COMBINATION OF STATISTICAL TESTS FOR MULTIVARIATE HYPOTHESES AGAINST RESTRICTED ALTERNATIVES

by

Pranab K. Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1494

December 1985

COMBINATION OF STATISTICAL TESTS FOR MULTIVARIATE HYPOTHESES AGAINST RESTRICTED ALTERNATIVES *

By PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

SUMMARY. Combination of (independent or dependent) tests for several multivariate hypotheses against restricted alternatives is considered. This includes a broad review of the classical Fisher method based on the observed significance levels (with necessary modifications), a variant form of the step-down procedure, a more plausible version of the Roy union-intersection principle and an extension of the well known Mantel-Haenszel procedure. The primary emphasis is placed on applications, and in view of that, various practical aspects of these diverse methods are discussed thoroughly. Nevertheless, the methodological aspects of this study rest on some theory which has not appeared in a systematic and unified form elsewhere, and hence, is included here. Connections of this theory to other related areas , such as, repeated significance tests and progressive censoring schemes, are also stressed.

1. INTRODUCTION

In a conventional setup, given $k(\geqslant 2)$ independent test statistics $T_1,\ldots,T_k$, for testing conformable null hypotheses $H_{01},\ldots,H_{0k}$, respectively (against suitable alternatives), a common problem is to pool the information from all the test statistics for testing the null hypothesis that all the H_{0i} are true . In this context, one may simply want to base such an overall test on the *observed significance levels (OSL)* of these component tests, or to utilize the T_i (or even the raw data sets from which the T_i were computed) in a more direct manner so as to obtain some optimal or desirable properties of the overall test. The OSL based

AMS Subject Classifications. 62F03, 62G10, 62H15

Key Words and Phrases: Combination of tests; Fisher's method; Mantel-Haenszel procedure; observed significance levels; progressive censoring schemes; repeated significance tests; restricted alternatives; step-down procedure; union-intersection principle.

*This work was supported partially by the Contract NO1-HV-12243-L from the U.S. National Institutes of Health.

tests work out very conveniently when the T_i are all independent. While there are quite a few overall tests belonging to this class, the classical one by Fisher (1932) is still in the lime light and possesses some (asymptotic) optimality properties too. For a nice review of OSL based combinations of independent tests, we may refer to Folks (1984) where detailed references to the relevant literature have also been cited. Bhattacharya (1961), Koziol and Perlman(1978) and Marden (1982), among others, considered the combination of independent chi square (or F-) statistics , while Monti and Sen(1976) considered locally optimal combinations of independent test statistics (of general forms). The OSL method as well as some of the other ones mentioned above work out well for the multivariate models. However, when testing against restricted alternatives, they may not be the most appropriate ones. The criterion of locally optimal combination of more appropriate test statistics appears to have more appeal in such a case.

Though there is some treatment of combination of non-independent tests in the literature [viz., Oosterhoff(1969)], there remains more room for methodological developments in this direction. For a variant form of the Roy(1958) *step-down procedure*, Mudholkar and Subbaiah (1980) worked out an OSL method based on the component Hotelling's T^2 statistics; a more general treatment of this procedure for general *multivariate analysis of variance (MANOVA) models* is due to Sen(1983). This extension of the Fisher (1932) OSL method works out conveniently for the incomplete multi-response models, growth curve models as well as for hierarchical designs. We will make more comments on it in a later section.

In multi-center clinical trials and many other (complex) experimental setups, one encounters the problem of combining tests for multivariate hypotheses where the alternative hypotheses are generally of some restricted types. For such restricted alternatives, the classical (parametric or nonparametric) test statistics may not be optimal, and some alternative test statistics having power superiority to the classical ones have been considered by various workers. Thus, from a practical point of view, it seems more appropriate to incorporate these alternative test statistics for the combination of tests. In this context, naturally, the OSL method may conveniently be applied, and we shall discuss such procedures in a later section. In the case where k, the number of tests to be combined, is large, the OSL method may not be that efficient (though for small values of k and large individual sample sizes, the Fisher (OSL) method has high Bahadur efficiency). For this reason, we shall consider some alternative procedures which generally behave better for moderate to large values of k. These methods to be considered are based on the combination of test statistics rather than their OSL. In this context, we shall exploit the Roy(1953) *union-intersection (UI-) principle* and also consider an extension of the Mantel-Haenszel (1963) procedure for some restricted alternatives.

In the context of multi-center clinical trials and other follow-up studies conducted in different centers, censoring is a natural phenomenan. For such a censored data set, the question of combination of tests also arises in a natural way, and we shall address to this question as well. In this context, it is not uncommon to have interim analysis of the accumulating data. Though , primarily, an interim analysis scheme eyes for a possible early termination of the follow-up study (particularly, when the null hypothesis is not true), in many epidemiological studies, such interim analyses are made without necessarily looking for any early curtailment of the study plan. In this context, both the concepts of *repeated significance tests* and *progressive censoring schemes* are relevant, and a substancial amount of research work has been accomplished in this area [viz. Sen(1985a) where other references are also cited]. Combination of repeated significance tests possibly arising in the context of progressive censoring schemes will also be considered here. In this context too, the Roy UI-principle plays a vital role.

2. OSL BASED COMBINATIONS OF INDEPENDENT TESTS

First, let us recapitulate briefly some existing tests. Suppose that $T_1, \ldots, T_k$ are $k (\geq 2)$ independent test statistics (for testing the null hypotheses $H_{01}, \ldots, H_{0k}$ respectively). We assume that these test statistics are all real valued and denote their sample realizations by $t_1, \ldots, t_k$, respectively. Then the *observed significance levels (OSL)* are defined by

$$p_i = P\{ T_i \geq t_i \mid H_{0i} \} , \text{ for } i = 1, \ldots, k. \tag{2.1}$$

Note that under the null hypotheses, these p_i are independent and identically distributed random variables (i.i.d.r.v.) each having the uniform (0,1) distribution , while the p_i are stochastically larger under suitable alternatives; this feature has been exploited in the combination of tests based exclusively on these OSL which would have then a completely known distribution under the overall null hypothesis H_0 that all the H_{0i} are true. Thus an OSL based combined test has a test statistic T_o^* of the form

$$T_o^* = \psi(p_1, \ldots, p_k) , \text{ for suitable real valued functions } \{\psi\}. \tag{2.2}$$

Birnbaum (1954) showed that if ψ is a monotone increasing function of each of its k arguments $(p_1, \ldots, p_k)$, then the test based on T_o^* is *admissible*. Among the various ψ , proposed in the literature, mention may be made of the following:

(i) The Fisher (1932) and Pearson (1933) method. Here

$$T_o^* = 2(\Sigma_{i=1}^k (-\log p_i)) , \tag{2.3}$$

so that under H_0 , T_o^* has the central chi-square distribution with 2k degrees of freedom, and moreover, ψ satisfies the Birnbaum-monotonicity condition. Littell and Folks (1971) studied the *Bahadur-efficiency* of the Fisher method and showed that whenever the individual sample sizes for the T_i are large (but, k is fixed), Fisher's test is optimal in the light of the Bahadur-efficiency. However, there are some concerns regarding the Bahadur-efficiency, and the Fisher method is not the only one to enjoy this asymptotic optimality property. Also, as k increases, the appropriateness of the Bahadur-efficiency criterion becomes dubious.

