INCOMPLETE MULTIRESPONSE DESIGNS AND SURROGATE
ENDPOINTS IN CLINICAL TRIALS

by

Pranab K. Sen

Department of Biostatistics, University of
North Carolina at Chapel Hill, NC.

INCOMPLETE MULTIRESPONSE DESIGNS AND SURROGATE
ENDPOINTS IN CLINICAL TRIALS

Pranab K. Sen

Departments of Biostatistics & Statistics

University of North Carolina

Chapel Hill, NC 27599-7400

Abstract.   Incorporating plausible surrogate variates, a general class of incomplete multiresponse designs for clinical trials is considered. In this context, the basic differences between the conventional linear models and survival analysis models are discussed thoroughly. Along with some parametric and semi-parametric procedures, suitable nonparametric ones are considered, and the related asymptotic theory is presented.

AMS Subject Classification: 62P10, 62K10, 62G99

Key words and phrases: Bootstrap; concomitant variates; conditional quantile, hierarchical design; incomplete multi-response design; linear model; logistic regression; nonparametrics; parametrics; semi-parametrics; validation sample.

Running Head: Imd & Surrogate Endpoints

1. Introduction

In clinical trials, from epidemiological perspectives, often, information is gathered on more than one response variable, and, in addition, on relevant concomitant variables too. Nevertheless, from clinical or medical perspectives, it is not uncommon to single out one of the response variates as the most important (or relevant) one, termed the primary variate (although, in some cases, there may be stringent cost constraints associated with its precise measurement). The multitude of other response variates through their mutual dependence can cast light on the primary variate too, and hence, their simultaneous measurements generally lead to comparatively more precise statistical conclusions. On the other hand, based on other practical considerations, such as relative cost, ease and/or precision of measurement, it may not be very convenient to include the entire battery of simultaneous measurements on all the response variates (as well as the covariates). For this reason, often, an incomplete multiresponse design (**imd**) or a hierarchical design (**hd**) is adopted. In the latter design, there is an hierarchy on the response variates in the sense that there is a (partial) ordering with respect to the number of experimental units on which their measurements are recorded. For example, suppose that there are $p(\geq 1)$ response variates, denoted by $Y_1,..., Y_p$, respectively. On a smallest set, say S_1, of experimental units, all these p responses are measured simultaneously; there is a larger set, say S_2, such that $S_1 \subset S_2$, and on the subset $S_2 \setminus S_1$, $Y_2,..., Y_p$ are recorded (not Y_1), and so on. For the largest set S_p, such that $S_1 \subset S_2 \subset ... \subset S_p$, on $S_p \setminus S_{p-1}$, only Y_p is recorded; $S_{p-1} \setminus S_{p-2}$, Y_{p-1}, Y_p are recorded and so on. Such a multiresponse design, determined by $\{S_p,..., S_p\}$ along with the inherent nesting: $S_1 \subset S_2 \subset ... \subset S_p$, is termed a **hierarchical design** [viz., Roy, Gnanadesikan and Srivastava (1971, Ch. 8)]. It may not be always desirable (or practical to impose the basic hierarchy condition mentioned above [vis., missing observation (response) problem in multiresponse data]. Nevertheless, it may be feasible to incorporate some incomplete response designs wherein the set $Y = \{Y_1,..., Y_p\}$ can be partitioned into various subsets $\{Y_{i_1},..., Y_{i_r}\}$, $1 \leq r \leq p$, $1 \leq i_1 < ... < i_r \leq p$, such that these subsets are adopted for possibly different number of experimental units. For example, for $p=2$, we have three possible subsets $\{Y_1\}$, $\{Y_2\}$ and $\{Y_1, Y_2\}$, and possibly different designs (say, D_1, D_2 and D_{12}) may be chosen for these subsets. Recall that in clinical trials one may be primarily interested in a comparative study of a placebo and one or more treatments, so that the designs $\{D\}$ are chosen in the conventional sense with due emphasis on these treatments.

In clinical trials, a primary (response) variate, in spite of being the most relevant one, may encounter some basic problems regarding its precise measurement (due to excessive cost or other limitations), and, therefore, it is not uncommon to make use of a very closely related (but, presumably,

relatively less expensive) variable (say Y_o), which is termed a *surrogate variate*. Generally, surrogate variates do not contain as much information as contained in the primary variate (say, Y_1), but their use may greatly enhance the scope of the clinical trial (subjected to some cost and time constraints). The use of such surrogate variates (response variables) in clinical trials (or medical investigations, in general) has generally been accepted by the allied medical community and has caught the attention of statisticians as well. A nice account of such uses (and abuses too) from a statistical perspective is given in a set of articles published in Statistics in Medicine, vol. 8, No. 2 (1989). For some more recent statistical work, we also refer to Pepe (1992), where other references are cited.

The primary objective of the current study is to focus on the design and analysis aspects of a general class of incomplete multiresponse designs, adoptable in clinical trials, which incorporate an appropriate surrogate variate to enhance statistical information without escalating (much) the cost of experimentation. The conventional surrogate model (incorporating a validation sample) considered by Pepe (1992) can be characterized as a special case of an imd wherein one has only one primary variate and a surrogate variate, and moreover, only a completely randomized design (treatment wise) is envisaged in her work. In Section 2, we formulate an imd incorporating a surrogate variate in a more general setup, and study various aspects of such a design.

