Statistica Sinica **30** (2020), 285-302
doi:https://doi.org/10.5705/ss.202017.0408

COST CONSIDERATIONS FOR EFFICIENT GROUP TESTING STUDIES

Shih-Hao Huang[1,2,3], Mong-Na Lo Huang[4] and Kerby Shedden[3]

[1]*National Central University,* [2]*Academia Sinica,*
[3]*University of Michigan and* [4]*National Sun Yat-sen University*

Abstract: In a group testing study, the researcher collects samples from multiple individuals, pools the results, and tests them as a group. A realistic cost model for such a study should consider both the costs of collecting the samples and those of running the assays. Moreover, an efficient design should accommodate inaccuracies in any prespecified nominal test sensitivity and specificity values, allowing them to vary with the group size. We derive locally optimal designs in this setting, and characterize their theoretical properties. We also provide a guaranteed algorithm for constructing designs on discrete design spaces. Several simulated examples based on a real-world group testing study show that the proposed designs have high efficiency. In addition, the designs are not strongly sensitive to the working parameter specification used to obtain the locally optimal design.

Key words and phrases: Budget-constrained design, dilution effect, group testing.

1. Introduction

Group testing, first discussed by Dorfman (1943), plays an important role in prevalence estimation and case diagnosis, and is likely to become increasingly important in areas such as public health, environmental monitoring, and risk surveillance as the use of sensors, assays, and data-driven risk monitoring proliferates; see, for example, Gastwirth (2000), Xie et al. (2001), Pilcher et al. (2005), and Liu et al. (2011). A successful group testing study should be based on an efficient and tractable design in order to maximize the information extracted from limited resources. A critical aspect of such an efficient design is the overall study cost (Turner, Stamey and Young (2009)), made up of separate costs incurred as a result of collecting samples and running assays. Another important issue is that the specificity and especially the sensitivity of the test may decline as the group size increases. These are known as *dilution effects* in the literature (Zenios and Wein (1998); McMahan, Tebbs and Bilder (2013)).

Many group testing studies for prevalence estimation utilize prespecified val-

ues for the sensitivity and specificity. As a result, their designs involve one group size only (Tu, Litvak and Pagano (1995); Liu et al. (2012)). However, Zhang et al. (2014) indicate that misspecified sensitivity and specificity may introduce bias into a prevalence estimate. Therefore, we estimate the prevalence while treating the sensitivity and specificity as nuisance parameters inferred from the data.

Huang et al. (2017) theoretically characterize optimal designs for group testing with uncertain testing parameters. However, because they do not incorporate costs for assays and subjects, the optimal designs may place untenably many subjects into large groups. In group testing, large groups are important for sensitivity estimation, but it is arguably unlikely that scarce samples would be utilized this way in practice. Therefore, introducing differential costs for subjects and assays and, in particular, placing a realistic cost on subject recruitment can lead to optimal designs in which the largest group sizes are moderated.

Here, we develop a theory and an algorithm that yield optimal designs for prevalence estimation in a realistic group testing setting. We allow for different costs for assays and subjects, and accommodate uncertain test accuracies that vary with the group size. Our results indicate that the optimal design depends substantially on the relative costs of assays and subjects. Therefore, a simplified approach in which either assays or subjects are taken to be cost-free may not be appropriate in many cases.

2. Preliminaries

Let $\theta = (p_0, p_1, p_2)^{\mathrm{T}}$, where p_0 is the prevalence (the proportion of diseased people in the population), and p_1 and p_2 are the sensitivity and specificity (true positive rate and true negative rate of the test, respectively). We first consider the case of unknown sensitivity and specificity values that do not change with the group size. We assume that $p_0 \in (0, 1)$, $p_1, p_2 \in (0.5, 1]$, and false positives and false negatives occur randomly with rates $1 - p_2$ and $1 - p_1$, respectively. Hence, the positive response probability (either true or false positive) of a trial with group size x is

$$\pi(x) = \pi(x|\theta) = p_1 - (p_1 + p_2 - 1)\,(1 - p_0)^x\,. \tag{2.1}$$

We consider designs subject to a known group size constraint, $1 \le x_L \le x \le x_U < \infty$, where the limits on the group sizes are driven by practical considerations, such as the feasibility of the test. Note that when the upper bound x_U is sufficiently large, it is often not a support point of the optimal design in our setting. As a result, it does not affect the design or the analysis.

To introduce costs, we let the total budget be C_0, and we assume that the costs of performing an assay and enrolling a subject are, respectively, q_0 and q_1. In practice, these are known, with $q_0, q_1 \geq 0$ and $q_0 + q_1 > 0$. Without loss of generality, we rescale the total budget, the cost for an assay, and the cost for a subject in terms of the cost for an individual test, $q_0 + q_1$. That is, the (rescaled) total budget is $C = C_0/(q_0 + q_1)$, and the (rescaled) costs for an assay and for a subject are $1 - q$ and q, respectively, for $q = q_1/(q_0 + q_1) \in [0, 1]$. We then model the cost of a trial with group size x as

$$c(x) = 1 - q + qx.$$

Under a fixed budget, setting $q = 0$ means that subjects incur no cost. Therefore, this is equivalent to the scenario with a fixed number of trials. Similarly, the scenario with $q = 1$ (i.e., assays are cost-free) is equivalent to that with a fixed number of subjects.