(ii) Tippett(1931) proposed the statistic $T_0^* = \min\{ p_i : 1 \le i \le k\}$, so that in this case, ψ reduces to the minimum of its arguments. This test is admissible. Wilkinson (1951) considered a generalization of Tippett's procedure: His statistic is the mth smallest value of the p_i, for some $m: 1 \le m \le k$.

(iii) Use of inverse distributional transformation: For a distribution function (d.f.) G , let $Y_i = G^{-1}(p_i)$, $i = 1,\ldots,k$. Then, for suitable G(.), one may use the Y_i in the construction of the overall test statistic. Lanchaster (1961) considered the case of chi squared (or Gamma) d.f. for G, Liptak(1958) considered the case of Normal G , and George and Mudholkar (1977) considered the case of logistic d.f. for G. These methods are generally inadmissible [c.f. Marden (1982)], although in the light of the Bahadur-efficiency, they are asymptotically efficient [c.f. Berk and Cohen (1979)].

There are other methods due to Good(1955) , Edgington (1972) and others, which will not be discussed here.

In (2.1) and (2.2), the test statistics T_i are of quite arbitrary form. They may be chosen on the ground of optimality against appropriate alternatives for the individual hypotheses. If we are interested in tests for multivariate hypotheses against restricted alternatives, and our problem is to combine several independent tests , then the OSL methods can be worked out very conveniently. All we have to do is to choose these T_i appropriately so that the corresponding tests have some optimality or desirability properties, and then proceed as in (2.1) and (2.2). However, the choice of appropriate T_i is very crucial in this context. For example, suppose that for each i (= 1,...,k) , the null hypothesis H_{0i} stands for a null mean vector of a multinormal distribution with unknown dispersion matrix $\underset{\sim}{\Sigma}$, and we have one sided alternatives that for each i , the mean vector is non-null and has nonnegative elements. In this case, for the individual samples, the Hotelling T^2-tests are not necessarily the best; tests having better power properties (and optimality in some sense) have been considered by Kudo(1963), Schaafsma(1968), Perlman(1969) and others. Thus, it seems quite natural to employ such alternative (and power-superior) test statistics in the OSL method for obtaining an overall test statistic. In fact, if we look at the Bahadur optimality of the Fisher OSL method, as has been studied by Littell and Folks (1971), we would notice that a pre-condition for this optimality is that the individual test statistics are optimal in some well defined manner. This basic condition may not be always true for testing multivariate hypotheses against restricted alternatives. In multivariate hypotheses testing, even in the case of global alternatives , there may not be a uniformly best test. Generally, the likelihood ratio tests have some asymptotic optimality properties characterized by the most stringency or the best average power (but restricted to the class of invariant tests), and in some specific direction (in the parameter space), some other test may have better power. Generally, the Lawley-Hotelling trace statistics behave similarly to the likelihood ratio test statistics, but Roy's (1953) largest root criterion is asymptotically non-equivalent to either of the others. Thus, it may not be the same picture if we use the likelihood ratio test statistics or the largest root criteria for the T_i in our combination procedure. The Bahadur optimality may be dependent on the direction of the alternatives, and uniformly, none may be the optimal in the light of this asymptotic efficiency criterion. The question is more pertinent in the case of restricted alternatives where the (asymptotic) optimality of the likelihood ratio test is not that universal. Indeed, it has been shown by Schaafsma (1968) that for such restricted alternatives, the likelihood ratio tests may not be most stringent or power-optimal in some other well defined manner. On the top of that the criterion of Bahadur-efficiency may not turn out to be very reasonable when k is large. For these reasons, while using the OSL method of combining independent test statistics for testing against restricted alternatives, proper care should be taken in the choice of the appropriate T_i as well as in the use of a proper measure of (asymptotic) efficiency. We do not recommend the use of likelihood ratio test statistics (for global or restricted alternatives) for the T_i when the alternative hypotheses are of restricted nature. In the nonparametric case, from the point of view of Pitman-efficiency, the T_i are generally chosen in such a manner that for some specific (parametric type) alternatives, they are asymptotically equivalent to the corresponding likelihood ratio test statistics (viz., Normal scores statistics for the two-sample location problem when the undelying distributions are ideally sought to be Normal), though they remain valid for any departure from this ideal situation and remain robust in efficiency too. In view of the shortcomings of the likelihood ratio statistics, discussed before, here also, in the context of combining independent nonparametric tests against restricted alternatives, we should choose alternative nonparametric statistics for the T_i having asymptotic power superiority to the nonparametric analogues of the usual likelihood ratio statistics. Roy's(1953) UI-principle may be used in this context to locate such alternative test statistics, and in a later section, we shall elaborate on this issue.

3. OSL BASED COMBINATIONS FOR DEPENDENT TESTS

For general multivariate hypotheses against restricted alternatives, the classical likelihood ratio test statistics (or some other allied ones) may not be optimal, and moreover, their exact distribution theory is generally quite complicated. Further, unlike the univariate case, in a general multivariate model, restriction to invariant tests may not yield an unique best test among them (even when testing against global alternatives): This leaves some room for other ad hoc procedures, and among these, Roy's(1953) UI-tests are often advocated. The *step-down procedure*, proposed by J. Roy(1958), is a tributary of the general UI-procedure of Roy (1953) in the sense that only a finite number of component hypotheses is involved in the formulation of the UI-principle. This procedure can be adapted to combine dependent tests either with the OSL method or with the usual combination of statistics method based on the quasi-independence (under the null hypothesis) of the test statistics at the successive steps.

According to this procedure, the overall null hypothesis H_0 is viewed as a finite intersection of $m(\geq 1)$ component hypotheses $H_{01},\ldots,H_{0m}$ (i.e., $H_0 = \bigcap_{j=1}^{m} H_{0j}$) and the alternative H as the union of component alternatives $H_1,\ldots,H_m$ (i.e., $H = \bigcup_{j=1}^{m} H_j$) , where the H_j themselves may be of restricted form. The overall hypothesis H_0 is then tested in steps: Testing for the jth component is made, if and only if, in all the preceding steps, the component null hypotheses were accepted, $1 \leq j \leq m$, so that the overall null hypothesis is accepted only when all the component null hypotheses are accepted. Step-down procedures are generally not invariant ones, and the ordering of the steps (or related component hypotheses) may sometimes be advocated on the ground of their relative importance (and, otherwise, in an arbitrary manner). In case of testing for H_0 against some restricted alternative H, often, a formulation of the component hypotheses H_{0j} and H_j , $j=1,\ldots,m$, can be so made that the H_j relate to some convenient one sided alternatives, for which some optimal tests exist, and this makes the step-down procedure amenable for such restricted alternative problems. For example, in the combination of independent tests for the normal mean vector against the positive orthant alternatives, as has been posed in the earlier section, we may choose $m = p$, the dimension of the mean vector, and at each step, we may use a combined test statistic for one sided alternatives (resulting in one-sided Student's t-tests instead of the F-tests or two-sided t-tests). In the step-down procedure, when the null hypothesis H_0 holds, the component hypotheses test statistics (say, $T_1^c,\ldots,T_m^c$) are quasi-independent, so that if we choose the levels of significance for the component tests as $\alpha_1,\ldots,\alpha_m$, and set for the overall level of significance α $(0 < \alpha < 1)$,

$$1 - \alpha = (1 - \alpha_1)\ldots(1 - \alpha_m) , \tag{3.1}$$

then the components can be tested in steps, leading to the overall significance level α . Note that each α_i is smaller than α , and the α_j are determined suitably either according to some index of preference for the m components or by the equality criterion : $\alpha_1 = \ldots = \alpha_m = 1 - (1 - \alpha)^{1/m}$.

