member of prestigious scientific societies. His email address is <ftrevino@hsc.unt.edu>

RAM SHANMUGAN is a Professor , Department of Health Services Research, Southwest Texas State University, Texas, USA. He is an elected member of prestigious scientific societies. His email address is <ram.shanmugan@swt.edu>

ALFED A. BARTOLUCCI is Professor, School of Public Health, University of Alabama at Birmingham, AL, USA. He is an elected member of prestigious scientific societies. His email address is <ftrevino@hsc.unt.edu>