For a study consisting of n_i trials with group size x_i, for $i = 1, \ldots, k$, we denote its *budget-constrained design* as $\xi = \{(x_i, w_i)\}_{i=1}^k$, where w_i is the proportion of the budget expended at group size x_i, expressed as

$$w_i = \frac{n_i c(x_i)}{C}, \tag{2.2}$$

and the total budget is given by $C = \sum_j n_j c(x_j)$. The log-likelihood function in θ is (omitting an unimportant additive constant)

$$\begin{aligned} L(\theta) &= \sum_{i=1}^{k} \{y_i \log(\pi(x_i|\theta)) + (n_i - y_i) \log(1 - \pi(x_i|\theta))\} \\ &= C \left(\sum_{i=1}^{k} \frac{w_i}{c(x_i)} \left\{ \frac{y_i}{n_i} \log(\pi(x_i|\theta)) + \left(1 - \frac{y_i}{n_i}\right) \log(1 - \pi(x_i|\theta)) \right\} \right). \end{aligned} \tag{2.3}$$

The maximum likelihood estimate (MLE) of θ, $\hat{\theta}$, is obtained by maximizing (2.3), and the covariance matrix of $\hat{\theta}$ is asymptotically proportional to the inverse of the information matrix of ξ, which is

$$M(\xi) = \sum_{i=1}^{k} w_i \lambda(x_i) f(x_i) f(x_i)^{\mathrm{T}}, \tag{2.4}$$

where

$$\begin{aligned} \lambda(x) &= \{c(x)\pi(x)(1 - \pi(x))\}^{-1}, \\ f(x) &= \left((p_1 + p_2 - 1)x(1 - p_0)^{x-1}, 1 - (1 - p_0)^x, -(1 - p_0)^x\right)^{\mathrm{T}}. \end{aligned}$$

Note that in equations (2.3) and (2.4), $c(x)$ plays the role of an inverse weight in

both the log-likelihood function and the information matrix.

Our main goal is to accurately estimate the prevalence, treating other unknown parameters as nuisance parameters. Therefore, we use the D_s-optimality criterion, which seeks a design that minimizes the asymptotic generalized variance of a given subset of model parameters. In this study, a D_s-optimal design maximizes

$$\Phi_s\{M(\xi)\} = -\log\left(M(\xi)^-\right)_{11} \tag{2.5}$$

among all designs under which p_0 is estimable. For a matrix M, M_{11} is the (1, 1)-entry and M^- is a generalized inverse of M. Note that the D_s-optimality above is equivalent to c-optimality with $c = (1, 0, 0)^{\mathrm{T}}$ (Atkinson, Donev and Tobias (2007, Chap. 17.5)), which minimizes the asymptotic variance of $c^{\mathrm{T}}\hat{\theta}$. The optimal group sizes of a D_s-optimal design may be nonintegers. Thus, for practical purposes, we further state that a design is D_s^I-optimal ("I" stands for "integers") if it is D_s-optimal among all designs supported on the positive integers $[x_L, x_U] \cap \mathbb{N}$. From (2.4) and (2.5), we can see that the optimality of a design depends on the unknown parameters $(p_0, p_1, p_2)^{\mathrm{T}}$ and the cost parameter q, but is invariant to the total budget C.

3. D_s-optimal Budget-constrained Designs

We first consider the design space as the interval $[x_L, x_U]$ to obtain an overview of the behavior of D_s-optimal budget-constrained designs. The main tools used in this section are the general equivalence theorem (Kiefer (1974)) and the following two lemmas. Note that the three results still hold when the design space $[x_L, x_U]$ is replaced by $[x_L, x_U] \cap \mathbb{N}$. We use these results to obtain the D_s^I-optimal designs in Section 3.1. For the D_s-criterion, we say that a design ξ with finitely many group sizes is *valid* if p_0 is estimable under ξ. The first result describes the collection Ξ of all valid designs (see Lemma 1). The proof for this lemma and other results are provided in the online Supplementary Material.

Lemma 1. *For the D_s-criterion* (2.5)*, Ξ consists of all designs with at least three support points in $[x_L, x_U]$.*

This lemma also shows that all valid designs under model (2.1) have nonsingular information matrices. Moreover, for three group sizes $x_1 < x_2 < x_3$, letting $F = (f(x_1), f(x_2), f(x_3))$, we have $(v_1, v_2, v_3)^{\mathrm{T}} = F^{-1} \cdot (1, 0, 0)^{\mathrm{T}}$, $u_i = \{\lambda(x_i)^{-1} v_i^2\}^{1/2}$, and $w_i^s = u_i / \sum_j u_j$, for $i = 1, 2, 3$. The following lemma determines the optimal weights on the three group sizes. Moreover, when a three-point

design is described by its support points, its weights are obtained from this lemma directly.

Lemma 2. *The weights* $\{w_i^s\}_{i=1}^3$ *are the unique optimal weights for the three group sizes,* $x_1 < x_2 < x_3 \in [x_L, x_U]$*, with*

$$\Phi_s\{M(\{x_i, w_i^s\}_{i=1}^3)\} = -2\log\sum_{i=1}^{3} u_i. \tag{3.1}$$

For completeness, we introduce the general equivalence theorem, as follows. For $x \in [x_L, x_U]$, let δ_x be the one-point design supported on x. For a design $\xi \in \Xi$ and $x \in [x_L, x_U]$, let $\phi_s(x, \xi)$ be the directional derivative of Φ_s at $M(\xi)$ in the direction $M(\delta_x)$:

$$\begin{aligned}\phi_s(x,\xi) &= \lim_{\alpha\to 0^+}\frac{1}{\alpha}\left(\Phi_s\{M(\xi)\} - \Phi_s\{M((1-\alpha)\xi + \alpha\delta_x)\}\right)\\ &= \lambda(x)f(x)^{\mathrm{T}}M^{-1}(\xi)f(x) - \lambda(x)f_s(x)^{\mathrm{T}}M_s^{-1}(\xi)f_s(x) - 1,\end{aligned} \tag{3.2}$$

where $f_s(x)$ is the 2×1 subvector of $f(x)$ after[18] deleting its first element, and $M_s(\xi)$ is the 2×2 submatrix of $M(\xi)$ after deleting its first row and first column. Then, we have the following general equivalence theorem.

Theorem 1. *For a design* $\xi_s \in \Xi$*, the following three assertions are equivalent:*

(a) $\Phi_s\{M(\xi_s)\} = \max_{\xi\in\Xi}\Phi_s\{M(\xi)\}$*;*

(b) $\max_{x\in[x_L,x_U]}\phi_s(x,\xi_s) = \min_{\xi\in\Xi}\max_{x\in[x_L,x_U]}\phi_s(x,\xi)$*;*

(c) for an arbitrary group size x_s *of* ξ_s*,* $\phi_s(x_s,\xi_s) = \max_{x\in[x_L,x_U]}\phi_s(x,\xi_s) = 0$*.*

Any linear combination of designs satisfying (a)–(c) also satisfies (a)–(c).