Let us make some specific comments on the scope of applicability of the step down procedure, as has been described before. For simplicity of presentation, we again refer to the combination of independent tests for the normal nean vector when the alternative hypotheses relate to the positive orthant of the p-dimensional Euclidean space. If we use the step down procedure discussed before, at each step, we need to combine independent statistics and we have an one-sided alternative. If all the dispersion matrices are assumed to be the same, then this can be done very conveniently. On the other hand, as may be the usual case, these dispersion matrices may not be the same, and an optimal combination of independent test statistics at each step will involve a Behrens-Fisher problem. To eliminate this difficulty, it may be advisable to use the OSL for each source statistics at each step, where the OSL are computed with reference to the restricted alternative situations in mind. Then, at each step, one may combine these OSL in the Fisher (1932) method , or in some other convenient manner , and then use the step-down procedure on these step-wise OSL based statistics (which will still be quasi-independent under H_0) . This modification can be done in a more general setup too. We refer to this as a *quasi-step down combination procedure*. In terms of the scope of applicability in practice, this modified procedure rests on comparatively less stringent assumptions on the individual source data sets, and hence, should be preferred to.

Both the step down and quasi-step down procedures require that the individual levels of significance, $\alpha_1 ,\ldots, \alpha_m$, for the component hypotheses, be fixed in advance , satisfying (3.1). As has been explained earlier, there remains some arbitrariness in the choice of $\alpha_1,\ldots,\alpha_m$, and this may be eliminated by the following alternative version of a step-down procedure, which we term a *Fisherian detour of the step down procedure*. For the Hotelling T^2-statistics, this version of the step-down procedure was considered by Mudholkar and Subbaiah (1980), and a detailed treatment of this methodology for general multivariate linear models is due to Sen (1983). Keeping in mind that we have $k(\geq 1)$ independent data sets, and for each data set, we have a multivariate null hypothesis against a restricted alternative, we may formulate the hypotheses as

$$H_0 = \bigcap_{i=1}^{k} \bigcap_{j=1}^{m} H_{0ij} \quad \text{vs} \quad H = \bigcup_{i=1}^{k} \bigcup_{j=1}^{m} H_{ij} \,, \tag{3.2}$$

where, for each $i(=1,\ldots,k)$, the H_{ij}, $j=1,\ldots,m$ relate to the reformulation of the alternative hypotheses in the usual setup of a step down procedure. For the ith source, bet $\mathcal{B}_{ij}$ the sigma-field generated by the random elements entering the first j steps, for $j=1,\ldots,m$, and let $\mathcal{B}_{i0}$ be the trivial sigma-field, for $i=1,\ldots,k$. Let T^c_{ij} be an appropriate (conditional) test statistic for testing H_{0ij} vs H_{ij} (given $\mathcal{B}_{ij-1}$), for $j=1,\ldots,m$ and $i=1,\ldots,k$. Let then

$$G_{ij}(t) = P\{ T^c_{ij} \le t \mid \mathcal{B}_{ij-1}, H_{0ij} \} \,,\ 1 \le j \le m;\ 1 \le i \le m \,, \tag{3.3}$$

and let p_{ij} be the OSL computed by using (3.3) on the observed values of the T^c_{ij}. Note that for restricted alternatives, the choice of these T^c_{ij} may be made in a more convenient manner in the setup of a step down procedure than in a multivariate case. Define then

$$U_{ij} = -2\log p_{ij} \,,\ j=1,\ldots,m;\ i=1,\ldots,k. \tag{3.4}$$

$$U^*_{ij} = U_{i1} + \ldots + U_{ij} \,,\ 1 \le j \le m \,;\ 1 \le i \le k \,; \tag{3.5}$$

$$U^o_j = U^*_{1j} + \ldots + U^*_{kj} \,,\ \text{for } j = 1,\ldots,m. \tag{3.6}$$

Note that,proceeding as in Sen (1983), under H_0, U^o_m has the central chi square distribution with 2km degrees of freedom, and we denote the upper $100\alpha\%$ point of this distribution by $\chi^2_{2km,\alpha}$. Further, note that the U_{ij} are nonnegative random variables, and hence, U^o_j is nondecreasing in j: $1 \le j \le m$. Thus, as a natural extension of Sen (1983), we propose the following procedure:

At the first step, compute U^o_1 (through the U_{i1}). If $U^o_1 \ge \chi^2_{2km,\alpha}$, stop at this stage along with the rejection of H_0, while, if $U^o_1 < \chi^2_{2km,\alpha}$, proceed to the second step. At the jth step (arising only when $U^o_\ell < \chi^2_{2km,\alpha}$, for every $\ell < j$), if $U^o_j \ge \chi^2_{2km,\alpha}$, the process terminates along with the rejection of H_0, while for $U^o_j < \chi^2_{2km,\alpha}$, one proceeds to the next step, for $j=1,\ldots,m$; for $j = m$, of course, the process terminates in favor of the null hypothesis H_0 when $U^o_m < \chi^2_{2km,\alpha}$. Thus, H_0 is accepted only when $U^o_m < \chi^2_{2km,\alpha}$.

When testing against restricted alternatives, often, the H_{ij} relate to some one-sided alternatives, for which the p_{ij} can be computed with considerable simplicity. Moreover, these p_{ij} when computed from one tail of the null distribution of the component statistics will be smaller than the corresponding values obtained by using both the tails, and hence, the U_{ij} in (3.4) would be larger than their counterparts computed from the two-sided tests, and this would make the U_j^o larger too. This explains why in this method of combination, for restricted alternatives, often, we would expect a better power by using one-sided tests whenever possible. For the usual two-sided tests for the T_{ij}^c , the Bahadur-efficiency of this proposed method can be computed as in Sen (1983), and it follows that it will enjoy the same Bahadur-optimality along with the classical Fisher (1932) OSL method based on the combination of k independent OSL from the different data sets. However, we may also note that for chi square distributions, the Bahadur-slopes are insensitive to the allied degrees of freedom, and hence, this Bahadur-optimality should be interpreted with some caution.