In spite of the isomorphism of design aspects, there is a basic difference between a conventional linear model and a so called regression model in clinical trials. This is apparent from the respective formulation of statistical functions on which inference is to be drawn in a valid and yet reasonably efficient manner. This may generally limit the scope of applicability of standard *parametric* procedures in such imd with surrogate variates. In that respect, *semi-parametric* models fare well, and in Section 3, we provide an outline of both such procedures. Our main concern is to develop appropriate *nonparametric* procedures which may not need the basic (and somewhat stringent) regularity conditions inherent in a parallel semi-parametric model. We propose to defer most of these technicalities to a subsequent communication. However, for the sake of illustration, in Section 4, we consider a simple imd (with one primary variate and a surrogate variate), and discuss the related nonparametrics in some detail.

2. Surrogate response variate and imd in clinical trials.

In many medical investigations or clinical trials, primary variates are difficult or expensive to obtain, and hence, it may be necessary (or more economic, at least) to use surrogate variates instead of the primary one(s). However, such a substitution may have serious effects on valid and efficient

statistical modeling and analysis, unless the surrogate variate is closely related to the primary one (in a statistical sense). The situation may particularly be very bleak when this statistical relation between the primary and surrogate variates may not be known, a situation that may arise when no specific data is available on simultaneous measurement of both these variates. For this reason, it seems very reasonable (if not essential) to design a study in such a way that for a majority of the experimental units only a surrogate variate is taken into account (along with relevant concomitant variates) while a smaller subset of experimental units, both the primary and surrogate variates are recorded (to throw light on their statistical relation). We may as well term this smaller subset of the experimental units as the *validation sample*, if this subset is chosen in a random manner (through an objective sampling scheme). However, such a randomness criterion may not be universally true. We shall see later on that this choice of a valid subset (on which simultaneous recordings are made) may have a good impact on the design as well as analysis aspects of such clinical trials. In the absence of a validation sample, we are left only with data on the surrogate variate as well as the concomitant variates (and design variables), and hence, any conclusion on the primary variate we may want to draw from the data set would demand rather stringent conditions on the surrogate variate. In the context of regression models in a clinical trial setup, Prentice (1989) has discussed the problem and laid down suitable regularity conditions on the surrogate variate which permit a valid statistical analysis. The situation changes drastically if a validation sample becomes available, and we may refer to Pepe (1992) for some nice work. The Prentice-Pepe setups can be characterized as both a hierarchical and incomplete multiresponse model with p=2 (where Y_1 refers to the primary variate and Y_2 to the surrogate one). We denote by $S=S_2$ the entire set of experimental units, and as in Section 1, we consider a subset $S_1(\subset S)$ where both the primary and surrogate variates are to be measured, while on the complementary set $S_2 \setminus S_1$, only the surrogate variates is to be recorded. For each experimental unit, information on the design and concomitant ones variables is gathered in the usual way. This design also fits with the imd, described in Section 1, where p=2 and $\{Y_1\}$ has a null set of experimental units. For a set A, let #A stand for its cardinality. Then, ideally, we may like to take

$$\# S_1 \ / \ \# S_2 \quad \text{as large as possible.} \tag{2.1}$$

However, this may create some operational impasse (from the clinical point of view), and hence, often we may have to settle (2.1) as a small number (say, c > 0), the extent of which may be determined by

the largeness of # S_2. Basically, # S_1 should be adequately large to enable one to study the statistical relationship between Y_1 and Y_2 in a satisfactory manner.

Many clinical trials encounter a more complicated study which have multiple endpoints resulting in multiresponse primary variates. We may refer to Wei, Lin and Weissfeld (1989) and Prentice and Cai (1992) for some nice statistical account of such problems. Here also, often, it may be possible to order these primary response variates on the basis of their importance as well as relevance. Thus, we may conceive of a vector

$$Y = (Y_1,..., Y_p)' \qquad (2.2)$$

of response variates, where Y_1 is regarded as the primary response variate, and the other coordinates (of Y) have decreasing order of importance. In such a case also, precise measurement of the primary variate Y_1 (on all the experimental units) may be difficult or costly, and hence, whenever a surrogate variate (say, Y_o) is available, it would be of good statistical importance to incorporate it in the design and statistical analysis. The other primary variate (i.e., $Y_2,..., Y_p$) may also share a similar difficulty, and hence, surrogate variates for them, whenever available, may also be taken into account in such designing of the study. Thus, in a comparatively more general setup, we may conceive of a q-vector, Y_0, of surrogate variates ($q \geq 1$), such that we have a total number of p+q response variates, some of which may be difficult to be recorded. Again, statistical relationship between Y and Y_o may not be adequately known unless they are simultaneously recorded for different subsets of experimental units. This provides the genesis of incomplete multiresponse designs in clinical trials. If all the surrogate variates can be conveniently measured for each experimental unit, while there is a natural ordering (of the relative ease of recording) of the p primary variates (with Y_1 being the most difficult one), we may conceive of the following hierarchical design:

$$S = S_o \supset S_p \supseteq \cdots \supseteq S_1, \qquad (2.3)$$

where on S_1, all the (p+q) responses are recorded, on $S_2 \setminus S_1$, (p+q − 1) responses (i.e., $Y_2,..., Y_p$ and Y_o), and so on; on $S_p \setminus S_{p-1}$, Y_p and Y_o are recorded, and on $S_o \setminus S_p$, only the surrogate variate Y_o is recorded. It may also be possible to induce a hierarchy on Y_o and partition $S_o \setminus S_p$ into appropriate (nested) sets to implement this hierarchy. In this setup, p and q need not be the same, and it may even be desirable to choose $q \geq$ p, so that additional information on the predictability of Y through Y_o can be extracted from this larger set of surrogate variates. Depending on the nesting in (2.3) [or some other similar hierarchical schemes], for the specific subsets,

$$S_1, \ S_2 \setminus S_1,..., \ S_p \setminus S_{p-1}, \ S_o \setminus S_p \ , \tag{2.4}$$

designs (primarily to draw statistical conclusions in an efficient manner) $\{D\}$ may depend very much on the nature of the response variates as well as concomitant variates pertaining to these subsets. As in Roy, Gnanadesikan and Srivastava (1971, Ch. 8), we may allow these designs to be of quite general form. However, such designs are to be set primarily on the basis of clinical factors, and as such, may be quite different from the conventional agricultural ones (where a comparatively more controlled setup can be taken for granted). There is a further fundamental difference between the two setups, and this will be discussed in detail in the next section.

