March 2008, Presentation

The Burden of Bioinformatics in Clinical Trials : Some Statistical Perspectives and Controversies

Pranab Kumar Sen

University of North Carolina at Chapel Hill, NC 27599-7420, USA

The ongoing evolution of genomics and bioinformatics has an overwhelming impact on medical and clinical research, albeit its development is often marked by genuine controversies as well as lack of scientific clarities and acumen. The search for disease genes and gene-environment interaction has drawn considerable interdisciplinary scientific attention: Statistical reasoning has a pioneering role in this venture while data mining resolutions are far from being statistically fully understood or interpretable. In the genomics context, clinical trials may be designed on chips and yet there are greater challenges due to the curse of dimensionality perspectives. This scenario is ap-

praised here with due consideration of some recent developments.

A little over a period of three decades, clinical trials have mushroomed in a variety of human health studies, with a variety of objectives, having a variety of interdisciplinary perspectives, and diverse implementational motives. Clinical trials are designed by human beings, mostly, for human beings, incorporating mostly human subjects, and, supposedly, for human benefit. Yet in this human venture there are some inhuman features which warrant critical appraisal. Using human subjects in scientific (and mostly exploratory) studies may generally trigger *medical ethics, cost-benefit perspectives* and a variety of other concerns. In order to control some of these disturbing concerns, often, subhuman primates are advocated as precursors or surrogates of human being, albeit there remains a basic query:

How to extrapolate stochastics from mice to man? Can the basic principles of animal studies or dosimetry be validated in clinical trials designated for human being?

There is a basic qualm on the main objective of a clinical trial: *symptomatic effects* versus true disease-disorder detection and cure.

Drug developers, pharmaceutical groups and regulatory agencies focus on treatments to relieve symptoms which may not totally or adequately match treatment objectives. Bioethics and public advocates have voiced concern on clinical trials in third-world countries, the affordibality of usual *high-cost drugs* being a major issue in this cost-benefit context. WHO and public health authorities all over the world are trying to identify effective and affordable regimens for many developing countries. These medical ethics, economic resources (affordability) and operational restraints often mar the routine use of standard statistical tools for drawing valid conclusions from clinical trials.

There are some basic differences between animal studies and clinical trials. The former can be conducted in a fairly controlled laboratory setups but human beings can not be put under such controlled environments, and as such, the enormous disparity in physical characteristics and many other epidemiologic endpoints, call for highly nonstandard statistical modeling and analysis. That is why *placebo-controlled trials* (PCT) are used extensively in development of new

pharmaceuticals. In the early phase of adoption of clinical trials, such PCTs were mostly advocated. However, there are allegations that PCT are invariably unethical when known effective therapy is available for the condition being treated or studied, regardless of the condition or the consequences of deferring treatments. The 1997 Helsinki Declaration by the World Medical Association (WMA) has clearly laid down the ethical principles for clinical trials: *In any medical study, every patient - including those of a control group, if any, should be assured of the best proven diagnostic and therapeutic methods.* Most often, in a PCT, this ethics is violated by the very composition of the placebo group. Based on this declaration, patients asked to participate in a PCT must be informed of the existence of any effective therapy, must be able to expore the consequences of deferring such therapy with the investigator, and must provide fully informed consent. *Active controlled equivalence trials* (ACET) have therefore been advocated for comparing an existing treatment with a targeted one. They may show whether a new therapy is superior (or inferior) to an existing one, but may not possess other characteristics

of PCTs (Temple and Ellenberg 2000, Sen 2001).

No matter it is a PCT or an ACET, there are numerous underlying constraints calling for novel *constrained statistical inference* (CSI) tools for statistical analysis. There is another feature common to both PCT and ACETs. It may be desirable in such a follow-up study to have *interim analysis* to monitor the accumulating clinical evidence in the light of statistical perspectives. While this feature has led to the evolution of *time-sequential* statistical methodology, there remains much to update this novel branch of CSI in the light of the underlying constraints and complications. It is usually desirable to look into the accumulating data sets at regular time intervals, and statistically deciding whether or not an *early termination* of the trial can be made in favor of the new therapy (if that is to be advocated in the drug market) so that patients can be switched to a better health perspective. Thus, usually, a *repeated signficance testing* (RST) scheme, often in a restrained setup, underlies statistical modeling and analysis of clinical trials. In conventional *group sequential tests* (GST) usually one assumes independent and homogeneous

increments for the associated stochastic processes. This is generally not the case in interim analysis related RST. *Progressively censoring schemes* (PCS) were introduced by Chatterjee and Sen (1973) to formulate the general methodology of *time-sequential* procedures; suitable martingale characterisations underlie most of these developments (Sen 1981, 1999a, 2001). With a need to update this approach in a more general framework to suit the ACET, let us consider the following statistical scenario.