4. COMBINATION OF TEST STATISTICS: UI-PRINCIPLE.

In the context of combining independent tests for ANOVA (or MANOVA), Monti and Sen (1976) considered a linear combination of the individual test statistics (or some monotone transforms of them) with the coefficients chosen in such a way that the overall test has locally optimal properties. Earlier, Bhattacharya (1961) considered some Monte Carlo studies of overall tests based on the sum of independent chi-square statistics. Koziol and Perlman (1978) studied in detail the combination of independent chi square tests, while Marden and Perlman (1982) and Marden (1982) considered the minimal complete class of procedures when combining independent non-central F- and chi square tests. When combining independent chi square statistics with degrees of freedom at least equal to 2, Koziol and Perlman (1978) have observed that the Fisher OSL method is 'relatively minimax' within a broad class of tests. Deeper results in this direction are contained in Marden (1982).

We may recall that in our study $H_0 = \bigcap_{i=1}^{k} H_{0i}$ and $H = \bigcup_{i=1}^{k} H_i$, where the H_i are generally of the restrictive type. For testing H_{0i} vs. H_i , an optimal (parametric) test statistic is not generally an F-statistic or a chi square one (or a monotone transform of either one). As such, while considering (convex) linear combinations of individual source optimal test statistics, we encounter the problem of other statistics which may not have the reproductive property (in the sense that their sums may not have the same type of distribution with the degrees of freedom added up). For example, for combining k independent tests for the normal mean vector against orthant alternatives, the individual source test statistics have the Beta-bar distributions [c.f. Barlow et al.(1972,p.181)]; these distributions are convex mixtures of several Beta distributions with varying degrees of freedom, and their convolution is not a member of the same class. Thus, we may use a linear combination of these optimal test statistics (for restricted alternatives) to construct an overall test, but that may not have a simple distribution or may not have the admissibility property. However, from the practical point of view, it makes a lot of sense to use an appropriate linear combination of the individual source optimal test statistics as an overall test statistic, and as in Monti and Sen (1976) to choose the coefficients in such a way that this overall test is locally optimal .

In this context, the Roy (1953) UI-principle may also be used to formulate alternative combination procedures in a very natural way. Suppose that it is possible to express,for each $i(=1,\ldots,k)$,

$$H_{0i} = \bigcap_{j\varepsilon\Gamma} H_{0ij} \quad \text{and} \quad H_i = \bigcup_{j\varepsilon\Gamma} H_{ij} , \tag{4.1}$$

where the subscript j belongs to an index set (that might even be a part of the continuum) and for each $j(\varepsilon\ \Gamma)$, (H_{0ij}, H_{ij}) form a natural pair of testable hypotheses. We write then

$$H^*_{0j} = \bigcap_{i=1}^{k} H_{0ij} \quad \text{and} \quad H^*_j = \bigcup_{i=1}^{k} H_{ij} , \text{ for } j \ \varepsilon\ \Gamma. \tag{4.2}$$

For testing H^*_{0j} against H^*_j , an optimal combined test may exist under general regularity conditions. At least, the admissibility of such a combined test may be established along the lines of Koziol and Perlman(1978) and Marden (1982). Also, in most of the cases, such a combined test statistic can be expressed as a linear combination of independent normal, beta or chi square variables. As such , through proper normalizations, we may consider some overall test statistic,say, T^*_{0j} , for every $j \in \Gamma$, with the property that at some level of significance α^* $(0 < \alpha^* < 1)$, $T^*_{0j} \geq C^* = C^*(\alpha^*)$ gives the size α^* critical region. Then, in the usual fashion of the UI-principle , we may set

$$T^* = \sup\{ T^*_{0j} : j \in \Gamma \} \quad \text{and} \quad \alpha = P\{ T^* \geq C^* \mid H_0\}. \tag{4.3}$$

Then T^* is the UI-test statistic with the overall significance level α . The problem is to choose α^* in such a way that α is predetermined.

To illustrate this procedure, we consider again the multinormal mean vector where the alternative hypotheses relate to the positive orthant against the null hypothesis that it is equal to $\underset{\sim}{0}$. In this setup, we may take Γ as the set of all p-vectors $\underset{\sim}{a} = (a_1,\ldots,a_p)$, such that $a_i \geq 0$, for every $i(=1,\ldots,p)$ and the norm of $\underset{\sim}{a}$ is unity (viz., $\underset{\sim}{a}'\underset{\sim}{a} = 1$). For every $\underset{\sim}{a}$, denoting by $\underset{\sim}{X}_{ij}$, $j \geq 1$ the observation vectors in the ith source, $\underset{\sim}{a}'\underset{\sim}{X}_{ij}$ has the univariate normal distribution with null mean (under H^*_{0a}) , one can construct an one-sided combined test based on a linear combination of Student's t-statistics (when the dispersion matrices are not assumed to be the same) or a Student t-statistic (when all the dispersion matrices are assumed to be the same). In either case, T^* can be obtained by maximizing these $T^*_{0\underset{\sim}{a}}$ over $\underset{\sim}{a} \in \Gamma$, and closed expression for this is available in the literature. In general, the solution obtained by this manner is not the same as the optimal linear combination of the individual source optimal test statistics (when testing against restricted alternatives). Moreover, the computation of T^* in (4.3) requires in turn the computation of the T^*_{0j} which, in general, demands the knowledge of some other sample statistics besides the individual test statistics (viz., the sample covariance matrices in addition to the Hotelling T^2 statistics in the example cited above). However, this should not pose a serious problem, because in most of the cases, these additional statistics are provided by the summary analysis of the individual source data sets. In passing, we may remark that if the set Γ in (4.2) has a finite number of elements, then instead of (4.3) one may also use a version of the step-down procedure, outlined in Section 3, so that the results in Section 3 may also be interpreted in the light of the UI-principle. Further, instead of the decomposition in (4.2), if we use some optimal test statistics for testing the individual H_{0i} against H_i (which may be of restricted type), then we may also use the UI-principle in the setup of this finite partitioning . In fact, if we use the maximum of these individual source test statistics, we have a procedure equivalent to the classical Tippett (1931) procedure based on the minimum OSL (and is known to be admissible under a very general setup). This explains how the OSL methods in Section 2 and the variants of the step-down procedures in Section 3 can all be interpreted in the light of Roy's(1953) UI-principle.

In the context of (multivariate) ANOVA and other models related to blocked designs , because of the block-effects and other possible nuisance parameters, the raw data sets from the different blocks may not be totally compatible. However, once these nuisance parameters are eliminated through proper transformations on the original data sets, the reduced data sets are very much comparable. In such a case, in testing for H_0 against a restricted type of alternative H , it may be of distinct advantage in using the UI-principle on the pooled reduced data set. In such a case, the distribution theory of the overall test statistic is generally much simpler than in the case of (4.3) where homogeneity of the different (reduced) data sets needs not be a part of the assumption. For example, in testing for the equality of treatment effects against an ordered alternative (for a randomized block design replicated in different locations or on different years), in the setup in (4.3), it is not necessary to assume the homogeneity of the error distributions from one location (year) to another, although such an assumption may not be that unreasonable. If we make this assumption then we would have an overall

UI-test (for the normal theory model) based on the so called Beta-bar distribution. On the other hand, without this homogeneity assumption, in the setup of (4.3), we would end up with a convolution of several of these Beta-bar distributions or in some other more complicated forms. Naturally, when this homogeneity assumption can be made with confidence, we would prescribe the UI-test based on the pooled reduced data set. Again, it may be noted that such an overall test is not a combination of the individual test statistics or their OSL , but an overall test based on the combined (reduced) data sets.