The hierarchical scheme in (2.3) - (2.4) is based on an ordering of the primary variates (i.e., the coordinates of $\mathbf{Y}$ are important in their own way (viz, they may stand for some endpoints which may be all of serious concern). In such a situation, some of the basic appeals of a (response wise) hierarchical design may be lost, and we may need to incorporate alternative designs which allow some flexibility in this respect. If we assume that all the $q(\geq 1)$ surrogate response variates are relatively easy to be recorded for the experimental units, there seems to be no good reason to drop them. Thus, response wise incomplete blocking may be solely applied to $\mathbf{Y}$. As in Section 1, we denote by $\mathbf{P} = \{1,$ ..., p} and consider the totality of subsets of $\mathbf{P}$, defined by $i_r = \{i_1,.., i_r\}$, for all possible $1 \leq i_1 < ... < i_r \leq p$ and $r = 0, 1,..., p; \ i_o = 0$. Thus, we would have a total number 2^p subsets of $\mathbf{P}$. Determined mostly by clinical and other extraneous considerations, we consider a proper subset $\mathbf{P}_0$ of $\mathbf{P}$, defined by

$$\mathbf{P}_0 = \{i_r : \ (r, i_r) \ \epsilon \ I_o\}. \tag{2.5}$$

Then the set $S = S_o$ of totality of experimental units is partitioned into a system

$$S(\mathbf{P}_0) = \{S_{i_r} : (r, i_r) \ \epsilon \ \mathbf{P}_0\} \ , \tag{2.6}$$

where for the subset S_{i_r}, the primary response variate(s) $Y_{i_2},..., Y_{i_r}$ along with the surrogate variate Y_o are recorded (and, in addition, appropriate design and concomitant variates are also recorded). For the subset S_{i_r}, an appropriate design D_{i_r} may be incorporated (for the treatment as well as design variates), so that we have the dual design sets

$$D(\mathbf{P}_0) = \{D_{i_r} : (r, i_r) \ \epsilon \ \mathbf{P}_0\} \ . \tag{2.7}$$

Then an incomplete multiresponse design for a clinical trial with the provision for surrogate end points can be formulated in terms of the pair

$$\{ \ S \ (\mathbf{P}_0) \ , \quad \mathbf{D} \ (\mathbf{P}_0) \ \}, \tag{2.8}$$

where the first set refers to the response wise design set, while the other one for the treatment wise design set. Obviously, we shall assume that the cardinality of $\mathbf{P}_0$ is greater than one, and also, we shall exclude the null set i_o from $\mathbf{P}_0$, unless we intend to have a set of experimental units where only the surrogate response variates (and covariates) are to be recorded. The design considered by Prentice (1989) is a special case of $p = 1$, $q = 1$; $\mathbf{P}_0$ refers to the set of no primary index and $S \ (\mathbf{P}_0) = S_o$ is the entire set of experimental units where only the surrogate variate is recorded. Pepe (1992) considered the case of $p = 1$, $q = 1$ and $i_o = \{0\}$, $i_1 = \{1\}$, so that one has a subset containing only the surrogate response, while a second subset contains both the primary and surrogate variates (*validation sample*).

Thus, in our formulation, we have a general class of imd's $\{ \ S(\mathbf{P}_0), \ D(\mathbf{P}_0)\}$, which may also be adoptable in a more general multiresponse setup (incorporating surrogate variates). In this manner, our findings extend the basic results of Roy, Gnanedesikan and Srivastava (1971, Ch. 9) to clinical trials (having possibly multiple surrogate endpoints). We may reiterate that unlike the case in conventional agricultural experiments, the choice of the so called optimal (or nearly optimal) designs of the type (2.8) may depend heavily on the clinical aspects (relating to the cost and/or difficulty of obtaining measurements on the response variates), and hence, a "cost-benefit" approach may as well be incorporated in this setup. This is a potential area of active research for biostatisticians as well as clinical scientist, and it is rather an immature opportunity to go into this realm in the current study.

We conclude this section with a further observation that when there are probably more than one surrogate endpoint, it may not be feasible (or practical) to have them recorded for all experimental units. In such a case, we may need to replace the set $\mathbf{P}$ by

$$\mathbf{P}^* = \{ \ i,..., \ p \ ; \quad 1,..., \ q \ \} \ , \tag{2.9}$$

and i_r by $(i_r, i_s) = \{ \ i_1,..., \ i_r \ ; \ i_1^*,..., \ i_s^* \ \}$, for all possible $1 \ \leq \ i_1 < ... < i_r \leq p$, $1 \leq i_1^* < ... < i_s^* \leq q$; $1 \leq r \leq p$, $1 \leq s \leq q$, and also include the null set i_o or i_o^* in this system. We replace $\mathbf{P}_0$ by $\mathbf{P}_0^* = \{ \ (\ i_r, i_s^*): \ (r, s, i_r, i_s^*) \ \epsilon \ I_o^* \ \}$. Then, we extend (2.6) to

$$S(P_0^*) = \{ \ S_{(i_r, \ i_s)}: \ (r, \ s, \ i_r, \ i_s) \ \epsilon \ P_0^* \ \}, \qquad\qquad (2.10)$$

wherein the subset $S_{(i_r, \ i_s)}$ the primary response variate(s) $Y_{i_1}, ..., Y_{i_r}$ and the surrogate response variates $Y^o_{i_1^*}, ..., Y^o_{i_s^*}$ are to be recorded on the experimental units. Although this extension covers a very broad class of clinical designs, we would not envisage a great generality of such designs, as it would lead to highly complex statistical analysis schemes, which would not only demand a much larger number of experimental units ($\#S$), but also will make it much less appealing to the clinical scientists who may not have that great apathy for such sophisticated statistical analysis. This point is made more clear in the next section.

