Department of Statistics Library

CENSORING IN THEORY AND PRACTICE: STATISTICAL PERSPECTIVES AND CONTROVERSIES

by
Pranab Kumar Sen
Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2143

January 1995

CENSORING IN THEORY AND PRACTICE : STATISTICAL PERSPECTIVES AND CONTROVERSIES

PRANAB KUMAR SEN
University of North Carolina at Chapel Hill, NC 27599-7400, USA

Abstract

Censoring schemes of diverse types arise in practice in clinical sciences as well as in reliability studies. Statistical analysis of such censored data rests on sound theory and methodology wherein some basic regularity assumptions are needed which may not always be tenable in practice. Therefore, there is a need to examine the impact of less than ideal regularity assumptions on validity and robustness of such statistical analysis schemes. In the context of some practical models, statistical perspectives and controversiesarediscussed.

1 Introduction

Essentially censoring schemes relate to life time data analysis. Any system, whether mechanistic or biologic, has a life span. Replacement of inactive or weak components by new ones prolongs the life of a mechanistic system, while in a biologic system such a replacement may not always be feasible, so that proper care may be needed to maintain life quality. In either context, censoring may be defined as a termination of the observation-life due to some cause other than the natural failure to which the system is subjected to. To appreciate fully this broad interpretation of censoring and to comprehend the usual complexities of statistical analysis procedures for such censoring schemes, it may be wiser to pressure a distinction between mechanistic and biologic systems. The underlying regularity assumptions may vary considerably, and that factor may call for somewhat different statistical analysis tools in different contexts. In such life-testing models, biologic systems may often pose some challenging statistical problems. The main thurst of this study is on such biologic systems for which standard statistical analysis of censored data may not work out well. Such biologic systems include human organs, blood, body (somatic/ germ) cells, and a variety of other organisms. Survival analysis is potentially adoptable in this broader domain, albeit there are some undercurrents which deserve a critical appraisal. On the other hand, reliability theory has its genesis in mechanistic systems and is more popular in operations research and systems analysis setups. Since for both mechanistic and biologic systems, lifetime is characterized by a failure (or death) of the system, there is scope for adaptation for both the branches. The main difference between the two is the very basic fact that in

biologic systems, sickness or substandfard quality of life may, often, precede a failure, and that may call for more complex models. The presence of numerous *surrogate response* variates, some of which may qualify for *concomitant* variates is another typical characteristic in biological systems.

In a conventional model, *censoring* may crop up in various ways. In the literature, *Type I censoring* (or *truncation*), *Type II censoring* and *random censoring* schemes have been studied extensively, and *progressive censoring* schemes are also often encountered in clinical trials and medical studies. For each such censoring schemes, appropriate regularity assumptions are needed to validate some standard statistical inference tools for drawing conclusions. The so called *proportional hazard models (PHM)*, introduced by Cox (1972), have invaded almost all walks of survival analysis, although the underlying regularity assumptions may not always be very appropriate, particularly, in the current context, and hence, there is a need to examine them critically from applications point of view. There is considerable diversity in the life and death phenomena of a system, so that diverse statistical modeling and analysis schems are needed for their studies. In dealing with biomechanistics, it may be wiser to take into account various extraneous factors and vital biological undercurrents to formulate appropriate statistical models and analysis tools, and, often, a controlled study in the conventional sense may not work out that effectively. For this reason, in Section 2, we introduce the basic differences in the setups of a mechanistic and biologic system. Section 3 deals with some common biologic systems where there is a considerable need for developing statistical models and analysis schemes. Section 4 is devoted to the statistical perspectives in this formulation, and the concluding section raises the general issues relating to censored data analysis for biological systems.

2 Mechanistics vs. Biomechanistics

In a mechanistic setup, a planned controlled study may generally work out well. In a biomechanistic setup, although a number of influential factors can be identified and their impacts on the system can be interpreted in a meaningful way, it may not generally be possible to achieve a *controlled expreimental design* setup that effectively. Thus, there may be a greater need to incorporate more sophisticated statistical tools for the purpose of planning, modeling and statistical analysis of biomechanistics. Both reliability theory and survival analysis are relevant and useful in this context. Yet these two broad disciplines may differ, often, drastically, in their foundation as well as operational approach. Recent advances in this broader domain have opened up new frontiers of statistical research and eliminated the impasses to a greater extent. There is ample room for further development of statistical research in this complex field, albeit the task is quite a challenging one.