Based on Lemmas 1 and 2 and Theorem 1, we characterize the D_s-optimal design in Theorem 2. This extends Theorem 3 in Huang et al. (2017) from the special case with cost parameter $q = 0$ to an arbitrary $q \in [0, 1]$.

Theorem 2. *The* D_s*-optimal design* ξ_s *for estimating the prevalence only is unique. It has three group sizes,* $x_L = x_1^s < x_2^s < x_3^s \le x_U$*, with weights as given in Lemma* 2*.*

Theorem 2 shows that some properties of Theorem 3 in Huang et al. (2017) continue to hold for $q \in [0, 1]$: (i) the unique D_s-optimal design has exactly three group sizes; running a design with four or more distinct sizes would reduce the efficiency for prevalence estimation; (ii) the information about the prevalence,

sensitivity, and specificity mainly comes from x_2^s, x_3^s, and x_1^s, respectively; (iii) a smaller value of x_L strictly improves the accuracy of the estimation of p_0; thus, x_L should be set to one whenever possible.

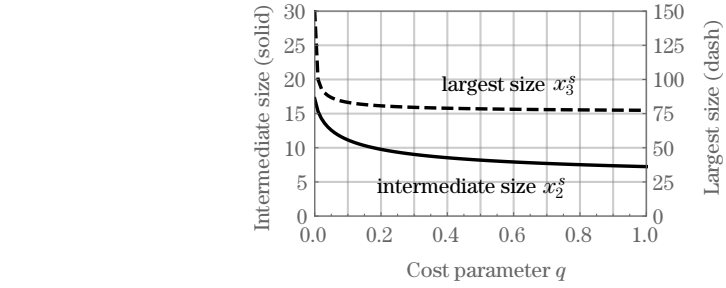
On the other hand, as q increases, the inverse weight $c(x)$ tends to penalize larger group sizes. Therefore, when $q > 0$, x_U may not be a support point of ξ_s. Thus, a two-dimensional optimization problem (x_2 and x_3 in equation (3.1)) needs to be solved in order to obtain ξ_s. In contrast, Theorem 3 in Huang et al. (2017) shows that when $q = 0$, x_3^s must be x_U, and x_2^s can be obtained using a one-dimensional root-finding algorithm.

Example 1. Let $\theta = (p_0, p_1, p_2)^{\mathrm{T}} = (0.07, 0.93, 0.96)^{\mathrm{T}}$ (this parameter setting is based on a chlamydia study described in McMahan, Tebbs and Bilder (2012)) and let $[x_L, x_U] = [1, 150]$. We obtain the D_s-optimal design for each $q \in [0, 1]$, as shown in Figure 1. First, we focus on the group sizes of the D_s-optimal designs. Theorem 2 shows that the smallest group size x_1^s of ξ_s must be the lower boundary x_L. In Figure 1(a), we observe that the intermediate and largest group sizes, x_2^s and x_3^s, respectively, decrease as the cost parameter q increases. Moreover, when x_U is as large as 150, the largest group size of a D_s-optimal design is strictly less than x_U, unless q approaches zero.

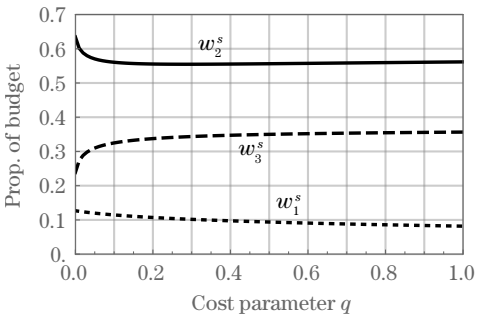
The optimal weights (proportions of the budget) $\{w_1^s, w_2^s, w_3^s\}$ of the D_s-optimal designs are shown in Figure 1(b). Under this parameter setting, the weight w_2^s at x_2^s always dominates the other two weights, w_1^s and w_3^s. From Figures 1 (a) and (b), we note that what really matters is whether $q \approx 0$. Furthermore, the D_s-optimal designs are quite stable when $q \geq 0.4$, which is roughly supported on $\{1, 7, 77\}$ with weights $\{0.09, 0.55, 0.36\}$.

Figure 1(c) shows another perspective on trial allocation related to cost. Roughly speaking, we find that as the cost parameter q increases, only the proportion of trials t_1^s at x_1^s increases, whereas the other two proportions decrease. Comparing Figures 1(b) and (c), we conclude that as q increases, a trial with a large group size becomes more expensive, and a greater proportion of the budget will be spent on groups with larger sizes in order to obtain sufficient information about p_1 to efficiently estimate p_0. On the other hand, the proportion of trials at the smallest size still increases, reflecting a preference for less expensive trials.

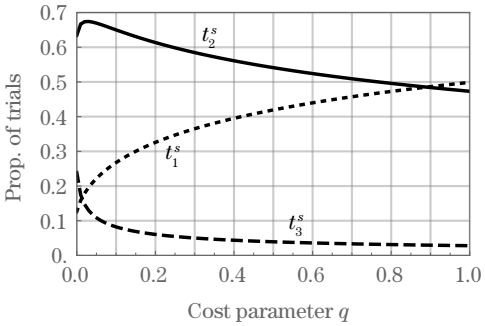
Remark 1. In the online Supplementary Material, Section S2, we also consider the D-optimal design under the setting of Example 1, where the D-criterion treats p_0, p_1, and p_2 as equally important. The D-optimal design also has exactly three group sizes, with a low boundary x_L and an intermediate size close to that of the



(a) Group sizes versus q



(b) Proportions of budget versus q



(c) Proportions of trials versus q

Figure 1. Properties of the D_s-optimal designs for Example 1.

D_s-optimal design. However, a D_s-optimal design places much greater weight (proportion of the budget) on its intermediate size (≥ 0.55 vs. 0.33).