5. MANTEL-HAENSZEL PROCEDURE: RAMIFICATIONS FOR RESTRICTED ALTERNATIVES

When combining test statistics from several multi-dimensional contingency tables, a cumulative chi square type test statistic can readily be obtained by adding over the individual sample test statistics, and this may fare quite well relative to the OSL based tests. In many cases, the homogeneity (in a broad sense) of the different contingency tables may be taken for granted, and, as a result, the Cochran-Mantel-Haenszel (CMH) procedure may be used with certain advantages. A comparatively more general and unified formulation of such CMH procedures has been considered by Landis, Heyman and Koch (1978). In the context of this investigation, we are primarily interested in combining tests against restricted alternatives, and in view of that, we shall consider some ramifications of the generalized CMH procedures for certain discrete multivariate observations.

Let us consider $k(\geq 1)$ independent $(s \times r)$ contingency tables, where for each i $(=1,\ldots,k)$, the observed frequencies $n_{ij\ell}$ in the contingency table may be represented as :

sub-population	Response variable categories 1	2	...	r	Total
1	n_{i11}	n_{i12}	...	n_{i1r}	$n_{i1.}$
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
s	n_{is1}	n_{is2}	...	n_{isr}	$n_{is.}$
Total	$n_{i.1}$	$n_{i.2}$	...	$n_{i.r}$	$n_{i..}$

In many cases, these contingency tables correspond to disctinct levels of a covariable or combinations of several concomitant variables, and the null hypothesis H_0 can be written as $H_0 = \bigcap_{i=1}^{k} H_{0i}$, where H_{0i} relates to the homogeneity of the s sub-populations with respect to the response variable for the ith table, for i=1,...,k. As in the earlier sections, in testing for this homogeneity , we may be interested in some restricted alternatives, so that we may write $H = \bigcup_{i=1}^{k} H_i$, where the H_i are of some restricted types (as will be more formally considered later on). As an example, we may consider the simple case where the response categories 1,...,r are ordered in a well defined sense, and we may have an ordered alternative relating to the stochastic ordering of the s distributions for the sub-populations with respect to these response categories.

Note that under H_0 (and conditional on the marginals $n_{ij.}$ and $n_{i\ell}$, given) the joint probability function of the $n_{ij\ell}$ is given by

$$\prod_{i=1}^{k} \{ \prod_{j=1}^{s} n_{ij.}! \prod_{\ell=1}^{r} n_{i.\ell}! \,/\, n_{i..}! \prod_{j=1}^{s} \prod_{\ell=1}^{r} n_{ij\ell}! \} . \tag{5.1}$$

As such , a conditionally distribution-free overall test statistic for testing H_0 (against global alternatives) is given by

$$\mathcal{L}_n^* = \mathcal{L}_n^{*(1)} + \dots + \mathcal{L}_n^{*(k)} , \tag{5.2}$$

where

$$\mathcal{L}_n^{*(i)} = \Sigma_{j=1}^{s} \Sigma_{\ell=1}^{r} (n_{ij\ell} - n_{ij.}n_{i.\ell}/n_{i..})^2(n_{i..}-1)/n_{ij.}n_{i.\ell} , \tag{5.3}$$

for i = 1,...,k. Note that under H_0 , for large sample sizes, each $\mathcal{L}_n^{*(i)}$ has closely the central chi square distribution with (s-1)(r-1) DF , so that $\mathcal{L}_n^*$ has asymptotically the central chi square distribution with k(s-1)(r-1) DF. Though this test is consistent against a broad class of alternatives, as an overall test (assuming some sort of homogeneity accross the k tables), it may not perform that well (mainly because of the k fold increase in the DF), and a better test is based on the Cochran-Mantel-Haenszel approach sketched below. Let us introduce the notations :

$$\underset{\sim}{n}_i = (n_{i11},\ldots,n_{i1r},\ldots,n_{is1},\ldots,n_{isr})' \ , \ i = 1,\ldots,k; \tag{5.3}$$

$$\underset{\sim}{m}_i = (m_{i11},\ldots,m_{isr})' \text{ with } m_{ij\ell} = n_{ij.}n_{i.\ell}/n_{i..} \ , \ 1\le j\le s, 1\le \ell\le r, 1\le i\le k, \tag{5.4}$$

$$v^{(i)}_{j\ell,j'\ell'} = n_{ij.}n_{i.\ell}(\delta_{jj'}n_{i..} - n_{ij'.})(\delta_{\ell\ell'}n_{i..} - n_{i.\ell'})/n^2_{i..}(n_{i..}-1), \tag{5.5}$$

for $j,j' = 1,\ldots,s;\ \ell,\ \ell' = 1,\ldots,r$; $\underset{\sim}{V}^{(i)} = ((\ v^{(i)}_{j\ell,j'\ell'}))$, for $i=1,\ldots,k$, where δ_{ab} stands for the Kronecker delta . Note that the conditional mean of $\underset{\sim}{n}_i$ (under H_0) is $\underset{\sim}{m}_i$, $i=1,\ldots,k$, and we denote by

$$\underset{\sim}{W}_i = (\underset{\sim}{n}_i - \underset{\sim}{m}_i\ ),\ i=1,\ldots,k\ ,\ \underset{\sim}{W}^* = \underset{\sim}{W}_1 + \ldots + \underset{\sim}{W}_k \ ; \tag{5.6}$$

$$\underset{\sim}{V}^* = \underset{\sim}{V}_1 + \ldots + \underset{\sim}{V}_k \ . \tag{5.7}$$

Then, the CMH test statistic can be expressed as

$$\mathcal{L}^{o}_{n} = (\underset{\sim}{W}^*)'(\underset{\sim}{V}^*)^-(\underset{\sim}{W}^*) \ , \tag{5.8}$$

where $\underset{\sim}{V}^{*-}$ is a generalized inverse of $\underset{\sim}{V}^*$ (of rank $(s-1)(r-1)$). Here also, the conditional distribution of $\mathcal{L}^o_n$ (given the marginal frequencies and under H_0) can be incorporated to construct a conditionally distribution-free test based on $\mathcal{L}^o_n$. Also, for large sample sizes, under H_0 , $\mathcal{L}^o_n$ has closely the central chi square distribution with $(s-1)(r-1)$ DF (compared to $k(s-1)(r-1)$ for $\mathcal{L}^*_n$), and this reduced DF often makes $\mathcal{L}^o_n$ more powerful than $\mathcal{L}^*_n$ when homogeneity accross the k tables may be taken for granted. [However, when this assumed homogeneity of the pattern accross the k tables is vitiated, the CMH procedure may perform rather poorly relative to the omnibus test in (5.3), and hence, the CMH procedure should be used with some cautions.]