3. Imd, Surrogate endpoints and Statistical analysis

In this section, we place side by side incomplete multi-response designs for conventional (mostly, agricultural) experiments and clinical trials (or medical investigations), and examine their relative complexities in the perspective of useful statistical modeling and efficient analysis. We may refer to Srivastava (1966, 1968) for some general discussion of the main issues involved in imd's, when there is no surrogate endpoint, and one has the conventional *multivariate general linear model (mglm)* setup in mind. The mglm coupled with the traditional assumption of the multinormality of the error vectors enable us to formulate the statistical models in a finite-dimensional parametric setup. We refer to Ch. 9 (Sections 9.2 and 9.3) of Roy, Gnanadesikan and Srivastava (1971) for an excellent account of statistical analysis of general imd's, wherein the multi-normality of the error vectors and the mglm structure both play vital roles. For possible departures from either of these two basic assumptions, even to a small extent, there can be serious effects on the validity and efficiency of their prescribed solutions. Or, in other words, for imd's, the standard parametric statistical analysis is likely to be much more nonrobust compared to the usual complete block (response) designs.

The appropriateness of mglm's in clinical trials is open to questions. First of all, the primary response variates may have highly skewed distributions (marginally) on the positive part of the real line, and they may also sometimes be *quantal* in nature (i.e., death/alive). In the first case, often a logarithm transformation can induce greater amount of symmetry for these (marginal) distributions, but claiming that the corresponding error (transformed) vector has closely a multivariate normal distribution may not be that rational. Moreover, any such transformation on the response variates affect the basic mglm. For example, if all the response variables are subject to logarithm

transformations, a mglm on the transformed variables relates to multiplicative model for the original model -- and this needs to be checked too. For a quantal response, a logistic regression model is often regarded as adequate. But, in the multi-response case, when not all the response variates are quantal in nature, there may be generally quite a bit of complications in adopting a logarithm/other conventional (power) transformations on a part and logistic regression setup for the other part. Secondly, generalized linear models (GLM) [viz., McCullah and Nelder (1989)], as have been developed quite satisfactorily mostly for uniresponse models, needs to be extended to such imd's. However, the main complication can arise due to (i) plausible non-identifiability of such GLM's -- a condition that is inherent in the adaptation of *quasi-likelihood* or *partial likelihood methods*, (ii) possible high dimension of the "asymptotic covariance" matrix, which generally may be quite complex in GLM's, and (iii) inadequately large sample sizes: With the increase in the number of parameters (say, t) in the GLM, the sample size (say, n(t)), need for a satisfactory adaptation of the asymptotic approximations, increases at a rate t^2 i.e., $n(t) >> t^2 n(1)$. Thirdly, in clinical trials, *censoring (type I, II or random)* is a common phenomenon. In simplest models, the celebrated Cox (1972) *partial likelihood* approach can be adapted with advantage. The performance characteristic of this approach depends quite a bit on the extent of censoring, and with moderate to high level of censoring, there may be very little information in the partial likelihood to induce any power of the test. The partial likelihood approach has led to significant developments on semi-parametric models (via a *counting process* modeling too). But, successful adaptation of a partial likelihood approach (often, termed the proportional hazard model or the multiplicative intensity model) may demand certain preconditions which may not be generally tenable in an imd, especially with a large number of covariates. Robustness considerations often recommend against an unrestricted use of such models. Even for the most simple case of p=q=1, Pepe (1992) has shown that a surrogate endpoint may grossly affect the validity and efficiency of statistical analysis when based on such a partial likelihood/"regression" model. We may reiterate at this stage that if the scatter of the covariates is not very localized (as is mostly the case), assuming a multiplicative intensity may cause a serious damage even for a small departure from the assumed model. For this reason, if the sample sizes are adequately large (as may generally be the case with multi-center multi-response (endpoints) clinical trials, it may be more appropriate to choose a model which is as flexible as possible, so that a consistent picture can be drawn from the statistical outcome of the trial. In this sense, a reasonably efficient statistical analysis based on a valid and broad model seems to be more appropriate than a more efficient statistical analysis pertaining to a more restrictive model. Viewed from this point, there seems to be a major interest in developing such "nonparametric" models pertaining to imd's in clinical trials with surrogate endpoints, and a steady growth of research work is anticipated in near future. We illustrate such a model based

analysis in the simplest model of p=q=1 in the next section.