Consider a typical constrained statistical interim analysis scheme relating to a comparative clinical trial relating to an existing therapy and a new one. The interim analysis related to monitoring of the accumulating evidence at time points $t_1 < \cdots < t_K$ for some specified K, spanning over a projected period of study $T = (0, t_K)$. If at an early time point t_k, there appears to be a significant difference (in favor of the new drug), then the trial is to be terminated at that point. The null hypothesis (H_0) relates to no difference over the entire period T while the alternative (H_1) to the new being better than the existing. We frame the null hypothesis H_{0r} that upto the time point t_r there is

no difference between the two therapies, and let H_{1r} be the alternative that for the first time, at time point t_r, there is a difference in favor of the new drug, for $r = 1, \ldots, K$. Then, restricted to the time domain T, we may note that there there is a nested nature of these hypotheses. The null hypothesis H_0 is accepted only when all the H_{0r} are accepted, while the alternative hypothesis H_1 is accepted when at least one of the K exclusive hypotheses $H_{1r}, 1 \leq r \leq K$ is accepted. Hence we write

$$H_0 = \cap_{r=1}^{K} H_{0r}, \quad H_1 = \cup_{r=1}^{K} H_{1r}. \tag{1}$$

Further, based on the accumulating data set upto the time point t_r, we construct a suitable test statistic $\mathcal{L}_r$ for testing H_{0r} vs $H_{1r},\ r = 1, \ldots, K$. This is essentially a RST problem in a constrained environment, and the natrure of the null and alternative hypotheses immediately calls for the *union intersection principle* (UIP). There is, however, some notable differences between the clinical trial and usual multiple hypothesis testing problems. The UIP having a finite intersection / union composition is more cumbersome to incorporate. Because of clinical and ethical undercurrents, first we appraise the potential con-

straints.

Restraint 1 : The component hypotheses are nested. For each $r(= 1, \ldots, K,\ H_{1r}$ is a one-sided alternative.

Restraint 2 : For different $r(= 1, \ldots, K)$, the different test statistics $\mathcal{L}_r$ are not independent, and the pattern of their dependence may not follow a Markov chain.

Restraint 3 : Early termination of the trial is associated with the acceptance of H_{1r}, for some $r < K$. It might be also due to significant adverse side-effects of the treatment, irrespective of the accumulating statistical evidence.

Restraint 4 : Explanatory variables provide useful statistical information, and hence, need to be included as far as possible, albeit increasing model complexity and CSI protocols.

Restraint 5 : Conventional (log-)linear regression models may not be appropriate. Some of the explanatory variables (viz., smoking, physical exercise, diabetic, etc.) may be binary, or at best, categorical. Even if they were quantitave, often for data recording,

they are are reported as categorical.

Restraint 6 : Informative censoring: Censoring due to noncompliance (e.g., drop-out or failure due to other causes) may not be independent of the placebo-treatment setup.

Restraint 7 : Surrogate end point: Often, the primary end point may be costly from data collection perspectives, and some closely related or associated (by symptoms, for example) variables, termed surrogate end points are used as substitute. The statistical model for the surrogate end point could be quite different from the primary one. Further, multiple end points may also crop up in such studies. Standard parametric multivariate CSI tools may not be usable properly.

Restraint 8 : Assessment of statistical quality of accumulating data with due respect to the underlying clinical and statistical restraints could be a major task.

Restraint 9 : Parametric models may not suit the purpose. Nonparametrics and semiparametrics may perform better. However,

the underlying restraints in semiparametrics may generally need critical appraisal. Nonparametrics may fare better but may require larger sample sizes to be of good quality and efficacy.

Restraint 10: Data mining : The advent of genomics is increasingly advocating for large number of end points and explanatory variables, and knowledge discovery and data mining (KDDM) tools are being advocated more and more. This does not, however, diminish the primary concern: To what extent statistical inference is not compromised or invalidated by data mining?