We may note that (5.8) is basically designed for an overall test against a general alternative. In the context of restricted alternatives, it seems quite possible to modify this procedure to yield even more powerful tests. In this context, instead of using the classical quadratic form for $\underset{\sim}{W}^*$ with the generalized inverse of $\underset{\sim}{V}^*$ as its discriminant, we may use the UI-principle (discussed in Section 4) to construct alternative test statistics. While the particular form of such an UI-CMH test statistic will depend specifically on the decomposition of the null and alternative hypotheses into appropriate component hypotheses, in many situations, these UI-test statistics can be expressed explicitly in terms of suitable quadratic forms in the $\underset{\sim}{W}_i$ (or appropriate transforms on them). For example, when the response categories 1,...,r are based on an implicit ordering and we are interested in testing for the homogeneity of the s sub-populations against shift alternatives , it may be quite appropriate to conceive of a vector $\underset{\sim}{a} = (a_1,\ldots,a_r)'$ of appropriate scores (which may be data adaptive too) and to define

$$\underset{\sim}{Z}_i = \underset{\sim}{A}\underset{\sim}{W}_i \ , \ i=1,\ldots,k; \quad \underset{\sim}{Z}^* = \underset{\sim}{A}\underset{\sim}{W}^* = \underset{\sim}{A}(\underset{\sim}{W}_1+\ldots+\underset{\sim}{W}_k) = \underset{\sim}{Z}_1 +\ldots+ \underset{\sim}{Z}_k \ , \tag{5.9}$$

where $\underset{\sim}{A}$ is a s ×sr matrix of the form:

$$\begin{pmatrix} \underset{\sim}{a}' & \underset{\sim}{0}' & \cdots & \underset{\sim}{0}' \\ \underset{\sim}{0}' & \underset{\sim}{a}' & \cdots & \underset{\sim}{0}' \\ \vdots & \vdots & \cdots & \vdots \\ \underset{\sim}{0}' & \underset{\sim}{0}' & \cdots & \underset{\sim}{a}' \end{pmatrix}$$

Let then

$$\underset{\sim}{V}^{o*} = \underset{\sim}{A}\underset{\sim}{V}^*\underset{\sim}{A}' = \underset{\sim}{A}(\underset{\sim}{V}_1+\ldots+\underset{\sim}{V}_k)\underset{\sim}{A}' \ . \tag{5.10}$$

Then, parallel to (5.8), one may consider here a CMH test statistic of the form

$$\mathcal{L}_n^{o*} = (\underset{\sim}{Z}^*)'(\underset{\sim}{V}^{o*})^-(\underset{\sim}{Z}^*) \ , \tag{5.11}$$

where $\underset{\sim}{V}^{o*-}$ is a generalized inverse of $\underset{\sim}{V}^{o*}$; then $\mathcal{L}_n^{o*}$ is a conditionally distribution-free statistic having asymptotically (under H_0) a central chi square d.f. with (s-1) DF. This modified CMH (score) test statistic is degined for the set of alternatives that the s sub-populations differ with respect to the average response variable , and for such alternatives, it would perform better than the original version in (5.8) [and this is mainly due to the reduction of the DF's from (s-1)(r-1) to s-1 without making the noncentrality parameter too small to compensate for this reduction in the DF.] However, the zone of consistency of this modified CMH test is more restricted than the zone for the original test in (5.8). Thus, there are other alternatives for which the test in (5.8) may perform much better than the one in (5.11). The statistic in (5.11) may be interpreted in the light of the Roy(1953) UI-principle. However, to illustrate such UI-tests based on the CMH approach, we may consider the case of 'ordered alternatives' where we want to test for the homogeneity of the s sub-populations against a stochastic ordering of them (with respect to the implicit ordering of the response variables). In such a case, on the vector $\underset{\sim}{Z}^*$ in (5.9) and the associated matrix $\underset{\sim}{V}^{0*}$ in (5.10), we may readily incorporate Roy's UI-principle and get the resulting test statistics. Towards this consider the s-vector $\underset{\sim}{b} = (b_1,\ldots,b_s)'$ and the space $B = \{ \underset{\sim}{b} : b_1 \leq \ldots \leq b_s\}$. For every $\underset{\sim}{b}$ in B , let

$$L_n^*(\underset{\sim}{b}) = (\underset{\sim}{b}'\underset{\sim}{Z}^*)/(\underset{\sim}{b}'\underset{\sim}{V}^{0*}\underset{\sim}{b})^{\frac{1}{2}} \text{ and let } L_n^{**} = \sup\{ L_n^*(\underset{\sim}{b}) : \underset{\sim}{b} \in B \}. \tag{5.12}$$

Then the UI-test statistic is L_n^{**} . This can be expressed in an explicit form as follows. Let $\underset{\sim}{Q}$ be a (s-1) ×s matrix with the elements

$$\begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}.$$

Let then

$$\underset{\sim}{Z}^{**} = \underset{\sim}{Q}\underset{\sim}{Z}^* \quad \text{and} \quad \underset{\sim}{V}^{**} = \underset{\sim}{Q}\underset{\sim}{V}^{0*}\underset{\sim}{Q}' . \tag{5.13}$$

Also, let m be any subset of $M = \{1,\ldots,s-1\}$, and $\bar{m}$ be the complementary subset. For each of the 2^{s-1} subsets $\{m : \emptyset \subseteq m \subseteq M \}$, $\underset{\sim}{Z}^{**}$ and $\underset{\sim}{V}^{**}$ are partitioned (following rearrangements, if needed) as

$$\underset{\sim}{Z}^{**'} = (\underset{\sim}{Z}^{**'}_{(m)}, \underset{\sim}{Z}^{**'}_{(\bar{m})}) \quad \text{and} \quad \underset{\sim}{V}^{**} = \begin{pmatrix} \underset{\sim}{V}^{**}_{(mm)} & \underset{\sim}{V}^{**}_{(m\bar{m})} \\ \underset{\sim}{V}^{**}_{(\bar{m}m)} & \underset{\sim}{V}^{**}_{(\bar{m}\bar{m})} \end{pmatrix}. \tag{5.14}$$

Also, let

$$\hat{\underset{\sim}{Z}}^{**}_{(m)} = \underset{\sim}{Z}^{**}_{(m)} - \underset{\sim}{V}^{**}_{(m\bar{m})}\underset{\sim}{V}^{**-1}_{(\bar{m}\bar{m})}\underset{\sim}{Z}^{**}_{(\bar{m})} , \tag{5.15}$$

$$\hat{\underset{\sim}{V}}_{(mm)} = \underset{\sim}{V}^{**}_{(mm)} - \underset{\sim}{V}^{**}_{(m\bar{m})}\underset{\sim}{V}^{**-1}_{(\bar{m}\bar{m})}\underset{\sim}{V}^{**}_{(\bar{m}m)} . \tag{5.16}$$

Then, following the lines of Sen (1982), L_n^{**} can be shown to be given by

$$L_n^{**} = \Sigma_{\{\emptyset \subseteq m \subseteq M\}} \{(\hat{\underset{\sim}{Z}}^{**'}_{(m)}\hat{\underset{\sim}{V}}^{**-1}_{(mm)}\underset{\sim}{Z}^{**}_{(m)})^{\frac{1}{2}} I(\hat{\underset{\sim}{Z}}^{**}_{(m)} \geq \underset{\sim}{0}) I(\underset{\sim}{V}^{**-1}_{(\bar{m}\bar{m})}\underset{\sim}{Z}^{**}_{(\bar{m})} < \underset{\sim}{0})\} , \tag{5.17}$$

where I(A) stands for the indicator function of the set A. Note that in (5.17), both the indicator functions are simultaneously non-zero for only one (random) of the 2^{s-1} subsets of M , so that (5.17) is really a single term expressed in this compact form. Like the case of (5.8) or (5.11), L_n^{**} is also conditionally distribution-free, and asymptotically it has (under H_0) the chi-bar distribution.

