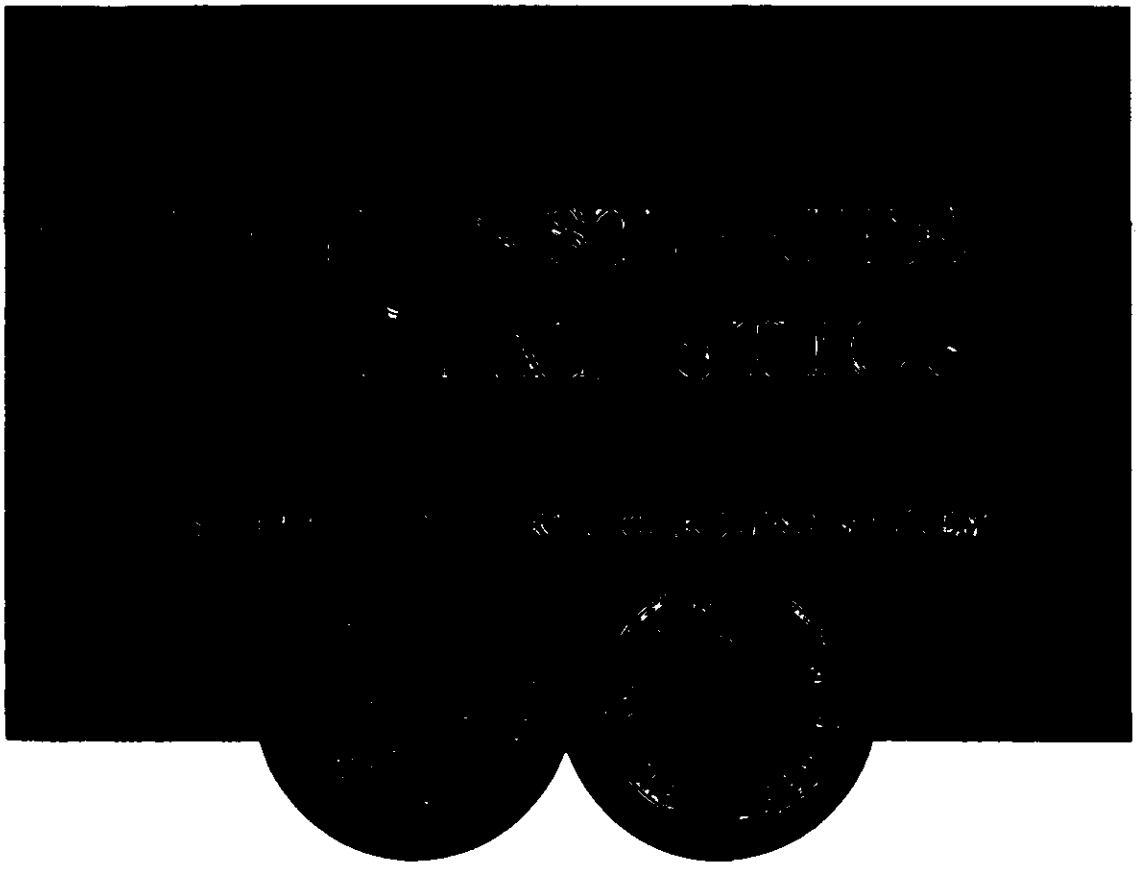

On Efficient Estimation of the Prevalence of Multiple Rare Traits

by: Jacqueline M. Hughes-Oliver
Department of Statistics
North Carolina State University

William F. Rosenberger
Department of Mathematics and Statistics
University of Maryland, Baltimore County

MIMEO SERIES #2505

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

MIMEO SERIES #2505
On Efficient Estimation of the Prevalence of Multiple Rare Traits
by: J. Hughes-Oliver, W. Rosenberger

NAME DATE

On Efficient Estimation of the Prevalence of Multiple Rare Traits

Jacqueline M. Hughes-Oliver
Department of Statistics,
North Carolina State University,
Raleigh, NC 27695, U.S.A
hughesol@stat.ncsu.edu

William F. Rosenberger
Department of Mathematics and Statistics,
University of Maryland, Baltimore County,
Baltimore, MD 21250, U.S.A.
billr@math.umbc.edu

Abstract

We consider a population with multiple traits of interest, where our goal is to estimate the proportions of individuals with the traits. When traits are rare, group testing can improve efficiency. Previous work of Hughes-Oliver and Swallow (Journal of the American Statistical Association 89, 982-993; 1994) developed an adaptive two-stage design for group testing of only one trait. We extend this work to the multi-trait case. We derive the optimal group sizes using compound D-optimum design theory. Estimation is based on maximum likelihood estimators, which are shown to be consistent and asymptotically normal. We apply our design to a problem of estimating the prevalence of HIV and hepatitis in Romanian children.

Some key words: Adaptive designs; Compound D-optimality; General equivalence theorem; Group testing; Optimal design; Pooled testing.

1 Introduction

Efficient estimation of the prevalence of a trait is an important problem in epidemiology and risk assessment. For rare traits, the advantages of group testing, where several units are pooled and a single test applied to the entire group, are well documented. Group testing has been successfully applied to identifying men with syphillis (Dorfman, 1943) and, more recently, to estimating seroprevalence of HIV (Emmanuel et al., 1988; Kline et al., 1989; Behets et al., 1990). For example, Behets et al. (1990) report a 78% cost saving from using groups of size 10 instead of individual testing.

Optimal determination of group size is always an issue, because it depends on the unknown prevalence. The two-stage adaptive scheme proposed by Hughes-Oliver and Swallow (1994) reduces the effect of a priori estimates on the final estimate, thus making group testing more robust to possibly misleading prior information. Measurement error is also a major concern in group testing. Tu, Litvak and Pagano (1995) model the effect of sensitivity and specificity of a diagnostic test on the estimation of prevalence and find that grouping improves the efficiency. At the same time, Behets et al. (1990) find that "pools resulted in 100% specificity compared to 99.8% for testing individual sera" so that testing in groups reduced the measurement error.