3.1. D_s^I-optimal designs

In practice, the group sizes in a group testing design must be supported on the finite set $[x_L, x_U] \cap \mathbb{N}$, rather than on the interval $[x_L, x_U]$. To obtain the optimal integer-valued group sizes, a natural approach would be to simply round the D_s-optimal design ξ_s. However, to attain optimality, we develop an efficient numerical search procedure that yields the D_s^I-optimal designs on $[x_L, x_U] \cap \mathbb{N}$.

Intuitively, a D_s^I-optimal design should be close to the corresponding D_s-optimal design ξ_s obtained from Theorem 2. Therefore, the three support points of ξ_s, after rounding, form a good initial design. Then, by Theorem 1, we know that either the initial design is optimal, or that it can be improved by adding a point that has a positive derivative (3.2). We then recalculate the weights by numerically optimizing (2.5). After excluding points with a zero weight, if any exist, we check the optimality of the new design. These steps are iterated until

optimality is attained.

The algorithm stops when the resulting design satisfies Theorem 1(c), which guarantees optimality; otherwise, the design obtained in each iteration is strictly better than the previous designs. Because $[x_L, x_U] \cap \mathbb{N}$ is finite, this algorithm must stop in a finite number of steps. In addition, owing to the convexity of the design criterion, this stepwise ascent algorithm converges to a global optimum. The details of the search algorithm for obtaining a D_s^I-optimal design ξ_I are provided in Section 4, together with the scenario with dilution effects, as Algorithm 1. In practice, the algorithm tends to converge in very few steps, because the initial design is often already close to (and in many cases, exactly equal to) a D_s^I-optimal design.

Note that heuristic numerical search (e.g., Zhang et al. (2014)) may yield incorrect results. When $\theta = (0.05, 0.95, 0.995)^{\mathrm{T}}$, $q = 1$, and $[x_L, x_U] = [1, 150]$, our algorithm obtains a D_s^I-optimal design with group sizes $\{1, 8, 113\}$. In contrast, a design supported on $\{1, 12, 150\}$ is reported by Zhang et al. (2014). From Theorem 1, our design is optimal, whereas the other is not.

3.2. Design implementation

In the approximate design framework, the optimal weights include the constraints $w_i > 0$ and $\sum w_i = 1$ only. For practical use, with a total budget C, equation (2.2) shows that the number of trials at each point x_i is $n_i = Cw_i/c(x_i)$, which should be a positive integer, introducing additional restrictions on the weights.

To implement a D_s^I-optimal design ξ_I, we obtain the number of trials using a variant of the efficient rounding procedure (Pukelsheim (2006)). For a budget-constrained design $\xi = \{(x_i, w_i)\}_{i=1}^k$ and a total budget C, let $\{n_i^0\}_{i=1}^k = \{\lfloor Cw_i/c(x_i) \rfloor\}_{i=1}^k$ and $C_1 = C - \sum_i n_i^0 c(x_i)$, where $\lfloor x \rfloor$ is the largest integer not greater than x. Then, we allocate C_1 at each x_i to obtain a design with trial counts $\{n_i^0 + \Delta_i\}_{i=1}^k$ with minimum variance of the prevalence estimator, where $\Delta_i \in \mathbb{N} \cup \{0\}$ for each i and $\sum \Delta_i c(x_i) \le C_1$. Note that $\sum \Delta_i c(x_i)$ may be strictly less than C_1 when the remaining cost is less than $c(\min(x_i))$. This is illustrated in the following example.

Example 2. Following Example 1, let $\theta = (0.07, 0.93, 0.96)^{\mathrm{T}}$, $q = 0.2$, and $[x_L, x_U] \cap \mathbb{N} = \{1, 2, \ldots, 150\}$. A D_s^I-optimal design ξ_I is supported on {1, 10, 81} with weights {0.104, 0.555, 0.341} and costs per test {1, 2.8, 17}. The asymptotic variance of the prevalence estimate from ξ_I is $0.137633/C$.

When the total budget is C = 10,000, we have $\{n_i^0\}_{i=1}^3$ = {1,042, 1,981,

Table 1. Allocations of C_1 on support points $\{1, 10, 81\}$ when $C = 10,000$.

x_i	1	10	81	remaining	$\mathrm{Var}(\hat{p}_0)$
$c(x_i)$	1	2.8	17	budget	$(\times 1/C)$
additional	0	4	0	0.0	0.137634
trials Δ_i	2	3	0	0.8	0.137645
	5	2	0	0.6	0.137642
	8	1	0	0.4	0.137640
	11	0	0	0.2	0.137638

200$\}$ and $C_1 = 11.2$. Table 1 shows all possible allocations of C_1. The variance attains a minimum when the additional trials are at $\{0, 4, 0\}$. Thus, we set the numbers of trials of the implemented design $\xi_I(C)$ to $\{1{,}042, 1{,}985, 200\}$, with a total number of trials of 3,227, and a total number of subjects of 37,092. Note that when C is sufficiently large, as in this example, the loss of design efficiency tends to be negligible, regardless of how we allocate C_1 in Table 1.

4. D_s^I-optimal Designs Under Dilution Effects

In Section 3, we treated the sensitivity and specificity as constants with unknown values. As noted in the introduction, dilution effects, which reduce the sensitivity or specificity for larger group sizes, are common, especially when the allowable range of group sizes $[x_L, x_U]$ is wide. In this section, we provide an algorithm for group testing with dilution effects.