4. Nonparametric inference using surrogate variates

We consider a uniresponse model with a true outcome variate, denoted by Y, and a set of concomitant variates, denoted by Z. Since it may be difficult or quite expensive to obtain measurements on Y, a closely related (but comparatively less expensive) variable, say X, may be used as a surrogate variate. In this setup, for a major part of the experimental units, measurements are recorded only on

$$(X_i , Z_i) , \ i \ \epsilon \ I. \tag{4.1}$$

This set I may be termed a *surrogate set*. This set I fails to provide any information on the relationship between Y and X, which is very much needed in drawing conclusions on the behavior of Y from that of X. For this reason, we may also consider another subset I^* (disjoint from I), such that for $i \ \epsilon \ I^*$, we record measurements on Y, X and Z all; this may be termed a validation sample. We may refer to Pepe (1992) for some nice examples of such designs in clinical trials. In this setup, we have, borrowing the notations of earlier sections,

$$S_o = I \ \bigcup \ I^* \ \text{and} \ S_1 = I^*, \tag{4.2}$$

so that we have a hierarchical design. We may also extend it to an imd by considering a third subset, say I^o, where we have measurements on (Y, Z) (but not X).

Suppose now that F(y | z) stands for the conditional distribution function (d.f.) of Y given Z = z, G(x | z) for the conditional distribution function of X given Z= z, and let H(y, x | z) be the conditional (joint) distribution function of Y and X, given Z= z. In clinical trials, it is customary to consider the survival functions, which are the complementary parts of the distribution functions. In a nonparametric setup, we consider a functional $\theta(z) = \theta(F(\circ | z))$ of the conditional d.f. $F(\circ | z)$ as a parameter of interest. For example, $\theta(z)$ may stand for the median of the conditional d.f. $F(\circ | z)$. Another popular measure is the conditional expectation of Y, given Z = z. However, this is highly non-robust, and we shall not pursue the same. In a similar manner, we define a conditional functional $\xi(z) = \xi(G(\circ | z))$ of the conditional d.f. $G(\circ | z)$, given Z = z. Since the primary variate Y and its surrogate X are generally closely related (and hence, statistically dependent too), it is quite

intuitive to set

$$\theta(z) \equiv \psi (\xi(z)), \quad z \in \mathcal{S}_z; \tag{4.3}$$

where $\psi (\cdot)$ is a suitable functional. In a conventional linear model setup, $\theta(z)$ and $\xi(z)$ are both taken as linear in z, so that $\psi (\cdot)$ may very conveniently be defined as a linear functional too. On the other hand, in clinical trials, under a nonparametric setup, it is not very reasonable to assume that $\theta(z)$ or $\xi(z)$ are themselves linear functions (or log-linear functions), and hence, allowing $\psi (\cdot)$ to be rather nonparametric seems to be a better alternative.

In a hierarchical design, $\xi(z)$ may be estimated from the data set pertaining to I and $\psi(\cdot)$ from the set I^*. In an imd, $\theta(z)$ may be estimated from the set I^o, $\xi(z)$ from I and $\psi(\cdot)$ from I^*. For simplicity of presentation, we consider explicitly the case where $\theta(\cdot)$ and $\xi(\cdot)$ are conditional quantiles, and the theory extends directly to more general functionals. As a surrogate variate is essentially a substitute for the primary variate, following Prentice (1989), it seems reasonable to require a surrogate for some true endpoint to have the potential to yield unambiguous information about differential treatment effects on the true endpoint. Hence, Prentice (1989) defined a surrogate endpoint to be a "response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint." To judge its relevance in the current context, let us partition the covariate vector Z as $(Z^{(1)}, Z^{(2)})$, where $Z^{(1)}$ contains the information on the treatment levels (mostly through dummy (nonstochastic) vectors, while $Z^{(2)}$ contains information on the genuine concomitant variates, some of which may as well be stochastic in nature. In the simplest model of 'placebo vs. treatment', often, $Z^{(2)}$ is taken as null, and $Z^{(1)}$ assumes only the two possible values (0 for placebo, and 1 for treatment). Thus, if $\theta (Z^{(1)}, Z^{(2)})$ and $\xi(Z^{(1)}, Z^{(2)})$ stand for the conditional functionals, the null hypothesis H_0 of interest is given by

$$H_0: \quad \theta(Z^{(1)}, Z^{(2)}) \equiv \theta (0 , Z^{(2)}), \tag{4.4.}$$

for all $Z^{(1)}$ and $Z^{(2)}$. If the null hypothesis is not tenable, the alternative one may relate to one - or two-sided discrepancy of the θ's. In this respect, we may formulate

$$H_0^*: \quad \xi(Z^{(1)}, Z^{(2)}) \equiv \xi(0 , Z^{(2)}), \tag{4.5}$$

for all $Z^{(1)}$ and $Z^{(2)}$, and hence we may require the "testability" of (4.4) through (4.5). This may

often call for a 'concordance' relation between the $\theta(\cdot)$ and $\xi(\cdot)$. Typically, if contrary to (4.4), $\theta(Z^{(1)}, Z^{(2)})$ depends on $Z^{(1)}$, then for the two nonidentical values, say, $Z_1^{(1)}$ and $Z_2^{(1)}$,

$$[\theta(Z_1^{(1)}, Z^{(2)}) - \theta(Z_2^{(1)}, Z^{(2)})] \, [\xi(Z_1^{(1)}, Z^{(2)}) - \xi(Z_2^{(1)}, Z^{(2)})]$$

should have the same sign, $\qquad\qquad (4.6)$

and moreover, a divergence in (4.4) should be somewhat isomorphic to that in (4.5), although it need not be a strictly linear pattern. If testing of a null hypothesis of no treatment effect is our primary goal, then based on a surrogate endpoint, a variety of nonparametric or semi-parametric tests can be constructed. Among these, the partial likelihood approach is quite popular, although genuine rank tests based on appropriate martingale characterizations [vis., Chapter 11 of Sen (1981)] can as well be extended to such **imd's**. There has been some attempts to restructure the Prentice (1989) formulation of a surrogate end-point [vis., Pepe (1992)], and partial likelihood or rank based tests work out neatly in such a case too. However, one may generally need to motivate such tests through the multivariate permutation principle [developed in detail in Chatterjee and Sen (1964) and unified further in Chapter 5 of Puri and Sen (1971)], so that such tests are generally permutationally (conditionally) distribution-free. The main advantage of such rank based procedures is that the concordance-condition in (4.6) can be verified without necessarily imposing a linearity constraint on $\theta(Z^{(1)}, \cdot)$, $\xi(Z^{(1)}, \cdot)$ or on their regression relation.