The foregoing results have been limited to prevalence of a single trait. However, Rudin et al. (1990) describe a study where samples of blood were taken from 169 orphans in Pascani and then subjected to eight tests: two for HIV, three for hepatitis B, one for hepatitis A, one for hepatitis C, and one for measles. The goal was to estimate the prevalence of all these diseases among the orphans in Pascani. The extension of grouping strategies to multiple responses, that is, multiple traits, is complicated by the possibly conflicting goals of estimation. Should several group sizes, one for each prevalence, be used, or is a single group size better? How should these group sizes be selected? These issues also

impact adaptive schemes, bringing into question the limiting behavior of the estimators. In this paper we integrate the theory of optimum design of experiments with determination of group sizes for estimating the prevalence of multiple traits. Compound D-optimality (Cook and Wong, 1994; Atkinson and Bogacka, 1997) is used to balance the influence of the different responses, even though it is usually applied to competing models for a single response. We also derive the limiting behavior of the optimal group sizes, relative to the dependent (adaptive) sampling scheme, and determine the limiting behavior of the prevalence estimators.

Consider a population with T traits of interest. We wish to estimate the proportions $p_t, t = 1, \ldots, T$, where $p_t \in (0, 1)$ is the proportion of individuals in the population having the t^{th} trait. Individuals from the population are placed in groups and these groups are then subjected to T tests, one for each trait. The group size is periodically updated to reflect the most current knowledge of the proportions. Each group must contain at least one individual and is restricted to a maximum size, say K_{max}, for practical reasons. For simplicity, we limit discussion to a two-stage approach. Extensions to multi-stage or fully sequential approaches are possible.

Suppose λN groups of sizes $k_1, \ldots, k_{\lambda N}$ are tested in Stage 1; that is, λN tests are performed for each of the T traits, where $\lambda \in (0, 1]$ is assumed known a priori, and groups are allowed to have distinct sizes. The results of the tests in stage 1 are tallied and used to decide group sizes for the second stage. In stage 2, $(1-\lambda)N$ groups of sizes $k_{\lambda N+1}, \ldots, k_N$ are tested for each of the T traits. The results from both stages are combined to yield estimates of the proportions $p_1, \ldots, p_T$. Because λ is specified a priori, and is not determined from some optimality criterion, it is easily selected to make λN and $(1 - \lambda)N$ integers.

We make two simplifying assumptions; extensions are discussed in Section 6. The first assumption is that there are no errors in testing, that is, if a group contains a positive

individual then this will be detected with a positive group result and if a group contains no positive individuals then the group result will be negative. The second assumption is that test results for the different traits are independent.

In Section 2, we derive the maximum likelihood estimator and use martingale techniques to show the usual asymptotic properties. In Section 3, we discuss optimality criteria for selecting the group sizes and discuss limiting behavior. In Section 4, we investigate the sensitivity of the optimal designs to user-supplied weights and prior information. In Section 5, we apply our techniques to the Rudin et al. (1990) data. Finally, we end with a discussion in Section 6.

2 Likelihood Results

Let $Y_i^t = 1$ if the i^{th} group tests positive for the t^{th} trait, and $Y_i^t = 0$ otherwise, for $i = 1, \ldots, N, t = 1, \ldots, T$. If $\mathbf{Y}_N^t \equiv (Y_1^t, \ldots, Y_N^t)$ and $\mathbf{p} = (p_1, \ldots, p_T)$, then the likelihood, up to the N^{th} group, is

$$\begin{aligned} \mathcal{L}_N(\mathbf{p}) &= \mathcal{L}\{\mathbf{Y}_N^1\}\mathcal{L}\{\mathbf{Y}_N^2\}\cdots\mathcal{L}\{\mathbf{Y}_N^T\} \\ &= \prod_{i=1}^{N}\prod_{t=1}^{T}\left[1-(1-p_t)^{k_i}\right]^{Y_i^t}\left[(1-p_t)^{k_i}\right]^{(1-Y_i^t)}, \end{aligned}$$

with

$$\begin{aligned} \frac{\delta}{\delta p_t}\ln\mathcal{L}_N(\mathbf{p}) &= \sum_{i=1}^{N}\frac{Y_i^t-\left[1-(1-p_t)^{k_i}\right]}{\left[1-(1-p_t)^{k_i}\right]\left[(1-p_t)^{k_i}\right]}k_i(1-p_t)^{k_i-1} \\ &\equiv \sum_{i=1}^{N}\frac{\delta}{\delta p_t}L_i(\mathbf{p}) \qquad t=1,\ldots,T. \end{aligned} \tag{1}$$

The maximum likelihood estimator $\hat{\mathbf{p}}$ of $\mathbf{p}$ is the root of the system of T equations in (1). If we consider only groups $1, \ldots, \lambda N$, then we have independent sampling, and all the usual estimation theory applies. If we consider all groups $1, \ldots, N$, then we have dependent

sampling because the group sizes of the second stage depend on the results of the first stage. The usual asymptotic results no longer apply, and so we use martingale theory to establish properties of $\widehat{\mathbf{p}}$.

Provided suitable conditions are met, $\widehat{\mathbf{p}}$ is consistent and $N^{1/2}(\widehat{\mathbf{p}} - \mathbf{p})$ is asymptotically multivariate normal with mean $\mathbf{0}$ and variance-covariance matrix $[\boldsymbol{\Sigma}(\mathbf{p})]^{-1}$, for an appropriately defined matrix $\boldsymbol{\Sigma}(\mathbf{p})$ having an inverse. Sufficient conditions for the validity of these asymptotic results are given by Rosenberger, Flournoy, and Durham (1997) for multi-parameter problems. The main requirement is that

$$\frac{1}{N}\sum_{i=1}^{N} E_{i-1}\left\{\frac{\delta}{\delta p_t}L_i(\mathbf{p})\frac{\delta}{\delta p_s}L_i(\mathbf{p})\right\} \xrightarrow{a.s.} [\boldsymbol{\Sigma}(\mathbf{p})]_{ts} \quad \text{as } N \to \infty, \text{ for } t,s = 1,\ldots,T, \tag{2}$$