To motivate the main points of difference between mechanistic and biomechanistic systems, let us start with the classical (limiting) *system availability* problem, treated in detail in a parametric mold by Barlow and Proschan (1975); for some related non-

parametrics, we may refer to Sen (1995). Consider a single-unit system supported by a single spare and a repair facility. When the operating unit fails, it is replaced instantaneously by the spare while the repair facility takes charge of the failed unit. Upon repair, the unit is transferred to the spare box. The system fails when an operating unit fails and the spare box is empty. This happens when the repairing time (Y) of the previously failed unit is larger than the life time (X) of the operating unit about to fail. It is assumed that upon repair a unit has the same life distribution (say, F, defined on R^+) as of the original component, and moreover, the repair time distributions of the successive failed units are the same (say, G, defined on R^+). Then, as in Barlow and Proschan (1975), the *limiting system availability* is defined as $A = E[\text{system life time}]/\{E[\text{system life time}] + (1-\alpha)E[\text{system downtime}]\}$, where

$$\alpha = P\{X \geq Y\} = \int_{R^+} G(x)dF(x). \tag{2.1}$$

The number of times the repair facility is used before a system failure occurs is a positive interger valued random variable, so that if the study is planned for a fixed duration of time, censoring may crop up in a natural manner. Barlow and Proschan (1975) assumed further that both F and G are exponential distributions with possibly different means, and this simplifies the expression for A considerably. In a general nonparametric case, a formulation is due to Sen (1995). In either case, if several independent copies of the system are in operation, censoring may occur when a system failure does not occur during the tenure of the study. For a mechanical system when the repair facility is enough sophisticated, the assumption of independent and identically distributed X_i (and Y_i) can be justified to a reasonable extent, while in the simple exponential model, a discounted life distribution for repaired units may also be worked out conveniently. Let us then examine how far such simple regularity assumptions are tenable for some common biomechanistic systems. As an illustrative example consider the following:

Operating Unit : Human Heart

Spare Box : Donors (?)

Repair Facility : Biologically highly complex.

There are some basic issues relating to the repairability (viz., heart transplants or bypass surgeries) which makes the problem far more complex than the one referred to above. Moreover, in the case of heart transplants, donors may not be instantaneously available, and there are many biological factors underlying the matching of donor hearts for the recipient. Age, general physical conditions, diet and smoking habits and many other *covariates* or *auxiliary variables* contain good deal of information on the response variate which are needed to be taken into account. Further, any such study can at most be carried out for a reasonable length of time, so that *heavy censoring* may occur. The assumptions tenable for a simple mechanistic system are generally not that appropriate here. In a mechanical setup, the units under study may all be procured at a common time, but in biomechanistics, generally not all the subjects (units) enter into the scheme at a common point of time, so that a *staggering entry plan* is generally encountered in practice. Moreover, because of withdrawl or drop out *noncompliance* is quite common

in any such study. In a standard statistical analysis scheme, such a noncompliance is tacitly assumed to be independent of the particular treatment to which a subject is administered. This assumption is likely to be violated in many cases. For example, in the same cardiovascular problem referred to earlier, suppose that the theme of the study relates to the impact of lowering of the *cholesterol level* on the *risk* of heart attacks, and the subjects are divided into a treatment group (where cholestorel level is reduced through some medical plan) and a placebo group (where no such medicine is used). It is likely to have higher a compliance rate in the treatment group than in the placebo group. This may therefore lead to a violation of one of the basic assumptions : *The censoring variable is distributed independently of the primary response variate.* Recall that in *random censoring* schemes, if X and C stand for the response and censoring variate respectively, then it is tacitly assumed that they are mutually stochastically independent and moreover, although X may have a distribution depending on the particular treatment group it is assigned, C would have the same d.f. for all such groups. In the possible negation of this basic regularity assumption, standard statistical analysis tools may not be validly applicable.

There are other biomechanic models where there is a more visible impact of standard mechanical models, albeit there are some hidden complications requiring more sophisticated statistical analysis. Metal imputation of hip-joints, artificial limbs etc. have clearly some non-living material component(s) which the host living organs may not accept as totally compatible. Thus biomechanic systems may inherit some features of mechanic systems but mingled with more complex stochastics which can only be pursued thoroughly with deeper biological or clinical backgrounds. These features, anticipating more complex types of censoring or incolmpleteness of acquired data sets, may introduce more complications in modeling a biomechanistic system and formulating appropriate statistical analysis schemes.

It is clear from the above discussion that in biomechanistics *analysis of censored data* is highly dependent on the formulation of *biomechanic models* which would be appropriate for the particular project. This calls for a more intensive study of the *robustness* aspects of statistical analysis procedures when departures from the assued model can take rather complex routes. For this reason, it may be imperative to look critically into the biological undercurrents with a view to making allowances for model departures for which robustness should be given due emphasis. The advent of modern *biotechnology* has opened the doors for more adaptations of mechanistics in biological systems, albeit at the cost of considerable complexities in modeling as well as analysis schemes. As such, before examining the statistical perspectives, we present the salient points of modern biomechanistics through a number of important examples.