We consider the dual problem of estimating $\theta(\cdot)$, $\xi(\cdot)$ and $\psi(\cdot)$ in a nonparametric fashion. In the simplest case, when we have $Z_i^{(1)} = 0$ or 1 according as the ith observation belongs to the placebo or treatment group, and there is no other concomitant variate, i.e., $Z^{(2)} = 0$, for the surrogate set I [in (4.2)], we have essentially two subsets corresponding to the placebo and treatment group. For each subset, we may then estimate $\xi\ (\cdot)$ (i.e., $\xi(0)$ and $\xi(1)$ by the conventional method (as the designated quantile of the respective sample distribution). For the validation set I* [see (4.2)], in this simple model, we may consider the two bivariate sample distributions (of X, Y) for the placebo and treatment groups. For each of these sample distributions, we consider the two marginal sample distributions which, in turn, provide the sample quantiles as estimators of $\theta(\cdot)$ and $\xi(\cdot)$. Thus, from the validation set I*, we have estimators of $\xi(0)$, $\xi(1)$, $\theta(0)$ and $\theta(1)$. We denote these estimators by

$$\widehat{\xi}_v(0), \ \widehat{\xi}_v(1), \ \widehat{\theta}_v(0) \ \text{and} \ \widehat{\theta}_v(1). \qquad\qquad (4.7)$$

The estimators of $\xi(0)$ and $\xi(1)$ obtained from the surrogate set are denoted by

$$\widehat{\xi}_s(0) \text{ and } \widehat{\xi}_s(1), \text{ respectively.} \tag{4.8}$$

Note that the estimators in (4.8) are based on comparatively larger sample sizes (than in (4.7)), and hence, in studying the regression relation between $\theta(\cdot)$ and $\xi(\cdot)$, we incorporate this additional information on the $\xi(\cdot)$. Consider the two (bivariate) points $(\widehat{\xi}_v(0), \widehat{\theta}_v(0))$ and $(\widehat{\xi}_v(1), \widehat{\theta}_v(1))$. By an appeal to standard multivariate nonparametric methods [Chapter 5 of Puri and Sen (1971)], it can be easily shown that

$$n_{v0}^{\frac{1}{2}} \begin{pmatrix} \widehat{\xi}_v(0) - \xi(0) \\ \widehat{\theta}_v(0) - \theta(0) \end{pmatrix} \sim \mathcal{N}_2(\underline{0}, \Gamma_0), \tag{4.9}$$

where n_{v0} is the sample size of the subset of I^* for which $Z = 0$, and letting n_{v1} be the parallel number for $Z = 1$,

$$n_{v1}^{\frac{1}{2}} \begin{pmatrix} \widehat{\xi}_v(1) - \xi(1) \\ \widehat{\theta}_v(1) - \theta(1) \end{pmatrix} \sim \mathcal{N}_2(\underline{0}, \Gamma_1), \tag{4.10}$$

and where the two (asymptotic) dispersion matrices Γ_0 and Γ_1 may both be consistently estimated from the validation subsets (when n_{v0} and n_{v1} are both large). In a variety of situations, it may even be quite plausible to assume that $\Gamma_1 = \Gamma_2 = \Gamma$ (say), so that Γ may be estimated with a better rate of convergence. Likewise, if n_{s0} and n_{s1} stand respectively for the sample sizes for the subsets of I for which $Z = 0$ and 1, then we have

$$n_{s0}^{\frac{1}{2}} (\widehat{\xi}_s(0) - \xi(0)) \sim \mathcal{N}_1(0, \gamma_{0s}), \tag{4.11}$$

$$n_{s1}^{\frac{1}{2}} (\widehat{\xi}_s(1) - \xi(1)) \sim \mathcal{N}_1(0, \gamma_{1s}), \tag{4.12}$$

where both γ_{0s} and γ_{1s} may be consistently estimated from the sample outcomes, and assuming further that $\gamma_{0s} = \gamma_{1s} = \gamma_s$, a better consistent estimator of γ_s may also be obtained along the same line. Further, note that (4.9) and (4.10) justify the (asymptotic) linearity of the estimators of $\theta(0)$ and $\theta(1)$ (by virtue of the linearity of regression lines). This leads us to consider the following two estimators of $\theta(0)$ and $\theta(1)$:

$$\widehat{\theta}_{vs}(0) = \widehat{\theta}_v(0) + \widehat{\beta}_0\{\widehat{\xi}_v(0) - \widehat{\xi}_s(0)\}, \tag{4.13}$$

$$\widehat{\theta}_{vs}(1) = \widehat{\theta}_v(1) + \widehat{\beta}_1\{\widehat{\xi}_v(1) - \widehat{\xi}_s(1)\}, \tag{4.14}$$