where $L_i(\mathbf{p})$ is defined in (1) and $E_{i-1}\{\cdot\}$ is expectation conditioned on the results up to and including the $(i-1)^{st}$ group test. Rosenberger et al. (1997) refer to (2) as condition (A3). Because $E_{i-1}\left\{\frac{\delta}{\delta p_t}L_i(\mathbf{p})\frac{\delta}{\delta p_s}L_i(\mathbf{p})\right\}$ is proportional to $Cov_{i-1}(Y_i^t, Y_i^s)$ and we assume that the traits are independent, then $E_{i-1}\left\{\frac{\delta}{\delta p_t}L_i(\mathbf{p})\frac{\delta}{\delta p_s}L_i(\mathbf{p})\right\} = 0$ for all $t \neq s$. When $t = s$,

$$E_{i-1}\left\{\frac{\delta}{\delta p_t}L_i(\mathbf{p})\frac{\delta}{\delta p_s}L_i(\mathbf{p})\right\} = \frac{k_i^2(1-p_t)^{2k_i-2}}{[1-(1-p_t)^{k_i}]\,[(1-p_t)^{k_i}]} \qquad i = 1,\ldots,N,$$

resulting in

$$\frac{1}{N}\sum_{i=1}^{N} E_{i-1}\left\{\frac{\delta}{\delta p_t}L_i(\mathbf{p})\frac{\delta}{\delta p_s}L_i(\mathbf{p})\right\} = \frac{1}{N}\sum_{i=1}^{N}\frac{k_i^2(1-p_t)^{2k_i-2}}{[1-(1-p_t)^{k_i}]\,[(1-p_t)^{k_i}]} \tag{3}$$

Hence verification of (2) is transformed to determination of the limiting behavior of group sizes $k_1,\ldots,k_N$. Because $0 < p_t < 1$ and $k_i \in [1, K_{max}]$ is bounded, the other conditions of Rosenberger et al. (1997) are easily verified for our model. Techniques of Rosenberger and Sriram (1997) can also be used to show that $\widehat{\mathbf{p}}$ converges almost surely to $\mathbf{p}$.

Group sizes are selected for achieving some inferential goal, usually following the theory of optimal design of experiments. For example, Hughes-Oliver and Swallow (1994) choose

group sizes to minimize mean squared error (MSE). For choosing $k^{(1)} \equiv k_1 = \ldots = k_{\lambda N}$, they evaluate the MSE using an a priori estimate of the proportion. The group sizes selected in this way are nonrandom, change with N, and approach a limit. For choosing $k^{(2)} \equiv k_{\lambda N+1} = \ldots = k_N$, they evaluate MSE using the estimate of the proportion obtained from the data of stage 1. These group sizes are random and change with N, but still converge almost surely to a constant. For now let us suppose that as $N \to \infty$ the group sizes $k_1, \ldots, k_N$ of our model converge almost surely to a member of the (finite) set $\{K_1, \ldots, K_m\}$. (For example, in Section 3 our design criterion is such that the group sizes $k_1, \ldots, k_{\lambda N}$ for the first stage converge almost surely to some constant K_1 and the group sizes $k_{\lambda N+1}, \ldots, k_N$ for the second stage converge almost surely to some constant K_2.) Consequently, $\hat{\mathbf{p}}$ is consistent and $N^{1/2}(\hat{\mathbf{p}} - \mathbf{p})$ is asymptotically multivariate normal with mean $\mathbf{0}$ and variance-covariance matrix $[\boldsymbol{\Sigma}(\mathbf{p})]^{-1}$, where $\boldsymbol{\Sigma}(\mathbf{p})$ is diagonal with the t^{th} diagonal element equal to the right hand side of (3), where $k_1, \ldots, k_N$ are replaced by their limits.

3 Selection and Limiting Behavior of Group Sizes

Careful selection of the size used for groups in both stages can lead to improved efficiency of $\hat{\mathbf{p}}$. In this sense, selection of group sizes is an exercise in experimental design. We use ξ^I to represent the continuous design having m^I distinct support points for Stage I, and ξ^{II} to represent the continuous design having m^{II} distinct support points for Stage II. In particular,

$$\xi^I = \left\{ \begin{array}{ccc} k_1^I & \ldots & k_{m^I}^I \\ w_1^I & \ldots & w_{m^I}^I \end{array} \right\},$$

where the k_j^I's are the group sizes and w_j^I is the proportion of the groups tested in Stage I that are of size k_j^I. A similar definition holds for ξ^{II}. In practice, for a given number of groups, say λN in Stage I, the design weights w_j^I may not yield an integer number of

groups $w_j^I \lambda N$ of size k_j^I. Some adjustments are often needed to find efficient designs for finite sample sizes. The method of efficient apportionment is one possibility. The reader is referred to Pukelsheim (1993, Ch. 12) for details.

3.1 Limiting behavior

Various criteria for optimal design of experiments exist in the literature (Atkinson and Donev, 1992; Pukelsheim, 1993). These are usually functions of the information matrix of the unknown parameters. In our case, $\boldsymbol{\Sigma}(\mathbf{p})$ as defined in Section 2 is the information matrix for $\mathbf{p}$.

For nonlinear dependence of the response on the parameters, as we have in our model, the information matrix is itself a function of the unknown parameters. Consequently, some prior knowledge of $\mathbf{p}$ is required for determining an optimal design. This prior information may be in the form of a single $\mathbf{p}_0 = (p_{01}, \ldots, p_{0T})'$ or a prior distribution $\pi(\mathbf{p})$ on $\mathbf{p}$. This prior information is used to provide group sizes for Stage I. At the end of Stage I, an estimate $\widehat{\mathbf{p}}_1$ is obtained and used to provide group sizes for Stage II. While the group sizes used in Stage I are nonrandom functions, the group sizes used in Stage II are random because they depend on the value of the stochastic process in Stage I.

Irrespective of the particular criterion used, we must determine the limiting behavior of the group sizes from both stages. Provided the criterion function does not change with N, the limiting behavior of groups sizes from Stage I is trivially obtained. The following result addresses the limiting behavior of group sizes from Stage II.