3 Mechanistics in Biomechanistics

With the phenomenal growth of research literature on biotechnology and ever-widening knowledge of physical, chemical and anatomical structures of our body, it will be no

exaggeration to depict ouselves as exceptionally complex pieces of machinery. Replacement of malfunctioning or damaged parts of the body with either organic transplants or even artificial equipments has been gaining popularity day by day, and there is good hope for more successful implementation in near future. *Heart pacemakers* are commonly used to mediate a regular heartbeat for damaged hearts; *bypass surgeries* have become a household word, and even *heart transplants* are instituted with increasing degree of confidence. The impact of fatty substance in blood on cardiovascular disorders has been thoroughly investigated and promise for a better quality of life is on the verge of realization. Bones and joints made of inert plastics and metal alloys are used to replace shattered or diseased bones. Infertility problems are now being taken care of by *fertility clinics* using modern biotechnology to a greater extent. Metabolic functions may now be adjusted or modified by using *capsules* and *pills* which release a necessary metabolite, such as *hormones* or *coenzymes*, over a length of times or under perticular conditions. The use of *estrogen* capsules following *hysterectomy* is quite prevalent in modern times. Food tablets and vitamines are quite acceptable by now to keep our bodies healthy. It is hoped that in the very near future, damaged *alveoli* of lungs would be replacable by new materials which would allow gaseous exchange and maintain the operability of our inhalation system which may otherwise be plunged by environmental or other toxicants. The use of *DNA* not only for medical purposes but also in criminal disputes has been gaining grounds everyday. Yet there are road blocks to a sweeping success for modern biotechnology. Side-effects of drugs or chemical reagents are surfacing at an alarming rate, and advances in *molecular biology* are producing alarming evidences mingled with *genotoxicological* explanations. The daily consumption of (drinking) water can not be reduced very much as it is necessary for many metabolic functions. But the *water contamination* problem is signaling serious threats to our vital life thrust. We can not live without adequate supply of oxygen, but the quality of air we breathe in is no longer safe; evironmental pollutants are steadily increasing the risk of health hazards and the inhalation factor is the prime one in this respect.The use of *antibiotics* has been on the caution: repeated use of antibiotics results in an increasing sequence of doses and becomes ultimately ineffective, and often, loaded with serious toxic side-effects. Plastic/metal replicates of human bones or joints do not have the *bone marrow* to produce the much needed blood cells, and there are thousand and one other accountable or unaccountable factors whic make the modern biotechnology still far from being *totally safe* for universal adoptation. While some of these pollutants can be controlled to a certain extent, we may not have sufficient overall control over our *environment.*

To iterate the impact of mechanistics in biomechanistics, let's consider some of the simpler models :

1. *Human Renal System*: Do the two kidneys function as if they are *in parallel/ in series* ? Are they really exchangeable in statistical sense ? How one of the two kidneys performs when the other one acquires some disease or disorder?

2. *Respiratory System*: Are the two lungs functionally and/or statistically exchangeable? What is the most probable site of development of a lung-tumor : Bronchus or Alveoli ? How to depict the movement of carcinogenic cells from the lungs to other parts of the body?

3. *Female Reproductory System*: Whether the two Ovaries function in parallel/ in series? How One accepts the role when the other one is diseased or contains a fertilized egg? Genesis of ovarian cancer!

4. *Human Optical System*: The two optical nerves are known to have somewhat different functional roles, and the question of exchangeability or independence remains open. How the work load is shifted to one of the eyes when the other one has malfunctioning ?

5. *The Heart*: Do the *Ventricles* have similar functional roles? The biological system of pumping in and out of blood is highly complex, and can a simple reliability model be ascribed to such a mechanism? What happens if one of the ventricles becomes nonfunctional or weak? What type of model may describe best the control of the heart on the arterial flow of blood? How LIPIDS affect the functioning of the heart and the cardiovascular system as a whole?

6. *Ear*: The two ears have been identified as having different roles, and hence, can they be regarded as exchangeable in a statistical sense? Is there good evidence of hereditory effects in this respect?

7. *Central Nervous System*: Is it not a highly complex network model? How to handle this extremely high dimensional system? Which sectors of the *brain* are to be selected for studying the overall and/or specific functioning of the CNS?

There are numerous other examples. Basically, in all these cases, the following salient features are to be highlighted:

a. Biological structures/constructions and restraints;

b. Biomedical undercurrents;

c. Environmental impacts;

d. Carcinogenicity aspects;

e. Tomography vs. Oncology ;

f. Unobservable factors: Identifiability problems;

g. Predominant stochastics in the phenomena.