where the regression coefficients $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are to be obtained by the classical (normal theory) linear inference procedures by using the estimates of Γ_0 , Γ_1 , γ_{0s} and γ_{1s} . Essentially, (4.13) and (4.14) incorporate adjustments to (4.9) and (4.10) based on the additional information in (4.11) and (4.12). This procedure extends directly to the case where $Z^{(1)}$ may have more than 2 values (as is the case with the usual more than one treatment vs. placebo studies). The picture becomes somewhat more complex when the concomitant vector $Z^{(2)}$ contributes additional variation in the outcome. If $Z^{(2)}$ contains coordinate variables which assume only discrete (or categorical) mass points (viz., smoking vs. nonsmoking, diabetic vs. not etc.), then on the basis of the possible values of $Z^{(2)}$ (say, denoted by a_r, $r = 1, ..., K$), we may classify the $\theta(\cdot)$ and $\xi(\cdot)$ as

$$\theta(Z^{(1)}, a_r), \quad \xi(Z^{(1)}, a_r), \quad r = 1, ..., K, \tag{4.15}$$

and for each of these functionals, we may proceed as in before (but based on the respective subset sample sizes). The larger is the value of K, we need a greater number of sample observations. In a general situation, some of the coordinate variables in $Z^{(2)}$ are continuous, while the others are discrete/ordinal. To handle this model, first, we consider the simplest case where $Z^{(2)}$ is a scalar random variable having a continuous distribution (e.g., cholesterol level). Suppose that we want to estimate $\theta(Z^{(1)}, x_0)$ for a given x_0 (e.g., cholesterol level, 200). We let

$$W_i = | Z_i^{(2)} - x_0 | , \quad i \, \epsilon \, I \, \text{ (and } i \, \epsilon \, I^*\text{)} . \tag{4.16}$$

Within each subset of I (and I^*) (as in (4.7)), we consider a set of observations for which the W_i have the smallest k values, where k is not very small, but k/n_{v0} (or k/n_{v1}, etc.) is small. Based on the (induced) observations for these k W_i, we proceed as in (4.7) through (4.14) and estimate $\theta(Z^{(1)}, x_0)$. This procedure can be repeated for various x_0 of particular interest. We may refer to Gangopadhyay and Sen (1992 a, b) for some details of the related statistical theory. In this context, the choice of k (relative to n_{v0} etc.) is an important task. Unlike in (4.9), (4.10), (4.11) or (4.12), we do not have a "square root n" order of convergence; they are typically of the order n^a for some $a \leq 2/5$. Also, for $a = 2/5$, the bias term may not be negligible, even asymptotically. As such, generally a is chosen as less that 2/5 [vis., Theorem 3.1 of Gangopadhyay and Sen (1992a)] .

We conclude this section with some remarks on the choice of the validation subset I^*. From what has been discussed before it is clear that the statistical modeling and analysis considered above rest on the tacit assumption that (X_i , Z_i) , $i \, \epsilon \, I$ have the same distribution as the (X_i , Z_i) , $i \, \epsilon \, I^*$, so that the relationship between $\theta(\cdot)$ and $\xi(\cdot)$ can be consistently estimated from the validation subset,

and then incorporated in the set I to provide more informative estimators of $\theta(\cdot)$. Such a tacit assumption may run to the contrary in a practical situation. For example one may adopt the following scheme: So long the surrogate variable X_i lies below a threshold value (say, x_0), measure (X_i, Z_i) and characterize these as belonging to I. On the other hand, if the surrogate variate exceeds x_0, its plausible crudeness may have more serious impact on the estimability of $\theta(\cdot)$ from $\xi(\cdot)$, and hence, a validation subset I^* should contain those observations for which $X_i \geq x_0$. A similar censoring may also be done for the covariates some of which may dictate the choice of the Y_i instead of the X_i. Essentially then the surrogate set I pertains to the conditional distribution $G(x \mid z)$ truncated from the right at x_0, while the validation subset I^* pertains to the bivariate conditional distribution $H(x, y \mid z)$ truncated from the left of x_0. Thus, we have a Type I truncation on the surrogate variate, so that the number (n_s and n_v) of experimental units belonging to the surrogate and validation sets are generally random variables assuming nonnegative integer values. Basically, for a given n (of preplanned number of experimental units), n_s/n and n_v/n converge to the respective probabilities of X being $\leq$ or $> x_0$. Thus, corresponding to the d.f. $G(x \mid z)$, we define the adjusted d.f.

$$G_R(x \mid z) = G(x \mid z)/G(x_0 \mid z) , \ x \leq x_0 , \tag{4.17}$$

and consider the functional

$$\xi_R(z) = \xi(G_R(0 \mid z)) . \tag{4.18}$$

Similarly, for the validation set I^*, we define

$$H_L(x, y \mid z) = \{H(x, y \mid z) - H(x_0, y \mid z)\}/\{1 - H(x_0, \infty \mid z)\}, \tag{4.19}$$

for $x > x_0$ and $y \epsilon R^+$. Then, the marginal d.f. of X (given $Z = z$) corresponding to $H_L(x, y \mid z)$ is

$$G_L(x \mid z) = [H(x, \infty, y \mid z) - H(x_0, y \mid z)]/\{1 - H(x_0, \infty \mid z)\}, \tag{4.20}$$

for $x > x_0$. Thus, parallel to (4.18), we set

$$\xi_L(z) = \xi(G_L(\cdot \mid z)) . \tag{4.21}$$

Also, the marginal d.f. of Y (given $Z = z$) is given by

$$F_L(y \mid z) = [H(\infty, y \mid z) - H(x_0, y \mid z)]/\{1 - H(x_0, \infty \mid z)\} , \qquad (4.22)$$