Theorem 1 *Suppose the criterion function $\Psi(\xi; \mathbf{p})$ is continuous in $\mathbf{p}$ and continuously differentiable and concave in ξ. Suppose further that while the criterion function depends on $\mathbf{p}$, it does not change with N. If $\xi^* = \arg\max_\xi \Psi(\xi; \mathbf{p}^*)$ and $\mathbf{p}^* \xrightarrow{a.s.} \mathbf{p}^0$, then $\xi^* \xrightarrow{a.s.} \xi^0$, where $\xi^0 = \arg\max_\xi \Psi(\xi; \mathbf{p}^0)$.*

The proof is in Appendix A.

The consequence of Theorem 1 is that the limiting behavior of the Stage II group sizes depends on the limiting behavior of $\widehat{\mathbf{p}}_1$. If $\widehat{\mathbf{p}}_1$ is the maximum likelihood estimator of $\mathbf{p}$ from Stage I data only, then $\widehat{\mathbf{p}}_1$ is strongly consistent for $\mathbf{p}$, independent of $\mathbf{p}_0$. Hence, assuming all other conditions of Theorem 1 are met, $\xi^{II} = \arg\max_\xi \Psi(\xi; \widehat{\mathbf{p}}_1) \xrightarrow{a.s.} \arg\max_\xi \Psi(\xi; \mathbf{p})$, where $\mathbf{p}$ is the true (unknown) vector of proportions.

We now consider specific criterion functions.

3.2 *D*-optimality and Compound *D*-optimality

D-optimum designs maximize the determinant of the information matrix $\boldsymbol{\Sigma}(\mathbf{p})$. Because we have two separate stages for which optimal designs are required, it is convenient to generalize the notation to emphasize the dependence of $\boldsymbol{\Sigma}(\mathbf{p}; \lambda)$ on λ. Design ξ^I is obtained by maximizing the determinant of $\boldsymbol{\Sigma}(\mathbf{p}; \lambda = 1)$, and design ξ^{II} is obtained by maximizing the determinant of $\boldsymbol{\Sigma}(\mathbf{p}; \lambda)$ for the actual value of λ used in data collection.

When there is unequal interest in the different parameters, for example, if there is general interest in the prevalence of three diseases, but only one of these diseases is life threatening, compound *D*-optimality (Cook and Wong, 1994; Atkinson and Bogacka, 1997) is an effective approach to balancing the conflicting goals of optimal design of experiments. Of course, no criterion can be best for every goal, so compound *D*-optimality uses a weighted combination of *D*-optimality criteria for the different goals. The weight for a particular goal represents the relative importance of that goal.

We consider estimation of p_t, $t = 1, \ldots, T$, where α_t is the relative importance of p_t and $\sum_t \alpha_t = 1$. The compound *D*-optimum criterion function to be maximized is

$$\Psi(\xi^I, \xi^{II}; \mathbf{p}, \lambda) = \sum_{t=1}^{T} \alpha_t \log[\boldsymbol{\Sigma}(\mathbf{p}; \lambda)]_{tt}, \tag{4}$$

where

$$[\Sigma(\mathbf{p};\lambda)]_{tt} = \lambda \sum_{j=1}^{m^I} w_j^I \frac{(k_j^I)^2(1-p_t)^{2k_j^I-2}}{\left[1-(1-p_t)^{k_j^I}\right]\left[(1-p_t)^{k_j^I}\right]} + (1-\lambda)\sum_{j=1}^{m^{II}} w_j^{II} \frac{(k_j^{II})^2(1-p_t)^{2k_j^{II}-2}}{\left[1-(1-p_t)^{k_j^{II}}\right]\left[(1-p_t)^{k_j^{II}}\right]} \quad (5)$$

is the t^{th} diagonal element of $\Sigma(\mathbf{p};\lambda)$. To select the optimum design for Stage I, set $\mathbf{p} = \mathbf{p}_0$ and $\lambda = 1$ then maximize (4); that is, find ξ to maximize $\Psi^I(\xi;\mathbf{p}_0) \equiv \Psi(\xi,\cdot;\mathbf{p}_0,1)$. To select the optimum design for Stage II, conditioned on the design ξ^I for Stage I, set $\mathbf{p} = \widehat{\mathbf{p}}_1$ and λ to be the value chosen a priori, then maximize (4); that is, find ξ to maximize $\Psi^{II}(\xi;\widehat{\mathbf{p}}_1) \equiv \Psi(\xi^I,\xi;\widehat{\mathbf{p}}_1,\lambda)$.

The general equivalence theorem of Kiefer and Wolfowitz (1960) can be used to construct conditions for global optimality of a continuous design. However, because of the heterogeneity of our responses, the standard formulas are not applicable. The following theorem provides the necessary conditions for our model.

Theorem 2 *Suppose*

$$f(k,p) = \frac{k^2(1-p)^{2k-2}}{[1-(1-p)^k][(1-p)^k]}.$$

A design ξ^I is optimal for Stage I if

$$d^I(k,\xi^I) = \sum_{t=1}^{T} \alpha_t \frac{f(k,p_{0t})}{\sum_{j=1}^{m^I} w_j^I f(k_j^I,p_{0t})} \quad (6)$$

is less than or equal to 1 for all $k \in [1,K_{max}]$, with equality holding only when k is one of the support points of the design.

A design ξ^{II} is optimal for Stage II if

$$d^{II}(k,\xi^{II}) = \sum_{t=1}^{T} \alpha_t \frac{\lambda\sum_{j=1}^{m^I} w_j^I f(k_j^I,\widehat{p}_{1t}) + (1-\lambda)f(k,\widehat{p}_{1t})}{\lambda\sum_{j=1}^{m^I} w_j^I f(k_j^I,\widehat{p}_{1t}) + (1-\lambda)\sum_{j=1}^{m^{II}} w_j^{II} f(k_j^{II},\widehat{p}_{1t})} \quad (7)$$

is less than or equal to 1 for all $k \in [1,K_{max}]$, with equality holding only when k is one of the support points of the design. In (7), the optimal design ξ^I for Stage I is assumed already known.

The proof is in Appendix B.