Both *Epidemiology* and *Biostatistics* are relevant for drawing conclusions from acquired data sets; but there are, often, fundamental differences in their foundations and operational procedures. There is, of course, scope for *Survival Analysis* as well as *Reliability Theory* in biomechanistics, but in view of the basic differences in biological setups, considerable amount of modifications of existing theory and methodology may be needed to validate actual adoptations in practice. With due emphasis on various censoring schemes, we shall illustrate some of the statistical problems by considering some specific biological models in a relatively more detailed fashion.

4 Statistical Perspectives and Controversies

Since censoring is most commonly encountered in survival analysis, we consider the following illustrative example, taken from Sen (1994a). It is a natural *aging process* which may lead to some blockage in the arterial channels, although such a phenomenon may have accelaration under *Arteriosclerosis.* In the case of the ascending aorta, carrying blood from the heart to the brain, a reduced blood (and oxygen) supply caused by such a constriction generally leads to malfunctioning of the central nervous system and the brian. A serious blockage may also lead to a *stroke* or a *cerebral thrombosis.* It is therefore necessary to eliminate this blockage to the extent possible, and among various clinical procedures to enact on this problem, (a) *medication* and (b) *surgery* are more popular. In a medication plan, persistent use of some drugs dilutes the congestions and increases the flow of oxygen rich blood. The medication may have some side-effects, often, quite serious to limit the use to a small time period, and hence, there may be deop outs or withdrawls. In some extreme cases, a surgery may be needed for life saving purposes. In a surgical plan, a sizeable part of the blockage is removed, and following the surgery an intensive care is needed for a short while. But all patients may not be strong enough for the surgery. Thus, there may be a basic question about the practicality of *randomization* in such a study, and sans randomization, the classical statistical analysis tools may not be validly applicable. In addition, there are generally numerous covariates and these are needed to be taken into account in statistical modeling and analysis of experimental outcomes. As in other clinical investigation, the study-period (including the follow-up time) is generally fixed, so that censoring is quite prevalent. It may therefore be quite appealing to take recourse to the Cox (1972) PHM to compare the relative risk of medication vs. surgery. However, there are some genuine concerns regarding an unrestricted use of PHM in this particular context. If we compare the *hazard functions* for the medication group and the surgery group, we may observe that in the former case, we have a smooth and usually nonincreasing function, while in the latter case, during the surgery or the period immediately following it, the hazard is likely to be quite high with a sudden drop after that. Thus, the two hazard functions may not be proportional even in a local sense. Therefore, from statistical analysis point of view, there are two basic (controversial) issues: (a) validation of randomization, and (b) appropriateness of the PHM or the *multiplicative-intensity counting process* approach.

Murphy and Sen (1991) have considered a *time-dependent coefficients* Cox regression model approach which eliminates the second drawback to a greater extent, but may generally entail a slower rate of convergence for the parameters (which are then functionals instead of being finite dimensional). Sen (1994a) has applied this model for a *change point* testing problem and their results are virtually applicable to a larger domian of suvival analysis models where the PHM assumptions may not be that appropriate. One advantage of this approach is that even the validation of randomization can be made under suitable experimental setups.

As a second illustrative example, consider the Arteriosclerosis problem with respect to the risk of heart disorders. High cholesterol is suspected to be associated with higher risk for heart diseases (including Angina Pectors, heart attacks, strokes, Parkinson's disease etc.), and it is also believed that hypertension, obesity, smoking, diabetes, irregular or insufficient physical exercise and disability may significantly contribute towards the increase of the cholesterol level in the blood. All of these hypotheses were based on empirical evidences by the medical and epidemiological investigators, and the acquired data sets had clearly *observational study* flavor. The National Heart, Lung and Blood Institute (NHLBI) planned a multicenter clinical trial to gather statistical evidence to support their conjectures in an objective manner, and the University of North Carolina, Chapel Hill was given the task of statistical formulations, data monitoring and analysis. A *controlled clinical trial* protocol was adopted with a view to incorporating randomization to the extent possible. The study was planned for a maximum duration of 12 years (1972-1984). The recruited subjects were all male between the ages 35 to 50, having high cholesterol level to start with but no previous heart problem. Since the 12 year mortality rate derived from from the US *Life Table* was estimated as close to 0.11, heavy censoring was taken for granted. For this reason, in order that a test for a suitable hypothesis has reasonable power, a much larger number of observations would be needed. In the current context, a sample of 3952 subjects was chosen from a pool of volunteers of more than 100,000. Recall that it was deemed to be a Phase III trial, so that no particular subject would be subjected to a higher risk because of being placed in the placebo or the treatment group. On the top of that, *medical ethics* prompted the consideration of an *early termination* of the trial if the accumulated evidence upto that stage indicates a clear cut difference between the two groups. Or, in other words, *interim analyis* on the experimental outcome was proposed as a part of the protocol. This resulted in the adoptation of the so called *time-sequential* statistical inference procedures, which are related to the conventional *repeated significance testing* schemes, although the decision rules need not be the same in both the setups. Keeping in mind the plausibility of heavy censoring and interim analysis, from a statistical perspective, it was deemed that *progressively censored schemes (PCS)*, developed by Chatterjee and Sen (1973), would be more appropriate than the usual fixed point right truncation or right censoring schemes. Moreover, drop-out or noncompliance was not ruled out, and hence, the random censoring mechanism was mingled with the progressive censoring scheme. Further, the recruitment of nearly four thousand eligible subjects