The most natural form of dilution is the decrease in sensitivity with an increase in group size (Zenios and Wein (1998)). For completeness, we also consider the presence of diluted specificity. When there is a dilution effect on the sensitivity or specificity, we work with the models,

$$p_1(x) = p_1(x|\alpha) = \mathit{link}(\alpha_0 - \alpha_1 \log(x)) \quad \text{and} \tag{4.1}$$

$$p_2(x) = p_2(x|\beta) = \mathit{link}(\beta_0 - \beta_1 \log(x)), \tag{4.2}$$

respectively, where $\mathit{link} : \mathbb{R} \to [0, 1]$ is a link function for probability. For convenience of interpreting the dilution models, we adopt a logistic regression in the following context: $\mathit{link}(u) = \mathrm{expit}(u) = \{1 + \exp(-u)\}^{-1}$ (see equation (4) in Zhang et al. (2014)). Thus, for instance, for the sensitivity model, $\mathrm{expit}(\alpha_0)$ is the baseline sensitivity $p_1(1)$, and the sensitivity has a nearly polynomial rate of decay, $\{1 + x^{\alpha_1} \exp(-\alpha_0)\}^{-1}$, as the group size x increases. In other scenarios, $\log(x)$ in equations (4.1) and (4.2) can be replaced by x, $\log^2(x)$, and so on, and another link function can be adopted. Here, we assume that $\alpha_0, \beta_0 > 0$ and

$\alpha_1, \beta_1 \geq 0$, such that $p_1(1), p_2(1) > 0.5$ and p_1, p_2 decrease monotonically as x increases.

When only the sensitivity has a dilution effect, the corresponding information matrix becomes a variant of (2.4):

$$M_\alpha(\xi) = \sum_{i=1}^{k} w_i \lambda(x_i) f_\alpha(x_i) f_\alpha(x_i)^{\mathrm{T}}, \tag{4.3}$$

where $f_\alpha(x) = H_\alpha(x) f(x) \in \mathbb{R}^4$, and $H_\alpha(x)$ is a 4×3 block-diagonal matrix with diagonal blocks $(1, \partial p_1(x)/\partial\alpha, 1)$. Similarly, when only the specificity has a dilution effect, or when both the sensitivity and the specificity have dilution effects, the corresponding information matrices are:

$$M_\beta(\xi) = \sum_{i=1}^{k} w_i \lambda(x_i) f_\beta(x_i) f_\beta(x_i)^{\mathrm{T}}, \quad \text{and} \tag{4.4}$$

$$M_{\alpha\beta}(\xi) = \sum_{i=1}^{k} w_i \lambda(x_i) f_{\alpha\beta}(x_i) f_{\alpha\beta}(x_i)^{\mathrm{T}}, \tag{4.5}$$

respectively, where $f_\beta(x) = H_\beta(x) f(x) \in \mathbb{R}^4$, $f_{\alpha\beta}(x) = H_{\alpha\beta}(x) f(x) \in \mathbb{R}^5$, $H_\beta(x) = \mathrm{diag}(1, 1, \partial p_2(x)/\partial\beta)$, and $H_{\alpha\beta}(x) = \mathrm{diag}(1, \partial p_1(x)/\partial\alpha, \partial p_2(x)/\partial\beta)$.

Extending the ideas in Section 3.1, our search algorithm is described as follows. From Theorem 2, the D_s-optimal design supported on $\{x_1^s, x_2^s, x_3^s\}$ can be used to efficiently estimate p_0 in the absence of dilution effects. However, when dilution effects exist, additional points should be added. Note that in Theorem 2, the information on p_1 and p_2 comes mainly from the larger and smaller group sizes, x_3^s and x_1^s, respectively. Therefore, to form an initial design, we add a size between x_2^s and x_3^s (or a size in (x_1^s, x_2^s)) if the sensitivity (or specificity) is diluted. The optimal weights for these group sizes can be obtained using Lemma 2 when there is no dilution effect, or using the following lemma when dilution effects do exist.

Lemma 3.

(a) For group testing with one dilution effect (either sensitivity or specificity), if the four distinct sizes $\{x_1, x_2, x_3, x_4\}$ are such that $F_* = (f_*(x_1), f_*(x_2), f_*(x_3), f_*(x_4))$ is invertible, where $f_* = f_\alpha$ or f_β, then the D_s-optimal weights at these sizes are proportional to $(\lambda(x_i)^{-1} v_i^2)^{1/2}$, for $i = 1, \ldots, 4$, where $(v_1, v_2, v_3, v_4) = F_*^{-1} \cdot (1, 0, 0, 0)^{\mathrm{T}}$.

(b) For group testing with two dilution effects (both sensitivity and specificity), if the five distinct sizes $\{x_1, x_2, x_3, x_4, x_5\}$ are such that $F_{\alpha\beta} = (f_{\alpha\beta}(x_1),$

$f_{\alpha\beta}(x_2), f_{\alpha\beta}(x_3), f_{\alpha\beta}(x_4), f_{\alpha\beta}(x_5))$ is invertible, the D_s-optimal weights at these sizes are proportional to $(\lambda(x_i)^{-1}v_i^2)^{1/2}$, for $i = 1, \ldots, 5$, where $(v_1, v_2, v_3, v_4, v_5) = F_{\alpha\beta}^{-1} \cdot (1, 0, 0, 0, 0)^{\mathrm{T}}$.

Based on the ideas above and applying Theorem 1 (where the information matrix (2.4) should be replaced by M_α, M_β, or $M_{\alpha\beta}$ if dilution effects exist), Algorithm 1 yields the D_s^I-optimal designs. The use of Algorithm 1 is demonstrated in the subsequent example.

Algorithm 1. Let $\Omega = [x_L, x_U] \cap \mathbb{N}$ and let $x_1^{(0)} < x_2^{(0)} < x_3^{(0)}$ be the three support points of ξ_s for $\theta = (p_0, p_1(1), p_2(1))^{\mathrm{T}}$ in Theorem 2, after rounding. Set $X_0 = \left\{x_1^{(0)}, x_\beta^{(0)}(\text{if } p_2 \text{ is diluted}), x_2^{(0)}, x_\alpha^{(0)}(\text{if } p_1 \text{ is diluted}), x_3^{(0)}\right\}$, where $x_\alpha^{(0)} = \lfloor (x_2^{(0)} + x_3^{(0)})/2 \rfloor$ and $x_\beta^{(0)} = \lfloor (x_1^{(0)} + x_2^{(0)})/2 \rfloor$. Set W_0 as the optimal weights obtained from Lemma 2 or Lemma 3 at the points in X_0. Set $\xi_0 = \{X_0, W_0\}$. For $j = 0, 1, \ldots,$ do:

Step 1. Set $x_j = \arg\max_{\Omega \setminus X_j} \phi_s(x, \xi_j)$. If $\phi_s(x_j, \xi_j) \le 0$, output ξ_j and stop.