Suppose Theorem 2 confirms ξ^I and ξ^{II} to be the optimal Stage I and Stage II designs. As previously mentioned, the limiting behavior of ξ^I is trivial (the limit is ξ^I), because the criterion function $\Psi^I(\xi; \mathbf{p}_0)$ is free of N and contains no random quantities. For Stage II, $\Psi^{II}(\xi; \widehat{\mathbf{p}}_1)$ is free of N but depends on the random $\widehat{\mathbf{p}}_1$. Because $\Psi^{II}(\xi; \mathbf{p})$ is continuous in $\mathbf{p}$ and continuously differentiable and concave in ξ for $k \in [1, K_{max}]$, and $\widehat{\mathbf{p}}_1$ is strongly consistent for the true $\mathbf{p}$, Theorem 1 says that ξ^{II} converges almost surely to $\xi^0 = \arg\max_\xi \Psi(\xi; \mathbf{p})$. Confirming the optimality of ξ^0 requires (7), except with $\widehat{p}_{1t}$ replaced by the true p_t.

The properties of $f(k, p)$ allow analytical formulations for the optimal designs. From (6), the Stage I optimal design ξ^I consists only of critical points (with respect to k) of $d^I(k, \xi)$. Because $f(k, p)$ is nonnegative and concave in k for $k \geq 1$ and $\alpha_t \geq 0$ for $t = 1, \ldots, T$, $d^I(k, \xi)$ is also concave in k and thus has a single critical point. Hence, ξ^I has only one support point k^I which solves

$$\sum_{t=1}^{T} \alpha_t \frac{\frac{\delta}{\delta k} f(k, p_{0t})}{f(k, p_{0t})} = 0. \tag{8}$$

Similarly, the Stage II optimal design ξ^{II} has only one support point k^{II} which solves

$$\sum_{t=1}^{T} \alpha_t \frac{\frac{\delta}{\delta k} f(k, \widehat{p}_{1t})}{\lambda f(k^I, \widehat{p}_{1t}) + (1-\lambda) f(k, \widehat{p}_{1t})} = 0, \tag{9}$$

and whose almost sure limiting value is the solution to

$$\sum_{t=1}^{T} \alpha_t \frac{\frac{\delta}{\delta k} f(k, p_t)}{\lambda f(k^I, p_t) + (1-\lambda) f(k, p_t)} = 0. \tag{10}$$

The specific asymptotic properties of $\widehat{\mathbf{p}}$, as outlined in Section 2, may now be addressed. The limiting behavior of the optimal group sizes is

$$k_1, \ldots, k_{\lambda N} \xrightarrow{a.s.} K_1, \quad \text{the solution of (8)},$$

and

$$k_{\lambda N+1}, \ldots, k_N \xrightarrow{a.s.} K_2, \quad \text{the solution of (10)}.$$

Consequently, $\hat{\mathbf{p}}$ is consistent and $N^{1/2}(\hat{\mathbf{p}}-\mathbf{p})$ converges in distribution to $N\left(\mathbf{0},[\boldsymbol{\Sigma}(\mathbf{p})]^{-1}\right)$, where $\boldsymbol{\Sigma}(\mathbf{p}) = \text{diag}\left\{\lambda f(K_1,p_t)+(1-\lambda)f(K_2,p_t)\right\}, t=1\ldots,T$.

4 Sensitivity of optimal design to weights and prior information

The methods of Section 3 require specification of weights α_t, $t = 1,\ldots,T$, to represent the relative importance of estimating p_t. Unfortunately, there is usually uncertainty in the relative importance of the various proportions, so that it is difficult to decide on a single value for $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_T)$. Because of this, it is more useful to assess the efficiency of a particular design with respect to several values of $\boldsymbol{\alpha}$. In this section we assess the efficiencies of three designs, obtained from using specific values of $\boldsymbol{\alpha}$, relative to the compound design for a range of values of $\boldsymbol{\alpha}$. In this section we illustrate the effect of under- or over-shooting the target $\mathbf{p}$.

Suppose there are $T = 2$ traits with true (unknown) prevalences of $p_1 = 0.04$ and $p_2 = 0.08$. Suppose also that the number of tests will be apportioned equally to both stages, that is, $\lambda = 0.5$. For given $\boldsymbol{\alpha} = (\alpha, 1-\alpha)$ and $\mathbf{p_0} = (p_{01}, p_{02})$, let the compound D-optimum design $\xi_C \equiv (\xi^I, \xi^{II})$ be obtained as in Section 3. We use this design as the basis for comparing three other designs: the D_s-optimum design ξ_1 where interest is limited to estimation of p_1 only; the D_s-optimum design ξ_2 where interest is limited to estimation of p_2 only; and the D-optimum design ξ_D where there is equal interest in estimation of both p_1 and p_2. These designs are constructed using the methods in Section 3 with $\boldsymbol{\alpha}$ equal $(1,0)$, $(0,1)$, and $(0.5,0.5)$, respectively.

The means of comparison is the asymptotic relative efficiency based on the model pre-

scribed by $\boldsymbol{\alpha}$. That is, to compare ξ_1 and ξ_C, we use

$$E_1 = 100\frac{\lambda f(k_C^I, p_1) + (1-\lambda) f(k_C^{II}, p_1)}{\lambda f(k_1^I, p_1) + (1-\lambda) f(k_1^{II}, p_1)},$$

where k_1^I, k_1^{II} are the Stage I and Stage II asymptotic group sizes from ξ_1 and k_C^I, k_C^{II} are the Stage I and Stage II asymptotic group sizes from ξ_C. Similarly, to compare ξ_2 and ξ_C, we use

$$E_2 = 100\frac{\lambda f(k_C^I, p_2) + (1-\lambda) f(k_C^{II}, p_2)}{\lambda f(k_2^I, p_2) + (1-\lambda) f(k_2^{II}, p_2)},$$

where k_2^I, k_2^{II} are the Stage I and Stage II asymptotic group sizes from ξ_2. To compare ξ_D and ξ_C, the asymptotic covariance matrices are of dimension two, hence we take the square roots of the determinants to remove the effect of dimension (Atkinson and Bogacka, 1997). The efficiency measure is defined as

$$E_D = 100\left[\frac{\{\lambda f(k_C^I, p_1) + (1-\lambda) f(k_C^{II}, p_1)\}\{\lambda f(k_C^I, p_2) + (1-\lambda) f(k_C^{II}, p_2)\}}{\{\lambda f(k_D^I, p_1) + (1-\lambda) f(k_D^{II}, p_1)\}\{\lambda f(k_D^I, p_2) + (1-\lambda) f(k_D^{II}, p_2)\}}\right]^{1/2},$$

where k_D^I, k_D^{II} are the Stage I and Stage II asymptotic group sizes from ξ_D.