took almost a two year period, so that effectively, there was a *staggering entry* pattern which introduced more complications in statistical modeling and analysis schemes. As the accumulating data set was acquired from multiple clinics scattered over the nation, there were variations between clinics as well as within clinics between patients who were quite heterogeneous with respect to their concomitant variates (and the latter was a plus point for the epidemiologists for their analysis). The Cox PHM was not totally ideal in this setup, although it could have been taken as an approximation if there were no staggering entries. The entry pattern was not very regular so as to reduce the case to a prespecified entry pattern model. Finally, as the projected study period was long (12 years), a change in the attitude of the sample units (subjects) towards diet, physical exercise, smoking etc. might have been possible, and some allowance for this effect should have been given in the statistical modeling as well as analysis schemes. These factors clearly conveyed the message that a simple Type I or II or random censoring model based data analysis would be inappropriate and grossly inadequate too. Moreover, a general nonparametric approach rather than the PHM model or the specific exponential model would be more appropriate from validity and robustness points of view. This led to the development of *time-sequential nonparametrics* in the late seventies, and some account of this development is contained in the last chapter of Sen (1981). To cope with the provision of interim analyses allowing possibly a moderately large number of *statistical looks* into the accumulating data set, instead of the conventional *asymptotic normality* results, *weak invariance principles* were established wherein weak convergence of suitably constructed stochastic processes to appropriate Gaussian functions was established and incorporated in the construction of suitable critical values for the test statistics to be used in interim analysis. Statistically speaking, if there are multiple tests for a common null hypothesis, use of the conventional coordinate test wise critical values may generally result in a much higher level of significance, and the above mentioned weak convergence result provides an easy way of maintaing the overall significance level at a preassigned level. Moreover, it plays a vital role in the study of the *stopping time* of a test procedure in interim analysis.

As a third illustration, let us present a biomedical model relevant to the study of *inhalation toxicolgy*. The human respiratory system consists of the vital components: nose, mouth, larynx, trachea, bronchea, bronchioles and the alveoli. Oxygen reach air is breated in and within the lungs, the alvioli allow the gaseous exchange of carnon dioxide and water vapour (to be breated out) with the oxygen which is conveyed by the pulmonary vein for circulation in the body. Dust particles, toxicants of various kinds and fumes from tobaco smoking, environmental smoking or automobile exhausts are trapped at various stages of the respiratory channel, resulting in the the most serious effect: blockage of the air sacs in the alvioli. This reduces the flow of oxygen rich air and is supposed to be an important factor in the formation of a *lung tumor* initiating the disorder *lung cancer*. The process is, however, a slow one, and any scientific study needs to allow for an adequate time duration, so that only a proportion of the subjects may have identified carcinogeneity during that time. Thus censoring is inherent in

such a study. The basic difference between such a study and a conventional laboratory experiment is that here we may not be in a position to measure the dose or level of the pollution/ toxicants accurately, and there is much less control on the individuals who are to be used as subjects in such a study. Nevertheless, such individuals generally vary in their metabolic capacities, genetic traits, exposure levels, and hence, a conventional *deterministic lung model* relating to the flow of toxicant and blockage of alvioli may be far from being realistic, so that stochastic modeling is essential. Again, censoring is a vital component of any acquired data set, and associated with this phenomenon are a number of factors: (a) *measurement errors* for doses, (b) *unobservable responses*, (c) *confounding* effects for a large number of 'dose' variables and many others. On the top of that it is not yet clearly known whether the site of carcinigeneity is a bronchus or an alveolus. The latter sounds more likely, but once a cancerous cell is formed it can move to any other part of the body. It has yet to be ascertained whether smoking is a cause of cancer or it acts as a catalytic agent! Given this omnibus situation, it is very difficult to recommend the adoptation of a simple censoring scheme, such as Type I, II or random one, and any statistical modeling or analysis scheme to be appropriate and valid in this setup must take into account the underlying biological and environmental factors to an adequate extent. Actually, *molecular biology* is a fundamental component in such a study, and any statistical analysis scheme must have adequate biological justification as a prerequisite.