Step 2. Set $X_{j+1} = X_j \cup \{x_j\}$, and obtain

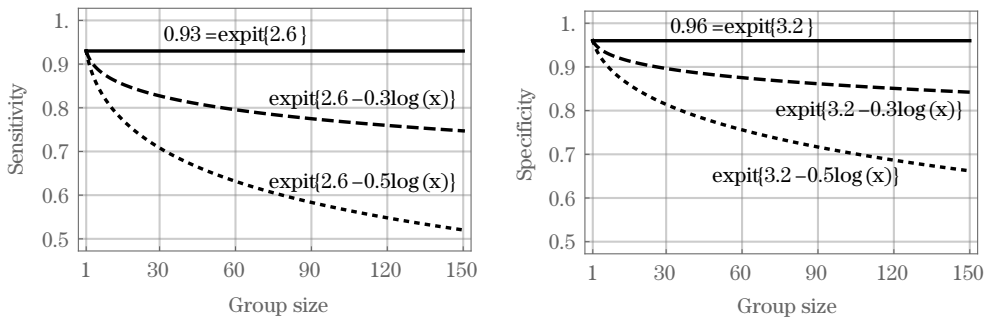
$$W_{j+1} = \arg\max_W \left\{ \Phi_s\{M(X_{j+1}, W)\};\ \min(W) \ge 0,\ \sum_{w \in W} w = 1 \right\}.$$

The weights W are available in closed form (Lemmas 2 and 3) if the design is minimally supported. Otherwise, W can be obtained by solving a convex optimization.

Step 3. Set $\xi_{j+1} = \{X_{j+1}, W_{j+1}\}$ after deleting those (x, w) with $w = 0$.

Example 3. In order to better understand the structure of the optimal designs in the presence of dilution effects, and how they relate to such designs when there is no dilution, we provide several numerical examples. Following Examples 1 and 2, we let $p_0 = 0.07$, $[x_L, x_U] \cap \mathbb{N} = \{1, 2, \ldots, 150\}$, and $q = 0.2$. We further let the sensitivity and the specificity be 0.93 and 0.96, respectively, at group size 1 ($\alpha_0 = 2.6$ and $\beta_0 = 3.2$), and let α_1 and β_1 vary from 0 to 0.5. Figure 2 shows how the sensitivity and specificity decay as the group size increases.

Table 2 shows the D_s^I-optimal designs for several settings, with and without dilution effects. When there is no dilution effect, the design supported on $\{1, 10, 81\}$ is D_s^I-optimal under the model with information matrix (2.4). When the experimenters include dilution effects in the group testing model, the information matrix becomes (4.3), (4.4), or (4.5).



(a) Sensitivity functions

(b) Specificity functions

Figure 2. The sensitivity and specificity functions for Example 3.

If the sensitivity is diluted, the new support point falls between x_2 and x_3, but does not approach either of the two. However, if the specificity is diluted, the new support point falls near the lower end of the range of group sizes. This is consistent with the fact that larger group sizes are more informative about sensitivity, whereas smaller group sizes are more informative about specificity. However, the new support points cannot approach the extremes of the range of allowable group sizes, because these points are already included in the design, and we need to observe results for sufficiently many distinct group sizes to be able to estimate the slope parameters α_1 and β_1.

We also considered how the population parameters for dilution effects influence the structure of the optimal designs. As the slope parameter α_1 increases, x_2 and x_α tend to decrease. However, when the slope parameter β_1 increases, x_β and x_2 tend to increase. These changes may allow for improved estimations of the sensitivity or specificity curves. However, because we are using the D_s-criterion that focuses on prevalence, the changes are not significant.

In the example above, it seems that the upper bound x_U is always present in a D_s^I-optimal design when the sensitivity is diluted. However, x_U is not necessarily present, especially when x_U is sufficiently large. For instance, under the same parameter setting as that in the previous example, with $\alpha_1 = \beta_1 = 0.5$, and moving x_U up to 1,000, the D_s^I-optimal design is supported on $\{1, 3, 14, 51, 674\}$.

5. Design Performance

In this section, we study the performance of the D_s^I-optimal design when the working parameter is moderately misspecified. We can see below that its perfor-

Table 2. D_s^I-optimal designs for several scenarios in Example 3. (Here, $\alpha_0 = 2.6$ and $\beta_0 = 3.2$).