The relative efficiencies are shown in Figure 1 where the prior information is given as $\mathbf{p_0} = 0.5\mathbf{p}, \mathbf{p}, 1.5\mathbf{p}$. The corresponding optimal asymptotic designs ξ_1, ξ_2, ξ_D are shown in Table 1. The Stage II asymptotic group sizes are not much affected by changing $\mathbf{p_0}$, but the Stage I asymptotic group sizes are, as we expect. To aid in intuition regarding these group sizes, an anecdotal comment is in order. For large $\mathbf{p_0}$, the optimum Stage I group size is "close to" $\arg\max_k f(k, \overline{p}_0)$, where $\overline{p}_0 = \sum_t \alpha_t p_{0t}$. For example, $\arg\max_k f(k, 0.09) = 16.90$ which is close to 16.53 for ξ_D with $\mathbf{p_0} = (0.06, 0.12)$. This works because $f(k, p)$ is "almost" flat in a neighborhood around its maximizer with respect to k and around $\overline{p}$.

When $\mathbf{p_0} = \mathbf{p}$, Figure 1(b), all efficiencies are at most 100 and ξ_D is most robust to changing values of α. When $\mathbf{p_0} = 1.5\mathbf{p}$, Figure 1(c), the robustness of ξ_2 to α improves, but ξ_1 loses robustness. In addition, some of the efficiencies exceed 100. This latter phenomenon is most evident when $\mathbf{p_0} = 0.5\mathbf{p}$, Figure 1(a). This unusual behavior makes perfect sense,

however. When the a priori information is incorrect, that is, $\mathbf{p}_0 \neq \mathbf{p}$, we are not maximizing the "correct" criterion function in (4), so that we do not actually obtain *the optimum* design. The implication is that it is possible to find better designs than the one we thought to be optimal.

The asymmetry of effect of under- versus over-shooting the true $\mathbf{p}$ is well documented; see, for example, Hughes-Oliver and Swallow (1994) and their references. Under-shooting leads to group sizes so large that most or all groups will test positive—a very uninformative outcome. Over-shooting leads to group sizes that are smaller than optimal, with the worst situation being one-at-a-time testing which can still be informative in large samples.

5 Application

We now apply our techniques to the Rudin et al. (1990) data on prevalence of several diseases in Romanian children as estimated shortly after the Romanian revolution. The region of interest, Pascani, is described as a "very poor region of north-eastern Romania (Moldavia)." Samples of blood were taken from 169 orphans in Pascani, then subjected to eight tests. The authors explain that "If more than 50 children lived in [an orphanage], blood was taken from the first 50 in alphabetical order." No reason is given for this, but it is likely that resources could not accommodate testing more samples. We consider only two tests: the western blot analysis (ELISA) for HIV, with a reported prevalence of 8.3%, and the hepatitis B surface antibody, with a reported prevalence of 89.3%.

Let $p_1 = 0.083$ = proportion of ELISA tests which are positive and $p_2 = 0.107$ = proportion of hepatitis B surface antibody tests which are negative. Suppose we are equally interested in these proportions, that is, $\alpha_1 = \alpha_2 = 0.5$, and that 84 tests will be done in Stage I, so that λ is approximately 0.5. Suppose further that our a priori information is $\mathbf{p}_0 = (0.04, 0.05)$, a sort of worst case scenario because we under-shoot the true proportions.

Then the asymptotic compound D-optimum (or just D-optimum) design uses a group size of 34.54 in Stage I and 15.75 in Stage II. The square root of the determinant of the scaled asymptotic covariance matrix of $\hat{\mathbf{p}}$, $1/|\boldsymbol{\Sigma}(\mathbf{p})|^{1/2}$, for this design is 0.0156, compared to .0853 for the one-at-a-time design used by Rudin et al. (1990). Thus grouping increases the (D-optimum) efficiency of the estimators in the same way that increasing the sample size by more than five times would allow.

Of course, one may argue that groups of sizes 35 and 16 are not feasible, because the ELISA test is only recommended for groups of size 15 or less (Kline et al., 1989). Even if groups of size two are used in Stages I and II, $1/|\boldsymbol{\Sigma}(\mathbf{p})|^{1/2} = 0.0449$, which is still much smaller than for the one-at-a-time design. Although small-sample results may differ from the asymptotic results discussed above, there is a distinct possibility that Rudin et al. (1990) would have benefited from the grouped, adaptive strategy presented in this paper. Testing in groups, as described in this paper, has the potential of producing much more efficient estimators of the prevalences, while using the same number of tests as Rudin et al. (1990).

6 Discussion

We have derived an optimum strategy for efficient estimation of the prevalence of multiple traits. While these methods may be applied to more prevalent traits, they are most advantageous when applied to rare traits. Application to the Rudin et al. (1990) data demonstrates that these methods can increase the efficiency of the estimators to a level that would ordinarily require many more tests. Because diagnostic tests are often expensive, these methods offer an alternative for achieving a high level of efficiency at a much lower cost.

The results obtained in this paper are based on assuming no measurement errors. As discussed in Section 1, Behets et al. (1990) argue that compared to individual testing, grouping improves specificity (that is, reduces the number of false positives) while not sacrificing

sensitivity (that is, maintaining the same number of false negatives). Nevertheless, we believe measurement error is an area which deserves additional investigation for the case of multiple traits. The level of accuracy of a diagnostic test is expected to affect the potential benefit of group testing, but is not expected to remove it and may actually increase it.

Another simplifying assumption made in this paper is that traits are independent. We typically do not expect this to be true. For example, HIV and hepatitis B are more realistically modeled as dependent, with a greater percentage of HIV positive individuals testing positive for hepatitis B. We are currently exploring a latent variable approach to binary group testing as a way of incorporating dependence of traits. Although our results are preliminary, we conjecture that strong correlation of traits will actually improve the efficiency of the group-testing estimators.