Suppose now that taking into account most of these restraints, albeit in approximate forms, it is possible to observe the partial data set $\mathcal{D}_t$ upto the time point t, so that $\mathcal{D}_t$ is nondecreasing (accumulating) in $t \in T$. Let $\mathcal{F}_t$ be the history process upto the time point t, so that $\mathcal{F}_t$ is nondecreasing in $t \in T$. Further, suppose that if all the (n)observations were available (i.e., the data set includes all responses and all explanatory variables), then for testing H_0 against a restricted alternative H_1, we would have a desirable test statistic

which we denote by $\mathcal{L}_n$. In a parametric setup, $\mathcal{L}_n$ could be a version of the likelihood ratio statistic or some of its variants like the partial-, penalized likelihood score etc. In semiparametrics, pseudo-, quasi-, or profile likelihood statistics might be usable. In nonparametrics, rank statistics have more appeal. We may set without any loss of generality $E\mathcal{L}_n|H_0) = 0,$ and dfefine

$$\mathcal{L}_n(t) = E_{H_0}\{\mathcal{L}_n \mid \mathcal{F}_t\}, t \geq 0. \tag{2}$$

Then, under fairly general regularity assumptions, we may note that under H_0,

$$\{\mathcal{L}_n(t), \mathcal{F}_t; t \geq 0\} \text{ is a zero mean martingale (array)} , \tag{3}$$

although this martingale characterization may not generally hold when the null hypothesis is not true. Also, even under the null hypothesis, $\mathcal{L}_n(t)$ may not have independent and stationary increments. Our task is to set a time sequential or RST procedure based on the discretized time-parameter process $\{\mathcal{L}_n(t_j),\ j \leq K\}$. Thus, we are confronted with suitable CSI procedures amenable to RST or interim analysis. Intuitively, we could conceive of an array of cut-off points:

$\{C_{nr}, r = 1l, \ldots, K\}$, **such that if** $\mathcal{L}_n(t_1) \geq C_{n1}$, **we stop the trial along with the rejection of** H_0; **if not, we go to the next time period** t_2 **and then if** $\mathcal{L}_n(t_2) \geq C_{n2}$, **we stop at that time along with the rejection of the null hypothesis. Otherwise we proceed to the next time period. In this way, the process continues, and if for the first time, for some** $k \leq K,\ \mathcal{L}_n(t_k) \geq C_{nk}$, **we reject the null hypothesis at that point and stop the trial. Thus, we proceed to accept the null hypothesis only when** $\mathcal{L}_n(t_j) < c_{nj}, \forall j \leq K$, **continuing the trial to its target time** t_K.

The basic problem is to control the overall Type I error rate without sacrificing much power in such interim analyses scheme. This, in turn, requires a skilfull choice of the cut-off points $C_{nr}, r \leq K$, **which generally depend not only on the** $t_k, k \leq K$ **but also on the accumulated statistical information at these points, and the latter is generally unknown or, at least, not properly estimable at the start of the trial. In this respect, we shall appraise the role of UIP along with other competitors. Group sequential tests (GST), formulated mostly in the late 1970's, make explicit use of normal distribution and equal increment assumptions which may not be generally true in**

such a time sequential setup. Even so, they needed extensive computation of the cut-off points. For some of these details, we refer to Sen (1999). Led by the basic weak convergence results for progressively censored linear rank statistics (Chatterjee and Sen 1973) some of these computational complexities have been eliminated considerably.

Typically, there exists a (random) time-parameter transformation by which the process $\{\mathcal{L}_n(t), t \in T\}$ can be written as $\mathbf{W}_{n,T} = \{W_{n,T}(u), u \in [0,1]\}$ such that under the null hypothesis, $\mathbf{W}_{n,T}$ converges weakly to a Brownian motion on $[0,1]$. By the same transformation, the calendar time points $t_r, r = 1, \ldots K$ are converted into (random) information time points $u_1 < \cdots < u_K$. Thus, we reduce the problem to a multivariate one-sided alternative hypothesis testing CSI problem for which the UIT sketched in earlier sections works out well. Basically, we have to construct the $W_{n,T}(u_r), r \geq 1$, and find a suitable cut-off point τ_{α^*} and a significance level α^* such that for a chosen α,

$$P\{W_{n,T}(u_r)/\sqrt{u_r} \ < \tau_{\alpha^*}, \forall r \ | H_0\} \leq \alpha. \tag{4}$$