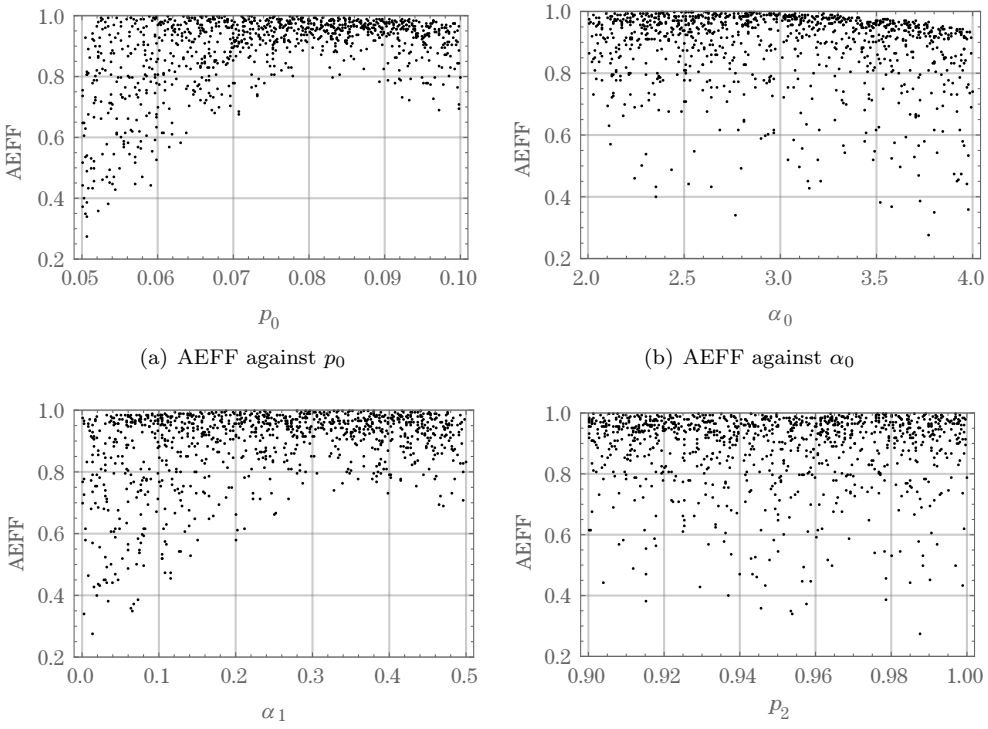
Model	α_1	β_1	x_1	x_β	x_2	x_α	x_3	w_1	w_β	w_2	w_α	w_3
(2.4)	–	–	1	–	10	–	81	0.11	–	0.55	–	0.34
(4.3)	0.0	–	1	–	11	53	150	0.04	–	0.27	0.38	0.31
	0.3	–	1	–	7	44	150	0.08	–	0.30	0.33	0.29
	0.5	–	1	–	6	38	150	0.10	–	0.31	0.31	0.28
(4.4)	–	0.0	1	3	18	–	89	0.07	0.19	0.44	–	0.30
	–	0.3	1	3	20	–	91	0.11	0.25	0.38	–	0.26
	–	0.5	1	3	22	–	92	0.15	0.27	0.34	–	0.24
(4.5)	0.0	0.0	1	3	15	57	150	0.05	0.14	0.29	0.31	0.21
	0.0	0.3	1	3	17	58	150	0.09	0.20	0.29	0.26	0.16
	0.0	0.5	1	3	18	58	150	0.13	0.24	0.27	0.23	0.13
	0.3	0.0	1	2	13	51	150	0.08	0.17	0.25	0.29	0.21
	0.3	0.3	1	3	14	52	150	0.09	0.21	0.27	0.26	0.17
	0.3	0.5	1	3	15	53	150	0.12	0.24	0.25	0.24	0.15
	0.5	0.0	1	2	12	46	150	0.08	0.17	0.24	0.29	0.22
	0.5	0.3	1	3	13	47	150	0.08	0.21	0.26	0.27	0.18
	0.5	0.5	1	3	13	46	150	0.12	0.25	0.25	0.23	0.15

Table 3. AEFF($\tilde{\xi}|\theta$) for selected $\theta \in \Theta$.

		$p_0 = 0.05$		$p_0 = 0.10$	
$p_1(x) = \text{expit}\{\alpha_0 - \alpha_1 \log(x)\}$		$p_2 = 0.9$	$p_2 = 1.0$	$p_2 = 0.9$	$p_2 = 1.0$
$\alpha_0 = 2$	$\alpha_1 = 0.0$	0.363	0.393	0.908	0.902
	$\alpha_1 = 0.1$	0.595	0.636	0.885	0.872
	$\alpha_1 = 0.5$	0.974	0.925	0.613	0.591
$\alpha_0 = 4$	$\alpha_1 = 0.0$	0.197	0.223	0.920	0.945
	$\alpha_1 = 0.1$	0.349	0.396	0.936	0.955
	$\alpha_1 = 0.5$	0.761	0.826	0.892	0.881

mance is relatively stable when the working parameter is not too different to the true value. Following Examples 1–3, and focusing on the most common setting where only the sensitivity is diluted, we let $[x_L, x_U] \cap \mathbb{N} = \{1, 2, \ldots, 150\}$ and $q = 0.2$, and let the working parameter $\tilde{\theta}_0 = (\tilde{p}_0, \tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{p}_2)^{\mathrm{T}} = (0.07, 2.6, 0.3, 0.96)^{\mathrm{T}}$. From Table 2 (Model (4.3), $\alpha_1 = 0.3$), the D_s^I-optimal design $\tilde{\xi}$ under $\tilde{\theta}$ is supported on {1, 7, 44, 150}.

In order to examine how a misspecified working parameter affects the performance of $\tilde{\xi}$, we consider that the true value of $\theta = \{p_0, \alpha_0, \alpha_1, p_2\}^{\mathrm{T}}$ comes from $\Theta = [0.05, 0.1] \times [2, 4] \times [0, 0.5] \times [0.9, 1]$, which covers $\tilde{\theta}$. The performance of $\tilde{\xi}$ under the true value of $\theta \in \Theta$ is measured by



(a) AEFF against p_0

(b) AEFF against α_0

(c) AEFF against α_1

(d) AEFF against p_2

Figure 3. AEFF($\tilde{\xi}|\theta$) under different θ for 1,000 draws.

$$\text{AEFF}(\tilde{\xi}|\theta) = \frac{\text{AMSE}(\tilde{\xi}|\theta)}{\text{AMSE}(\xi_\theta^I|\theta)} \in [0,1],$$

where ξ_θ^I is the D_s^I-optimal design under θ, and $\text{AMSE}(\xi|\theta) = M(\xi|\theta)_{11}^{-1}$ is the (scaled) asymptotic mean squared error (AMSE) of the prevalence estimator under ξ, which is also its (scaled) asymptotic variance.

Table 3 shows AEFF($\xi|\theta$) for selected $\theta \in \Theta$, and Figure 3 shows AEFF($\xi|\theta$) for θ drawn randomly from Θ. Under this parameter setting, the accuracies of the prespecified p_0 and α_1 are important factors within this range of parameters. Figure 4 further shows how the true values of p_0 and α_1 affect the performance of $\tilde{\xi}$. Note that when $\tilde{p}_0$ and $\tilde{\alpha}_1$ are misspecified in the same direction, especially when both are over-specified, the AEFF decreases rapidly. Roughly speaking, when the true value of $\theta \in \Theta$ falls between the two dashed lines in Figure 4, $\tilde{\xi}$ performs well, with an AEFF close to or greater than 80%.