ACKNOWLEDGMENT

William F. Rosenberger is also appointed in the Department of Epidemiology and Preventive Medicine of the University of Maryland School of Medicine, Baltimore, MD, U.S.A.

APPENDIX 1

Proof of Theorem 1

Define $N_\delta(\xi^0)$ to be the neighborhood of ξ^0 having radius $\delta > 0$. By definition of ξ^0, for all $\delta > 0$ there exists $\epsilon = \epsilon(\delta) > 0$ such that

$$\Psi(\xi^0; \mathbf{p}^0) - \Psi(\xi^n; \mathbf{p}^0) > \epsilon \qquad \text{for any } \xi^n \in N_\delta(\xi^0). \tag{11}$$

Moreover, because $\mathbf{p}^* \xrightarrow{a.s.} \mathbf{p}^0$ and $\Psi(\xi; \mathbf{p})$ is continuous in $\mathbf{p}$, then $\Psi(\xi, \mathbf{p}^*) \xrightarrow{a.s.} \Psi(\xi; \mathbf{p}^0)$. Hence, with probability one there exists N_ϵ such that for all $N > N_\epsilon$,

$$|\Psi(\xi^0, \mathbf{p}^*) - \Psi(\xi^0; \mathbf{p}^0)| < \epsilon/2 \qquad \text{and} \qquad |\Psi(\xi^n, \mathbf{p}^*) - \Psi(\xi^n; \mathbf{p}^0)| < \epsilon/2. \tag{12}$$

Combining (11) and (12) yields

$$\Psi(\xi^0;\mathbf{p}^*) - \Psi(\xi^n;\mathbf{p}^*) > \Psi(\xi^0;\mathbf{p}^0) - \Psi(\xi^n;\mathbf{p}^0) - \epsilon > 0 \tag{13}$$

with probability one for all $N > N_\epsilon$ and for any $\xi^n \in N_\delta(\xi^0)$.

Finally, because $\Psi(\xi,\mathbf{p}^*)$ is continuously differentiable in ξ, by (13) its derivative with respect to ξ equals zero with probability one for all $N > N_\epsilon$ and for some $\xi \in N_\delta(\xi^0)$. This implies $\xi^* \in N_\delta(\xi^0)$ with probability one for all $N > N_\epsilon$, since $\Psi(\xi;\mathbf{p}^*)$ is concave in ξ and so admits a unique (global) maximizer. But δ is arbitrary, so let $\delta \to 0$ to obtain $\xi^* \xrightarrow{a.s.} \xi^0$. □

Appendix 2

Proof of Theorem 2

Let $\overline{\xi}$ be a design with a single support point k and suppose $\phi^I(k,\xi)$ is the derivative of $\Psi^I(\xi;\mathbf{p}_0)$ in the direction of $\overline{\xi}$. The general equivalence theorem states that ξ^I is optimal for Stage I if and only if $\phi^I(k,\xi^I)$ achieves its maximum of zero at the points of the design (Atkinson and Donev, 1992, p. 96). Following the notation of Atkinson and Donev (1992, p. 96), set $\mathbf{p} = \mathbf{p}_0$ and let

$$M_t^I(\xi) = [\Sigma(\mathbf{p};\lambda=1)]_{tt} = \sum_j w_j f(k_j, p_{0t}), \qquad t = 1,\ldots,T.$$

Then

$$\Psi^I(\xi;\mathbf{p}_0) = \sum_{t=1}^{T} \alpha_t \log M_t^I(\xi)$$

and

$$\begin{aligned}
\phi^I(k,\xi) &= \lim_{\gamma\to 0+} \frac{1}{\gamma}\left[\Psi^I\{(1-\gamma)\xi + \gamma\overline{\xi};\mathbf{p}_0\} - \Psi^I\{\xi;\mathbf{p}_0\}\right] \\
&= \lim_{\gamma\to 0+} \frac{1}{\gamma}\left[\sum_{t=1}^{T} \alpha_t \log\left\{(1-\gamma)M_t^I(\xi) + \gamma M_t^I(\overline{\xi})\right\} - \sum_{t=1}^{T} \alpha_t \log M_t^I(\xi)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{\gamma \to 0+} \left[\sum_{t=1}^{T} \frac{\alpha_t}{(1-\gamma) M_t^I(\xi) + \gamma M_t^I(\bar{\xi})} \left\{ -M_t^I(\xi) + M_t^I(\bar{\xi}) \right\} \right] \\
&= -1 + d^I(k, \xi),
\end{aligned}
$$

so that $\phi^I(k, \xi^I) \leq 0$ is equivalent to $d^I(k, \xi^I) \leq 1$.

Similarly, let $\phi^{II}(k, \xi)$ be the derivative of $\Psi^{II}(\xi; \mathbf{p}_0)$ in the direction of $\bar{\xi}$. The general equivalence theorem states that ξ^{II} is optimal for Stage II if and only if $\phi^{II}(k, \xi^{II})$ achieves its maximum of zero at the points of the design. Following the notation of Atkinson and Donev (1992, p. 96), and assuming the design ξ^I for Stage I is given, set $\mathbf{p} = \hat{\mathbf{p}}_1$ and let

$$
M_t^{II}(\xi) = [\boldsymbol{\Sigma}(\mathbf{p}; \lambda)]_{tt} = \lambda \sum_j w_j^I f(k_j^I, \hat{p}_{1t}) + (1-\lambda) \sum_j w_j f(k_j, \hat{p}_{1t}), \qquad t = 1, \ldots, T.
$$