As a final illustration, let us consider the case of *antibiotics* which were introduced nearly 50 years ago and offered an arsenal of magic bullets to knock out the *germs* that cause anything from *Pneumonia* to *Gonorrhea*, *typhoid* to *T.B.*. These drugs have been widely used by the physicians, often, disregarding the excessive doses with which they were to be administered on successive occasions. After this long period of practice, it has become evident that (i) these antibiotics may generally have serious side-effects, (ii) they are becoming less *potent* for the treatment of specific diseases for which they were highly effective, and (iii) repeated use of such antibiotics requires an increasing sequence of doses, setting limitations on the number of times thay can be administered on the same subjects. *Genetical mutations* of these *microbs* provide a good explanation of the fighting back of these invisible legions of drug-resistant organisms. In practice, any attempt to study the complexities of these antibiotics for treatment of some specific type of disease (viz, *bubonic plague*) must therefore take into account these factors; these call for nonstandard models wherein censoring is an essential element for the experimental outcome. Drop-out or noncompliance may no longer be independent of the treatment group, and moreover, replications over time may not yield comparable copies of experimental outcomes. Elimination of severe side-effects may induce considerable incompleteness in the design of a study, and randomization may not be effectively implementable. Thertefore, standard Type I, II or random censoring schemes may not appear to be that relevant in this context, and a more micro-biologically compatible comprehensive censoring scheme has to be formulated. Thus, some nonstandard statistical analysis schemes are to be envisaged.

5 Statistical Prospects and Tasks

Perhaps, the discussions in the preceding two sections endow us with a good deal of insights on the basic controversies undelying the adaptation of standard censored data analysis tools for most of the biomedical and clinical studies commonly encountered in practice. Nevertheless, it may be emphasized that statistical principles are necessary for proper modeling and valid statistical analysis of experimental outcomes. In this context, our prospects become more visible if we keep in mind the following general features relating to censored data models in this general field:

i. In laboratory setups, experiments are generally conducted with reasonable amount of *control* over the experimental units. This basic feature is generally not tenable in survival analysis. In dealing with experiments involving human subjects who may come from different walks of life, there may not be much experimental control. Thus, at the very begining, we may wonder : *whither control*?

ii. In a conventional agricultural or physical/chemical study, generally, a *control* and one or more *treatment* groups are compared with respect to the response variable(s) pertaining to the experimental scheme. In environmental as well as biomedical studies, often, such a *treatment vs. control* setup may be difficult to implement, and hence, the priciples of *design of experiment* may not be of sufficient importance in the planning or statistical analysis of such studies. Thus, we may ask: *Whither treatment vs. control* ?

iii. In a conventional experimental setup, it is generally taken for granted that some sort of *linear models* hold for the response variate, where even it may be assumed that the error components are normally distributed in a homoscedastic pattern. In practice, often, a *transformation* on the response and dose variates (termed the *response meta-meter* and *dose meta-meter* or *dosage* , induces more homoscedasticity and close normality of the errors, although we may have to be careful that the *additivity* of the effects is not vitiated by this transformation. The classical Box-Cox transformations are quite popular in biological assays and other biomedical applications. Nevertheless, it may not be able to have simultaneous attainment of the *linearity* of the *dose-response* relation and normality of the error distribution for the transformed mode. Thus, we wonder: *Whither linearizing/normality-inducing transformations in biomedical and biomechanical applications*?

iv. In reliability theory and survival analysis, to deemphasize the role of exponential failure distributions (i.e., constant failure rates), various classes of life distributions characterized by their hazard or failure rates have been formulated, and often, they provide valuable guidelines to statistical modeling and analysis schemes. Within this setup, the Cox (1972) PHM occupies a focal point, and it has led to a phenomenal development of statistical methodology encompassing suitable counting processes which are commonly encountered in lifetime studies. Yet, we have seen

earlier that such models may not be universally appropriate in all biomechanic models. Thus, we wonder: *Whither PHM in biomechanistics*?

v. In view of (i)-(iv), there is a genuine need to develop more *adaptive designs* for biomedical and biomechanic models, and this, in turn, calls for development of more *robust statistical analysis* schemes which would allow more flexibility with respect to model-based regularity assumptions. In this context, nonparametric approaches have greater scope, but it needs to be checked whether such nonparametric methods are appropriate for the given size of the experimental outcome and set experimental objectives. Thus, we have: *Whither parametrics, semi-parametrics or nonparametrics*?

vi. Generally there a is greater scope for models incorporating *measurement errors* as well as *misclassifications* of states in biomedical and biomechanic studies. This feature may lead to some *identifiability* problems with respect to model specifications as well as valid statistical analysis. Therefore, any proposed model should be capable of dealing with measurement errors and misclassication of states in an identifiable manner, and within this broad setup, it should lend itself to some reasonably simple, robust and valid statistical analysis schemes. A majority of sophisticated mathematical tools and models for the classical case of i.i.d. random vectors may not be strictly usable in this context.