Since a Brownian motion process $W(t), t \in [0,1]$ has irrerular behavior

with respect to the square root boundary as $t \to 0$, technically, we need that u_1 is away from 0. If the u_r are scattered over $(0, 1]$ and K is large, a more convenient way of computing the cut-off points would be to appeal to the boundary crossing probability of standard Brownian motion over one-sided square root boundaries; DeLong (1981) has provided detailed tables for these. This approximation is quite good when K is larger than 10, as is often the case of clinical trials with long-range follow-up time. Here also, the tabulated critical values correspond to some small truncation at 0 (i.e., over the range $[\epsilon, 1]$, for some positive ϵ (small). This weak invariance principle also avoids the need to specify the exact information times needed for the GST. There is an allied RST procedure considered by Chatterjee and Sen (1973) [and DeMet and Lan (1983)] where the weak convergence to Browinan mogtion has been incorporated in the utilization of (one-sided) linear boundaries [and a more general spending function approach]. For rank based procedures, often, for not so large samples, permutation tools provide scope for good approximations. The spirit of UIP is inherent in such interim analysis too.

With the advent of genomics and bioinformatics, in general, clinical trials are also encountering some challeging tasks. Instead of the conventional symptomatic effect approach, there is a new emphasis on pharmacogenomics dealing with the drug responses and the detection of disease genes along with the gene-environment interaction. Recalling that there may be thousands of genes which in a polygenic mode may not have individually significant impact but a large number of them in synergy may have significant (joint) impact, clinical trials are charged with not only finding the genes associated (causally or statistically) with a specific (group of) disease(s) but also their pharmacokinetics and pharmacodynamics with specific drug development. Instead of clinical trils with human subjects it calls for additional refinements: microarray and proteomics studies in clinical trials setup at the molecular level with tissues or cells. While this subject matter is beyond the scope of the present study, at least, it could be emphasized that becuase of enormous cost in conducting such trials, multi-center trials are needed for pooling small information from the individual centers and also multiple end points typically arise in such

composite studies. Typically, we encounter a matrix of statistics, individually from the centers and within each center, for the multiple end points. Although these centers may be treated as independent, the intra-center responses for the different end point are not. Confined to within center perspectives, typically, we have a vector valued stochastic process, and as before, we have a constrained environment. Therefore, even if we are able to construct a martingale array (in a multi-dimensional setup), formulating CSI procedures in a proper manner could be a formidable task. Bessel process approximations for multi-dimensional stochastic processes in clinical trials have been studied in the literature (viz., Sen (1981, Chapter 11)). There is a challenging task of incorporating such distributional approximations in the formulation of statistical inference procedures for restrained environments. The prospects for multivariate CSI analysis, displayed in detail in Silvapulle and Sen (2004) need to be appraised further. It is our belief that UPI, because of its flexibility and amenity to more complex models, would be most suitable in this context too.

We conclude with some pertinent remarks on the role of UPI in

meta analysis, as is currently adapted in multi-center clinical trials and genomic studies. Multi-center clinical trials, although, generally conducted under not so homogeneous environment (e.g., different geographical or demographic strata, age / culture differences), have a common objective of drawing statistical conclusions that pertain to a broader population. Consider in this vein, $C(\geq 2)$ centers, each one conducting a clinical trial with the common goal of comparing a new treatment with an existing one or a control or placebo. Since such centers pertain to patients with possibly different clutural, racial, demographic profiles, diet and physical exercise habits etc. and they may have somewhat different clinical norms too, the intra-center test statistics $\mathcal{L}_c,\ c = 1, \ldots, C$, used for CSI/RST, though could be statistically independent, might not be homogeneous enough to pull directly. This feature may thus create some impasses in combining these statistics values directly into a pooled one to enhance the statistical information. Meta analysis, based on observed significance levels (OSL) or p-values, is commonly advocated in this context. Recall that under the null hupothesis (which again can be interpreted as the

intersection of all the null hypotheses), the p-values have the common uniform $(0, 1)$ distribution, providing more flexibility to adopt UIP in meta analysis. Under restricted alternatives, these OSL values are left-tilted (when appropriate UIT are used) in the sense that the probability density is postively skewed over $(0, 1)$ with high density at the lower tail and low at the upper. Let us denote the p-values by

$$P_c = P\{\mathcal{L}_c \geq \textbf{the observed value } |H_0\}, c = 1, \ldots, C. \tag{5}$$