for $y \in R^+$, and we may set

$$\theta_L(z) = \theta(F_L(\cdot \mid z)) . \qquad (4.23)$$

Because of right and left truncations for X (in I and I^* respectively), $\xi_L(z)$ and $\xi_R(z)$ are generally not the same. Moreover, the relationship between $\theta_L(z)$ and $\xi_L(z)$ may be different from that between $\theta(z)$ and $\xi(z)$. Thus, the asymptotic normality results [as in (4.9) and (4.10)] on the estimates of $\theta_L(z)$ and $\xi_L(z)$ and on $\xi_L(z)$ [as in (4.11) − (4.12)] may not lead to improved estimators [as in (4.13) − (4.14)] of $\theta(z)$, unless $\xi_L(z)$ and $\xi_R(z)$ and $\theta_L(z)$ and $\theta(z)$ are related by some functional forms (which are estimable from the two sets I and I^* of observations). To enforce such estimability conditions, it may be quite tempting to consider some simple semi-parametric models [such as the proportional hazard model of Cox (1972)] which relate to a functional form of $\xi_L(z)$, $\xi_R(z)$ and $\xi(z)$, and $\theta_L(z)$ and $\theta(z)$ as well as on $\psi_L(\cdot)$ and $\psi(\circ)$. However, from robustness considerations such semi-parametric models may not be ideal; even a small departure from the assumed model may cause a great distraction of such simple relations, and may thereby lead to significant biased (if not inefficient) pictures. As a partial remedy it may be better to extend the validation set I^* by incorporating some additional observations for which the X_i are $\leq x_0$ and then recording the Y_i for them as well. This will enable one to estimate the relationship between $\xi_L(z)$ and $\xi_R(z)$ in a natural nonparametric way, and also the relationship between $\theta(z)$ and $\xi_L(z)$, $\xi_R(z)$ can be estimated in a nonparametric fashion from this extended validation set. From robustness considerations, we would endorse such a validation set I^* instead of the left truncated one considered before. In this context, optimal design problems are quite important ones to study systematically, and we propose such studies for future research. However, in passing, we may stress the importance of a proper selection of a validation set I^* which should not only meet the practicality of the measurements of the primary and surrogate variates in the given context but also be capable of eliminating bias and inefficiency to the extent possible. Thus, they should be robust. Therefore optimality (even in an asymptotic setup) of such clinical designs should not only take into consideration the relevant imd factors but also robustness of the underlying model. Naturally, such solutions are generally more complex. Nevertheless, they should not be outside the reach of statisticians primarily working in complex (multicenter) clinical trials. Finally, it may also be possible to have for a primary endpoint, a set of surrogate endpoints, so that X_i's are p-vectors for some $p \geq 1$. Because of some technical problems in defining a multivariate quantile function

(other than as a vector of the marginal quantiles), it may be wiser to consider a *cutting function* $X_i^* = h(X_i)$ [when the X_i are p-vector] and treating the X_i^* as the surrogate variate. The choice of $h(\cdot)$ may not be that evident, but from various clinical and other extraneous considerations, often, this can be made conveniently. Once, this is accomplished, we have no problem in incorporating the theory developed earlier to provide the desired methodology. The choice of an optimal cutting function may therefore have a significant role in the contemplated design and statistical analysis aspects of the study.

References

Chatterjee, S.K. and Sen, P.K. (1964). Nonparametric tests for the bivariate two sample location problem. Calcutta Statis. Assoc. Bull. 13, 18-58.

Cox, d.R. (1972). Regression models and life tables (with discussion). J. Roy. Statist. Soc. B 34, 187-220.

Dabrowska, D.M. (1988). Kaplan-Meier estimate on the plane. Ann. Statist. 16, 1475-1489.

Gangopadhyay, A.K. and Sen, P.K. (1992a). Contiguity in nonparametric estimation of a conditional functional. In Nonparametric Statitics and Related Topics (ed. A.K.M.E. Saleh), North Holland, Amsterdam, pp. 141-161.

Gabgopadhyay, A.K. and Sen, P.K. (1992b). Contiguity in Bahadur type representation of a conditional quantile function and application in conditional quantile process. (Preprint).

Gill, R.D. (1990). Multivariate survival analysis. Tech. Report 621, Department of Mathematics, University of Utrecht.

Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis. Ann. Statist. 18, 1501-1555.

Hanley, J.A. and Parnes, M.N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. Biometrics 39, 129-139.

McCullagh, P. and Nelder, J. (1989). Generalized Linear Models, Chapman and Hall, London.

Pepe, M.S. (1991). Inference for events with dependent risks in multiple endpoint studies. J Amer. Statist. Assoc. 86, 770-778.

Pepe, M.S. (1992). Inference using surrogate outcome data and a validation sample. Biometrika (to appear).

Prentice, R.L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. Statist. Med. 8, 431.440.

Prentice, R.L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. (To appear).

Puri, M.S. and Sen, P.K. (1971). Nonparametric Methods in Multivariate Analysis. John Wiley, New York.

Roy, S.N., Gnanadesikan, R.N. and Srivastava, J.N. (1971). Analysis and Design of Certain Quantitative Multiresponse Experiments. Pergamon Press, New York.

Sen, P.K. (1981). Sequential Nonparametrics. John Wiley, New York.

Srivastava, J.N. (1966). Incomplete multiresponse designs. Sankhya A28, 377-388.

Srivastava, J.N. (1968). On a general class of designs for multiresponse experiments. Ann. Math. Statist. 39, 1825-1843.

Wei, L.J. , Lin, D.Y. , and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J. Amer. Statist. Assoc. 84, 1065-1073.

Wittes, J. , Lakatos, E. and Probstfield, J. (1989). Surrogate endpoints in clinical trials: Cardiovascular diseases. Statist. Med. 8, 415-425.