vii. In biomedical investigations, applications of experimental treatments (or doses) may result in a change in the distribution of the response variables for the *primary end point*, as well as numerous other auxiliary variates related to the main scheme of study. As such, there is a need to examine whether these auxiliary variates qualify for *concomitant* variates (whose distribution should not be affected by treatment differences) or they need to be treated in a *multivariate* setup so as to possibly allow their distributions to be dose-dependent. Inclusion of a large number of concomitant variates may result in greater scope for significant departures from conventional linear, log-linear or even the proportional hazard models. As such, *model-robustness* is highly desirable. Similarly, inclusion of a number of response variates in a general multivariate setup generally leads to a loss of *power* of suitable statistical tests or *efficiency* of estimators. Moreover censoring may take place in the primary response variate / concomitant variates, or noncompliance may even result in a complete missing observation on all the variates. This can be compensated by increasing the sample size considerably, and often, this may run contrary to the inherent size limitations of such experimentations. Development of *finite sample size* analysis tools for censored data arising in such multi-response multi-factor biomedical studies is indeed a challengig task.

viii. *Undesirable effects,* such as severe side effects of a drug or toxicity of some treatment, crop up in many biomedical studies. Such a phenomenon may even lead

to a curtailment of the study, resulting in a censoring. The main difference between such censoring and random censoring in the conventional sense is that the censoring variable may not be stochastically independent of the response variates. Thus, the conventional Kaplan-Meier (1958) *product limit* estimator based statistical methodology may not work out well in such cases. There is thus a genuine need to explore statistical methodology beyond the conventional frontiers of Type I, II or random censoring schemes. Hanley and Parnes (1983) explored the conventional multivariate survival function estimation in the presence of censoring, and yet there is an acute need to examine nonstandard situations which may typically arise in biomedical investigations.

ix. *Toxicity* in biomedical experimentations may have a more serious aspect, namely, the *genetic toxicology* or *mutagenesis*. Molecular biology plays a basic role in the assessment of such toxicity which goes far beyond the usual mode of chemical reaction, and for the genetic elements one needs to look into the impacts of molecular biology. Again, more complex models are generally needed to characterize the prognosis of such toxicity, and without due attention to such compatible models, the usual conventional censored data analysis tools may not be of much worth.

x. Since in biomedical studies, sample units are human subjects, and often, the primary response variate is death (failure), it may be too costly or impractical to record this primary end-point. In medical studies, often, *surrogate end-points* are used to draw valid statistical conclusions. However, in order that a set of auxiliary response variables may qualify for surrogate variates, it is necessary to check some statistical model constraints. Although this is largely an open area of research for (bio-)statisticians, there has been some good progress in realistic formulations of statistical methodolgy for surrogate end-point analysis of censored data; a special issue of *Statistics in Medicine*, vol.8 (1989), no.2, is devoted to this important aspect of statistical modeling and analysis. For further developments, we may refer to Pepe (1992) and Sen (1994b), among others.

xi. *Carcinogenicity* is prevalent in most biomechanical and medical systems. The devopment of carcinogens in a system, formation of tumors and prognosis of cancer are yet not totally scientifically traced out, although good progress has been made in the assessment of impacts of various catalytic agents (such as smoking, consumption of fatty foods, environmental pollutants etc.,). In terms of statistical modeling and valid statistical analysis of experimental data sets, there are some challenging problems. These relate not only to *heavy censoring* but also to the appropriateness of conventional biologic models when the etiology of the disorder is not that precisely known. The censoring pattern may highly be dependent on the progress of the disease, and an approprite model encompassing a rational prognosis and its relation to censoring is necessary for a valid statistical analysis of experimental outcome data sets. For example, in the case of lung cancer, the development of a tumor may not be detectable until at a very late stage, when the

disorder may be out of control for any effective treatment. Thus, any guess that can be made for the *growth of a tumor* model and its relation to noncompliance may have a strong impact on statistical modeling as well as analysis, so that the latter schemes are quite vulnerable to possible discrepancies between the assumed and the true models. In this context too, noncompliance may be highly dependent on the state of the disorder, so that the conventional independence assumption for the censoring variate may not be tenable.