The well-known Fisher's test is based on the statistic

$$F_n = \sum_{c=1}^{C} \{-2 \log P_c\}, \tag{6}$$

which, under the null hypothesis, has the central chi-square distribution with $2C$ degrees of freedom. This test has some desirable asymptotic properties. There are many other tests based on the OSL values. The well known step-down procedure (Roy 1958) has also been adapted in this vein (cf. Subbaiah and Mudholkar 1980, Sen 1983), and they have been amended for CSI and RST as well (cf. Sen 1988). One technical drawback observed in this context is the insensitivity (to small to moderate deepartures from the null hypoth-

esis) of such tests (including the Fisher's) when C is large, resulting in nonrobust and, to a certain extent, inefficient procedure. Thus, alternative approaches based on the OSL values have been explored more recently in the literature.

In the evolving field of bioinformatics and genomics, generally, we encounter an excessively high dimensional data set with inadequate sample size to induce the applicability of standard CSI or even conventional statistical inference tools. On top of that, in genomics, the OSL values to be combined (corresponding to different genes) may not be independent, creating another layer of difficulty with conventional meta analysis. This led to the development of multiple hy[potheses testing in large dependent data models based on OSL values. This field is going through an evolution, and much remains to be accomplished. In this spectrum, the Simes (1986) theorem occupies a focal point. Let there be K null hypotheses (not necessarily independent) $H_{0k},\ k = 1, \ldots, K$ with respective alternatives (which possibly could be restricted or constrained as in clinical trials or microarray studies) $H_{1k},\ k = 1, \ldots, K$. We thus come across the same UIP scheme

by letting H_0 as the intersection of all the component null hypotheses, and H_1 as the union of the component alternatves. Let $P_k, k = 1, \ldots, K$ be the OSL values associated with the hypotheses testing H_{0k} vs. H_{1k}, for $k = 1, \ldots K$. We denote the ordered values of these OSL values by $P_{K:1}, \cdots, P_{K:K}$. If the individual tests have continuous null distributions then the ties among the P_k (and hence, among their ordered values) can be negelected, in probability. Assuming independence of the P_k, Simes theorem states that

$$P\{P_{K:k} > k\alpha/K, \forall k = 1, \ldots, K | H_0\} = 1 - \alpha. \tag{7}$$

Interestingly enough, the Simes theorem is a restatement of the classical Ballot theorem, developed some twenty years before (cf. Karlin 1969). In any case, it is a nice illustration how the UIP is linked to the extraction of extra statistical information through ordered OSL values.

It did not take long time for applied mathematical statisticians to make good uses of the Simes-Ballot theorem in CSI and multiple hypothesis testing problems. The above results pertains to tests for an

overall null hypothesis in the UIP setup. Amonmg others, Hochberg (1988) considered a variant of the above result:

$$P\{P_{K:j} \geq \alpha/(K-j+1), \forall j = 1, \ldots, K | H_0\} = 1 - \alpha, \tag{8}$$

and incorporated this result in a multiple testing framework. Benjamini and Hochberg (1995) introduced the concept of false discovery rate (FDR) in the context of multiple hypothesis testing, and illustrated the role of the Simes-Ballot theorem in that context. The past ten years have witnessed a phenomenal growth of research literature in this subfield with applications to genomics and bioinformatics. The basic restraint in this respect is the assumption of independence of the $P_j, j = 1, \ldots, K$, and in bioinformatics, this is hardly the case. Sarkar (1998) and Sarkar and Chang (1997) incorporated the MTP_2 (multivariate total positivity of order 2) property to relax the assumption of independence to a certain extent. Sarkar (2000, 2002, 2004) has added much more to this develiopment with special emphasis on controlling of FDR in some dependent cases. The literature is too large to cite adequately, but our primary emphasis here is to

stress how UIP underlies some of these developments and to focus on further potential work.