Then

$$
\Psi^{II}(\xi; \hat{\mathbf{p}}_1) = \sum_{t=1}^{T} \alpha_t \log M_t^{II}(\xi)
$$

and

$$
\begin{aligned}
\phi^{II}(k, \xi) &= \lim_{\gamma \to 0+} \frac{1}{\gamma} \left[\Psi^{II}\{(1-\gamma)\xi + \gamma \bar{\xi}; \hat{\mathbf{p}}_1\} - \Psi^{II}\{\xi; \hat{\mathbf{p}}_1\} \right] \\
&= \lim_{\gamma \to 0+} \frac{1}{\gamma} \left[\sum_{t=1}^{T} \alpha_t \log \left\{ (1-\gamma) M_t^{II}(\xi) + \gamma M_t^{II}(\bar{\xi}) \right\} - \sum_{t=1}^{T} \alpha_t \log M_t^{II}(\xi) \right] \\
&= \lim_{\gamma \to 0+} \left[\sum_{t=1}^{T} \frac{\alpha_t}{(1-\gamma) M_t^{II}(\xi) + \gamma M_t^{II}(\bar{\xi})} \left\{ -M_t^{II}(\xi) + M_t^{II}(\bar{\xi}) \right\} \right] \\
&= -1 + d^{II}(k, \xi^{II}),
\end{aligned}
$$

so that $\phi^{II}(k, \xi^{II}) \leq 0$ is equivalent to $d^{II}(k, \xi^{II}) \leq 1$. □

References

Atkinson, A.C. and Bogacka, B. (1997). Compound, D- and D_s-optimum designs for determining the order of a chemical reaction. *Technometrics* **39**, 347–356.

ATKINSON, A.C. AND DONEV, A.N. (1992). *Optimum Experimental Design*, Oxford: Oxford University Press.

BEHETS, F., BERTOZZI, S., KASALI, M. et al. (1990). Successful use of pooled sera to determine HIV-1 seroprevalence in Zaire with development of cost-efficiency models. *AIDS* **4**, 737–741.

COOK, R.D. AND WONG, W.K. (1994). On the equivalence between constrained and compound optimal designs. *Journal of the American Statistical Association* **89**, 687–692.

DORFMAN, R. (1943). The detection of defective members of large populations. *Annals of Mathematical Statistics* **14**, 436–440.

EMMANUEL, J.C., BASSET, M.T., SMITH, H.J. et al. (1988). Pooling of sera for human immunodeficiency virus (HIV) testing: an economical method for use in developing countries. *Journal of Clinical Pathology*, **41**, 582–585.

HUGHES-OLIVER, J.M. AND SWALLOW, W.H. (1994). A two-stage adaptive group-testing procedure for estimating small proportions. *Journal of the American Statistical Association* **89**, 982–993.

KIEFER, J. AND WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12**, 363–366.

KLINE, R.L., BROTHERS, T.A., BROOKMEYER, R. et al. (1989). Evaluation of human immunodeficiency virus seroprevalences in population surveys using pooled sera. *Journal of Clinical Microbiology* **27**, 1449–1452.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*, New York: John Wiley & Sons, Inc.

ROSENBERGER, W.F., FLOURNOY, N., AND DURHAM, S.D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *Journal of Statistical Planning and Inference* **60**, 69–76.

ROSENBERGER, W.F. AND SRIRAM, T.N. (1997). Estimation for an adaptive allocation design. *Journal of Statistical Planning and Inference* **59**, 309–319.

RUDIN, C., BERGER, R., TOBLER, R., NARS, P.W., JUST, M., AND PAVIC, N. (1990). HIV-1, hepatitis (A, B, and C), and measles in Romanian children. *Lancet* **336**, 1592–1593.

TU, X.M., LITVAK, E. & PAGANO, M. (1995). On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV screening. *Biometrika* **82**, 287–297.

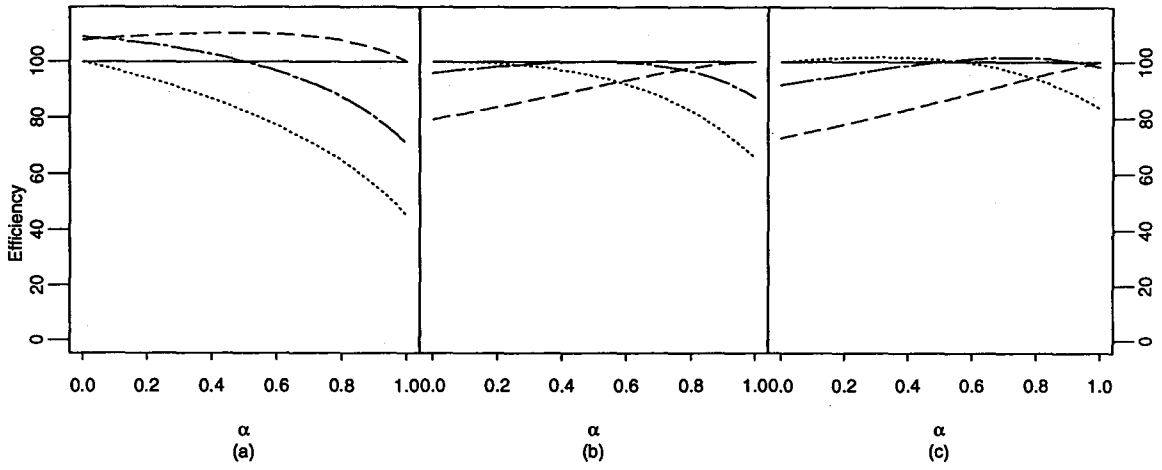

Figure 1: Efficiency of compound design ξ_C relative to ξ_1 (dashed line), ξ_2 (dotted line), and ξ_D (dotted-dashed line), for $\lambda = 0.5$, $\mathbf{p} = (0.04, 0.08)$, and $\mathbf{p}_0 = 0.5\mathbf{p}$ (a), $\mathbf{p}_0 = \mathbf{p}$ (b), $\mathbf{p}_0 = 1.5\mathbf{p}$ (c).

Table 1: Stage I and Stage II group sizes for optimal designs ξ_1, ξ_2, and ξ_D, based on $\mathbf{p} = (0.04, 0.08)$ and $\lambda = 0.5$.

	$\mathbf{p}_0 = 0.5\mathbf{p}$		$\mathbf{p}_0 = \mathbf{p}$		$\mathbf{p}_0 = 1.5\mathbf{p}$	
	k^I	k^{II}	k^I	k^{II}	k^I	k^{II}
ξ_1	78.8817	39.0384	39.0384	39.0384	25.7554	39.0384
ξ_2	39.0384	19.1124	19.1124	19.1124	12.4664	19.1124
ξ_D	51.4188	23.7537	25.2524	25.2524	16.5252	25.8279