xii. *Interdisciplinary Approach.* Because of the presence of significant stochastic elements in the menifestation of biomechanical or biomedical responses, ordinary biological models are generally not adequate. Communicability with biological and/or medical scientists is, however, essential in tracing down the important factors influencing the flow of the events, identifying the stochastics associated with them, and in view of these, to formulate the basic objectives of the study as lucidly as possible (in scientific as well as statistical terms). Often, this runs contrary to practice where scientists in other experimental sciences plan their experiments first, conduct them, gather scientific data-sets, and then seek the assistance of statisticians for statistical analysis. Planning of the experiment is by far the most important task, and this needs a lot of statistical expertise too. In dealing with experiments with significant censoring, such a planning becomes even more statistically important. Basically, censoring in interdisciplinary research is more complex in nature, and it needs attention from all pertinent directions.

xiii. *Meta-analysis.* During the past ten years considerable attention has been paid to the prospects of *pooling of opinion* from several independent but coceptually related experimental data sets, and in this context too, censored data sets can be fitted quite well. Thus, for multi-center studies relating to possibly censored observations, but of a common research objective, it seems quite intuitive to adopt the techniques of meta-analysis in arriving at general statistical conclusions with enhanced confidence (or reduced risks for making incorrect decisions). In the conventional case, one deals with independent studies having some concordance , and the *observed significance level* (OSL) based methodolgy can be directly adopted from the earlier work of Fisher (1932). In this context, it has been observed [viz., Sen (1983)] that the classical *step down procedure* can be incorporated to modify the Fisher's method so as to allow possible dependence across the sources; actually, this Fisherian detour of the step down procedure works out well for multivariate models including *longitudinal data models.* The last feature makes it particularly adaptable in biomedical and clinical setups where follow-up is a common phenomenon. The classical Mantel and Haenszel (1959) has also find its way in meta-analysis where the component sources may not have independent data sets; we may refer to Sen (1988) for some detailed discussion of such procedures. There are other procedures available in the literature (including some genuine Bayesian ones). But most of these have been worked out for conventional equal probability

sampling schemes which may not be that tenable in biomedical or biomechanical setups. This is largely an open area and there is an excellent chance for further useful developments.

Let me conclude with an optimistic note on the scope of statistical reasonings for censored data analysis pertaining to such complex biomedical and/or biomechanic systems. The wealth of research work already accomplished for standard models (and Type I, II and random censoring schemes) should serve as the foundation of any attempt to build more realistic statistical modeling and analysis schemes for such complex problems. This is indeed a challenging task. The statisticians have always come up with comprehensive methodology and analysis tools in many other nonstandard situations, and hence, my expectation is that incorporation of the complex biological background into planning of the study, sampling schemes, and data collection and monitoring protocols would lead us to the right path for valid and efficient statistical analysis of censored data arising in complex biological models. In this venture, mathematical sophistications are, of course, useful, but they may often preclude statistical foresights. We really need a blending of biological background with statistical reasonings in order to harness mathematical sophistications towards the desired resolutions.

References

[1] BARLOW, R.E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models,* Holt, Rinehart and Winston, New York.

[2] CHATTERJEE, S.K. and SEN, P.K. (1973). Nonparametric testing under progressive censoring. *Calcutta Statist. Assoc. Bull.* **22**, 13-58.

[3] COX, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc.* **B34**, 187-220.

[4] HANLEY, J.A. and PARNES, M.N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* **39**, 129-139.

[5] Fisher, R.A. (1932). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 4th ed.

[6] KAPLAN, E.L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481, 562-563.

[7] MANTEL, N. and HAENSZEL,W. (1959). Statistical aspects of the analysis of data from retrospective studies of disese. *J. Nat. Cancer Inst.* **22**, 719-748

[8] MURPHY, S.A. and SEN, P.K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochast. Proc. Appl.* **39**, 153-180.

[9] PEPE, M.S. (1992). Inference using a surrogate outcome data and a validation sample. *Biometrika* **79**, 355-365.

[10] SEN, P.K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*, J. Wiley, New York.

[11] SEN, P.K. (1983). A Fisherian detour of the step-down procedure. In *Contribution to Statistics: Essays in Honour of Norman L. Johnson* (ed. P.K. Sen), North Holland, Amsterdam, pp. 367-377.

[12] SEN, P.K. (1988). Combination of statistical tests for multivariate hypotheses against restricted alternatives. In *Proc. Internat. Confer. Multivariate Statistical Analysis* (eds. S. Dasgupta and J.K. Ghosh), Indian Statist. Inst., Calcutta, pp.377-402.

[13] SEN, P.K. (1994). Some change-point problems in survival analysis: Relevance of nonparametrics in practice. *J. Appl. Statist. Sc.* **1**, 425-444.

[14] SEN, P.K. (1994). Incomplete multiresponse designs and surrogate endpoints in clinical trials. *J. Statist. Plan. Infer.* **42**, 161-186.

[15] SEN, P.K. (1995). Statistical analysis of some reliability models: Parametrics, semi-parametrics and nonparametrics. *J. Statist. Plann. Infer.* **43**, in press.