Combining OSL values, in whatsoever manner, may generally involve some loss of information when the individual tests are sufficiently structured to have coherence that should be preserved in the meta analysis. We have seen earlier how guided by the UIP, progressive censoring in clinical trials provided more efficient and interpretable testing procedures. The classical Cochran-Mantel-Haenszel (CMH) procedure is a very notable example of this line of attack. In a comparatively more general multiparameter CSI setting, Sen (1999b) has emphasized the use of the CMH procedure in conjunction with the OSL values to induce greater flexibility. The field is far from being saturated with applicable research methodology. The basic assumption of independence or specific type of dependence is just a part of the limitations. A more burning question is the curse of dimensionality in CSI problems. Typically, there K is large and the sample size n is small, i.e., $K >> n$. In the context of clinical trials in genomics setups, Sen (2006) has appraised this problem

with due emphasis on the UIP. Conventional test statistics (such as the classical LRT) have awkward distributional problems so that usual OSL values are hard to compute and implement in the contemplated CSI problems. Based on the Roy (1953) UIP but on some nonconventional statistics, it is shown that albeit there is some loss of statistical information due to the curse of dimensionality, there are suitable tests which can be implemented relatively easily in high-dimension low sample size environments. In CSI for clinical trials in the presence of genomics undercurrents, there is a tremendous scope for further develiopments along this line. Some recent work (Sen et al. 2007, JASA, Sen 2007-08: IMS Collection 2, and Kang and Sen 2008) have incorporated the Chen-Stein theorem extended to a sequential setup in rank statistics which seems to have a favorable performance picture.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discov-

ery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B57,*, 289-300.

Chatterjee, S.K. and Sen, P.K. (1973). Nonparametric testing under progressive censoring. *Calcutta Statist. Assoc. Bull. 22*, 13-50.

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

DeLong, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process. *Commun. Statist. A 10*, 2197-2213.

DeMets, D.L. and Lan, K.K.G. (1983). Discrete sequential boundaries for clinical trials. *Biometrika 70*, 659-663.

Hochberg, Y. (1988). A sharfer Bonferroni procedure for multiple tests of significance. *Biometrika 75*,800-802.

Karlin, S. (1969). *A first Course in Stochastic Processes*, Academic Press, New York.

Roy, J. (1958). Step-down procedures in multivariate analysis. *Ann. Math. Statist. 29*, 1177-1188.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.24*, 220-238.

Sarkar, S.K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture.*Ann. Statist.26*, 494-504.

Sarkar, S.K. (2000). A note on the monotonicity of the critical values of a step-up test. *J. Statist. Plann. Infer. 87*, 241-249.

Sarkar, S.K. (2002). Some results on false discovery rate in multiple testing procedures. *Ann. Statist. 30*, 239-257.

Sarkar, S.K. (2004). FDR-controlling stepwise procedures and their false negativesc rates. *J. Statist. Plann. Infer. 125*, 119-137.

Sarkar, S.K. and Chang, C.-K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc. 92*, 1601-1608.

Sen, P.K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference* John Wiley, New York.

Sen, P.K. (1983). A Fisherian detour of the step-down procedure. In *Contributions to Statistics: Essays in honour of Norman L. Johnson* North Holland, Amsterdam, pp. 367-377.

Sen, P.K. (1988). Combination of statistical tests for multivariate hy-

potheses against restricted alternatives. In *Advances in Multivariate Statistical Analysis* (eds. S.Dasgupta and J.K. Ghosh), Ind. Statist. Inst. pp. 377-402.

Sen, P. K. (1999a). Multiple comparisons in interim analysis. *J. Statist. Plann. Infer. 82*, 5-23.

Sen, P.K. (1999b). Some remarks on the Stein-type multiple tests of significance. *J. Statist. Plann. Infer. 82*, 139-145.

Sen, P. K. (2001). Survival analysis: Parametrics to semiparametrics to pharmacogenomics. *Brazilian J. Probab. Statist. 15*, 201 - 220.

Sen, P. K. (2006). Robust Statististical inference for high-dimension low sample size problems with applications to genomics. *Austrian J. Statist.* in press.

Silvapulle, M.J. and Sen, P. K. (2004). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions* John Wiley, New York.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*, 751-754.

Subbaiah, P. and Mudholkar, G. S. (1980). Testing significance of

a mean vector - a possible alternative to Hotelling T^2. *Ann. Inst. Statist. Math. 32*, 43-52.

Temple, R. and Ellenberg, S.S. (2000). Placebo-controlled trials and active-controlled trials in the evaluation of new treatments, I: Ethical and scientific issues. *Ann. Inter. Med. 133*, 455-463.

Tsai, M.-T. and Sen, P.K. (2005). Asymptotically optimal tests for parametric functions against ordered functional alternatives. *J. Multivar. Anal. 95*, 37-49.