This chi-bar distribution is a convex combination of several chi distributions (with DF 0,1,...,s-1) where the mixing coefficients are given by the orthant probabilities of certain multivariate normal distribution [we may refer to Barlow et al. (1972) for some of these details]. Note that by (5.11) and (5.12), $(L_n^{**})^2 \leq L_n^{o*}$, although under the restricted alternative these two would tend to be close to each other. On the other hand, the critical value of $(L_n^{**})^2$ will be smaller than that of L_n^{o*} (at the same significance level), and hence, the UI-test in (5.17) is generally power-superior to the other CMH-test in (5.11) when we restrict ourselves to the restricted alternatives. For some simple non-parametric tests against restricted alternatives, this power-superiority picture has been studied by Chatterjee and De (1974) and Chinchilli and Sen (1981b), among others. The proposed UI versions of the CMH tests extend their findings to discrete (ordered categorical) multivariate setups, and share the basic properties.

For simplicity of presentation, we have discussed specifically the case of ordered alternatives where the UI-principle works out very conveniently. The UI-principle can also be applied to other forms of restricted alternatives. All we need to do is to identify a set B (such as in before (5.12)) and for each element $\underset{\sim}{b}$ in B , consider the normalized form of a suitable CMH-type test statistic, and then finally to find out the maximum of all such statistics where $\underset{\sim}{b}$ is restricted to the set B. In this process, often, the well known Kuhn-Tucker-Lagrange point formula theorem in non-linear programming can be used with advantage, and the end-product can be expressed in a form similar to (5.17). In categorical data models, such UI-test has not been applied so extensively for restricted alternatives, and the current prescription enables us to advocate this possibility and to stress on its scope and efficiency. The test statistic in (5.17) may also be regarded as an extension of the UI-test for ordered alternatives (in the nonparametric case) when there are ties in the observations or when one has purely ordinal data.

6. INTERIM ANALYSES AND COMBINATION OF TESTS

In multi-center clinical trial and other follow-up studies, generally, the response occurs sequentially over time. Although, unlike the classical (time-) sequential case, in follow-up studies, because of time and cost constraints, censoring is a common phenomenan. In this context, it is not uncommon to have interim analyses of the accumulating data. In many follow-up studies, from the medical ethics or other considerations, one would like to have an early termination of the study if a clear statistical decision can be made at that stage; indeed, when the null hypothesis (of homogeneity or other forms of compatability) is not tenable (at an earlier stage), it may be wise to curtail the experimentation and switch the patients to the better treatment for their benefit. Also, from the epidemiological point of view, such statistical monitoring provides information on the general trend even when no early termination is contemplated. In course of this statistical monitoring, it may be desired to have a statistical assessment of the accumulating data at regular time intervals or after certain fractions of the data becore available. In this context, repeated significance tests are used [viz., Armitage, McPherson and Rowe (1969)]. Technically, if we decide a priori to look into the accumulating data sets $k(\geq 1)$ times, and if these test statistics are denoted by $T_1,\ldots,T_k$, respectively, then we have a situation very similar to the combination of tests (as discussed in earlier sections). However, these test statistics are based on accumulating data , and hence, are generally not independent of each other. Although a similar situation arises in the classical sequential analysis, there is a fundamental difference: In the classical sequential analysis, one usually encounters a stochastic process with independent and homogeneous increments, whereas in the context of repeated significance tests, neither this independence nor the homogeneity of the increments of the underlying stochastic processes can be taken for granted. On the top of that,in clinical trials, often a statistical monitoring is conducted in the so called continuous manner, so that one encounters a so called progressive censoring scheme [viz., Chatterjee and Sen (1973)] where k is either large or may even be technically uncountable. The main difficulty with the repeated significance tests arising in a multiple or progressive censoring scheme is the determination of the significance levels of the individual tests leading to a specified significance level for the overall test. In an asymptotic setup, this difficulty can be overcome by using suitable invariance principles for the related stochastic processes,and thereby, appealing to some standard results on Wiener processes (or other forms of processes with independent and homogeneous increments) ; an adequate treatment of this methodology is given in Sen (1981 ,1985a). As in the earlier sections, in testing for the null hypothesis,in a repeated significance testing or progressive censoring setup, we may be interested in some restricted alternatives. In this context, the Roy UI-principle can again be used with advantage to find out suitable testing procedure having some power-superiority properties over the tests for global alternatives. A general class of progressively censored tests for restricted alternatives based on the UI-principle has recently been considered by Sen(1985b). Though, primarily, this theory is developed for rank based tests, it remains applicable for the Cox (1972) proportional hazard models as well as to other forms of (partial or total) likelihood ratio type tests. In the remaining of this section, we consider the combination of independent repeated significance tests arising typically in multi-center clinical trials. For simplicity, we consider the case of a finite number of component tests for each of the centres.

Let there be $q(\geq 1)$ centres at each of which we have a clinical trial, and let there be $k(\geq 1)$ time-points $\tau_1, \ldots, \tau_k$, at which for each center, we have a test statistic for testing a common null hypothesis against some (possibly restricted) alternative ones. We denote these test statistics by T_{ij}, $i=1,\ldots,q$; $j=1,\ldots,k$, where k and q are fixed in advance. The problem is to combine these tests to obtain a suitable overall test for the null hypothesis accross the centers.

A crude way of doing this is to construct for each $i(=1,\ldots,q)$ a repeated significance test statistic T_i^o and to compute the corresponding OSL. Then the Fisher method (or some other one),discussed in Section 2, can easily be adopted to combine these OSL for an overall test. In this context, for the T_i^o , one may use suitable test statistics designed for restricted alternatives, provided the corresponding OSL may be computed in a convenient manner. If we use the Fisher OSL method, the overall test statistic will have (under H_0) the central chi square distribution with 2q DF. Let us look at the same problem from the point of view of a step-down procedure (as outlined in Section 3). At time-point τ_1 , we consider an overall test for the null hypothesis (against the desired alternatives) based on the data collected from all the q centers; we denote this test statistic by T_1^* . Similarly, at time-point τ_j , we consider a suitable conditional test T_j^* based on the data set from all the q centers upto the time-point τ_j and conditional on the test statistics at the earlier time-points τ_r ,$r \leq j-1$, for $j = 1,\ldots,k$. Under fairly general regularity conditions, these test statistics at the successive time-points can be shown to be quasi-independent (as in Section 3), so that one may either use the conventional theory for the usual step-down procedures and test for these component hypotheses at significance levels $\alpha_1 ,\ldots, \alpha_k$, respectively, where [as in (3.1)] for the overall significance level α , we have $1 -\alpha = (1-\alpha_1)\ldots(1-\alpha_k)$. Alternatively, as in Sen (1983) ,we may use the Fisher OSL method on these quasi-independent test statistics , and the resulting test statistic will have (under H_0) the central chi-square distribution with 2k DF. In either way, we allow the possibility of an early termination of the study (when for some $j < k$, T_j^* exceeds its $100\alpha_j\%$ critical level or the partial sum of the OSL exceeds the $100\alpha\%$ point of the chi square d.f. with 2k DF). If q is large compared to k , this alternative method will generally lead to more efficient tests than the earlier OSL method based on the T_i^o , $i=1,..,q$. Actually, when some homogeneity of the patterns accross the centres in a multi-center clinical trial can be presumed, pooling of the tests at each time-point τ_j seems to be very reasonable, and in such a case, if early termination is not sought, one may even use the theory of UI-principle (outlined in Section 4) to formulate an optimal overall test based on the successive steps tests . Such overall repeated significance tests from independent sources work out well when k is not that large. If k is large , but q is not, then, we may prescribe the earlier OSL method. On the other hand, if both k and q are large, then either of the two methods described above may have certain drawbacks ; in fact, they may become quite inefficient. Since such a situation may arise in the context of progressive censoring schemes (where k is generally very large) related to a coordinating trial involving a large number of clinics, we conclude this section with some remarks on this case too.

For the single clinic case, in a progressive censoring setup, nonparametric tests against restricted alternatives have been considered in Sen (1985b). Typically, in such a problem, the underlying stochastic process converges in distribution to a Bessel process; for restricted alternatives, we may have the so called Bessel-bar process. If we adopt the second approach then at each failure point, we may construct an overall test statistic (by pooling the information accross the clinics), and for the stochastic process relating to this sequence of overall test statistics, we would comeup with suitable functionals of the Bessel process. Thus, suitable overall tests can be constructed by reference to the distribution theory of such functionals for the classical Bessel process. Though in some simple cases, these distributions are available in the literature, in a general setup, there is a good deal of needs for the development of the related probability distributional theory for the exit times for such functionals of Bessel processes. Fortunately, in most of these cases, simulation methods work out very conveniently.

REFERENCES

ARMITAGE, P., MCPHERSON, C.K. and ROWE, B.C. (1969). Repeated significance tests on accumulating data. *J. Roy. Statist. Soc. A. 132*, 235-244.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M. and BRUNK, H.D. (1972). *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression.* New York, Wiley.

BERK, R.H. and COHEN, A. (1978). Asymptotically optimal methods of combining tests. *J. Amer. Statist. Assoc. 74*, 812-814.

BHATTACHARYA, N. (1961). Sampling experiments on the combination of independent χ^2-tests. *Sankhya, Ser. A 23*, 191-196.

BIRNBAUM, A. (1954). Combining independent tests of significance. *J. Amer. Statist. Assoc. 49*, 559-574.

CHATTERJEE, S.K. and DE, N.K. (1974). On the power superiority of certain bivariate location tests against restricted alternatives. *Calcutta Statist. Assoc. Bull. 23*, 73-84.

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

CHINCHILLI, V.M. and SEN, P.K. (1981a). Multivariate linear rank statistics and the union-intersection principle for hypotheses testing under restricted alternatives. *Sankhya, Ser. B 43*, 135-151.

CHINCHILLI, V.M. and SEN, P.K. (1981b). Multivariate linear rank statistics and the union-intersection principle for the orthant restriction problem. *Sankhya, Ser. B 43*, 152-172.

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

EDGINGTON, E.S. (1972). A normal curve method for combining probability values from independent experiments. *J. Psych. 82*, 85-89.

FISHER, R.A. (1932). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd, 4th ed.

FOLKS, J.L. (1984). Combination of independent tests. In *Handbook of Statistics, Vol.4 : Nonparametric Methods* (ed. P.R. Krishnaiah and P.K. Sen), Amsterdam: North Holland, pp. 113-121.

GEORGE, E.O. and MUDHOLKAR, G.S. (1977). The logit method for combining independent tests. *Bull. Inst. Math. Statist. 6*, 212.

GOOD, I.J. (1955). On the weighted combination of significance tests. *J. Roy. Statist. Soc. B 17*, 264-265.

KOZIOL, J.A. and PERLMAN, M.D. (1978). Combining independent chi square tests. *J. Amer. Statist. Assoc. 73*, 753-763.

KUDO, A. (1963). A multivariate analogue of the one-sided test. *Biometrika 50*, 403-418.

LANDIS, J.R., HEYMAN, E.R. and KOCH, G.G. (1978). Average partial association in three-way contingency tables: a review and discussion of alternative tests. *Inter. Statist. Rev. 46*, 237-254.

LIPTAK, T. (1958). On the combination of independent tests. *Magyay Tud. Akad. Mat. Kutato Int. Kozl. 3*, 171-197.

LITTELL, R.C. and FOLKS, J.L. (1971). Asymptotic optimality of Fisher's method of combining independent tests. *J. Amer. Statist. Assoc. 66*, 802-806.

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

MARDEN, J. I. (1982). Combining independent noncentral chi squared or F tests. *Ann. Statist. 10,* 266-277.

MONTI, K.L. and SEN, P.K. (1976). The locally optimal combination of independent test statistics. *J. Amer. Statist. Assoc. 71,* 903-911.

MUDHOLKAR, G.S. and SUBBAIAH, P. (1980). Testing significance of a mean vector- a possible alternative to Hotelling T^2. *Ann. Inst. Math. Statist. 32,* 43-52.

OOSTERHOFF, J. (1969). *Combination of One Sided Statistical Tests.* Amsterdam: Math. Centre.

PEARSON, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika 25,* 379-410.

PERLMAN, M.D. (1969). One-sided testing problems in multivariate analysis. *Ann. Math. Statist. 40,* 549-567.

ROY, J. (1958). Step-down procedure 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.

SCHAAFSMA, W. (1968). A comparison of the most stringent and the most stringent somewhere most powerful test for certain problems with restricted alternative. *Ann. Math. Statist. 39,* 531-546.

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

SEN, P.K. (1982). The UI-principle and LMP rank tests. *Coll. Mat. Soc. J. Bolyai 32* , 843-858.

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

SEN, P.K. (1985a). *Theory and Application of Sequential Nonparametrics.* Philadelphia: SIAM publ.

SEN, P.K. (1985b). Nonparametric testing against restricted alternatives under progressive censoring. *Sequen. Anal. 4,* 247-276.

TIPPETT, L.H.C. (1931). *The Methods of Statistics.* London: Williams & Norgate.

WILKINSON, B. (1951). A statistical consideration in psychological research. *Psych. Bull. 48,* 156-158.