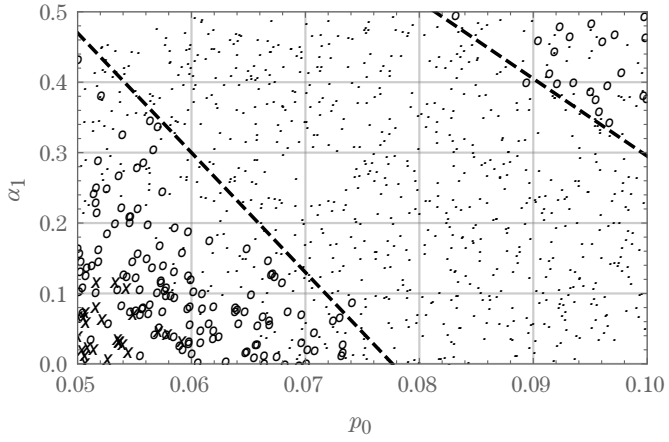


Figure 4. AEFF($\tilde{\xi}|\theta$) vs. various true values of p_0 and α_1 ($\cdot$: AEFF $\geq 80\%$; $\circ$: AEFF $\in [50\%, 80\%)$; $\times$: AEFF $< 50\%$), where $\tilde{p}_0 = 0.07$ and $\tilde{\alpha}_1 = 0.3$.

6. Conclusion

In this work, we develop efficient group testing designs that accommodate real-world complexities, including differing subject and assay costs, and uncertain sensitivity and specificity, which may include dilution effects. We characterize these designs and present an algorithm that is guaranteed to yield an optimal design on a discrete design space, as is encountered in practice. We found that accounting for subject costs yields designs with a smaller maximum group size than those of previously published optimal designs in which the subjects were considered to be cost-free (Huang et al. (2017)). Our results reveal that if the ratio of the cost per subject relative to the cost per assay increases even moderately, the largest group size of the resulting design and its proportion of trials decrease rapidly, although its proportion of the budget still increases.

As a practical illustration, we provided examples of optimal allocations, with integer-valued trials at the optimal group sizes. Although the locally optimal designs depend on the working parameters, our results based on a real-world setting show that the proposed designs are robust against a misspecification of the working parameters and exhibit good asymptotic efficiency. When there are major concerns about a possible misspecification of the working parameters, our optimal designs can be utilized with a multistage adaptive approach (Hughes-Oliver and Swallow (1994)). In the first stage, the working parameters may be specified using domain knowledge, and in subsequent stages, they are estimated from the previous stages. Alternatively, a Bayesian or minimax optimal design approach

(Dette et al. (2014)) can be adopted. Here, a Bayesian approach seeks designs that maximize the D_s-optimality criterion (2.5), averaged over the parameters with respect to a prior distribution, and a minimax approach minimizes the largest possible variance of the prevalence estimator.

The most flexible model for group testing would allow the sensitivity and specificity to be estimated from the data and, potentially, to vary with the group size. However, the sensitivity and specificity parameters are nuisance parameters in practice, and are nonorthogonal to the prevalence, which is the primary parameter of interest. As a result, estimating these nuisance parameters increases the variance of the prevalence estimate, but eliminates any bias that would result from misspecifying them in a "plug-in" approach. The increase in variance is large for small numbers of trials. Therefore, it is unlikely to be favorable to estimate the sensitivity and specificity parameters in practice if the budget is small. However, if the budget is sufficiently large, the risk of bias due to a misspecification dominates the increase in variance due to the additional parameter estimation. Therefore, our results provide guidance to practitioners, suggesting that for smaller-scale research, a plug-in approach may be suitable, but that researchers conducting larger studies should consider allowing the sensitivity and specificity parameters to be estimated from the data.

Increased interest in near real-time safety monitoring for disease epidemics, terror attacks, food safety, and environmental risks may provide new opportunities for group testing in future. If the cost considerations differ from the disease prevalence estimation that has dominated group testing to date, larger pools or larger total sample sizes may be practical. This could provide a setting in which the additional cost of estimating the dilution effects along with the prevalence is modest. Our results may also be applied to evaluate the feasibility of such a procedure.

Supplementary Material

The online Supplementary Material provides technical proofs of the theorems and lemmas, as well as a discussion on D-optimal group testing designs under cost considerations.

Acknowledgments

The first two authors were supported by the Ministry of Science and Technology, Taiwan, with grants MOST 105-2917-I-564-017 and MOST 103-2118-M-

110-002-MY2, respectively. The third author was supported by the National Center of Theoretical Sciences, Taiwan.

References

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Dette, H., Kiss, C., Benda, N. and Bretz, F. (2014). Optimal designs for dose finding studies with an active control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76**, 265–295.

Dorfman, R. (1943). The detection of defective members of large populations. *Ann. Math. Statist.* **14**, 436–440.

Gastwirth, J. L. (2000). The efficiency of pooling in the detection of rare mutations. *Amer. J. Hum. Genet.* **67**, 1036–1039.

Hughes-Oliver, J. M. and Swallow, W. H. (1994). A two-stage adaptive group-testing procedure for estimating small proportions. *J. Amer. Statist. Assoc.* **89**, 982–993.

Huang, S.-H., Huang, M.-N. L., Shedden, K. and Wong, W. K. (2017). Optimal group testing designs for estimating prevalence with uncertain testing errors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79**, 1547–1563.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849–879.

Liu, A., Liu, C., Zhang, Z. and Albert, P. S. (2012). Optimality of group testing in the presence of misclassification. *Biometrika* **99**, 245–251.

Liu, S.-C., Chiang, K.-S., Lin, C.-H., Chung, W. C., Lin, S.-H. and Yang, T. C. (2011). Cost analysis in choosing group size when group testing for Potato virus Y in the presence of classification errors. *Ann. Appl. Biol.* **159**, 491–502.

McMahan, C. S., Tebbs, J. M. and Bilder, C. R. (2012). Informative Dorfman Screening. *Biometrics* **68**, 287–296.

McMahan, C. S., Tebbs, J. M. and Bilder, C. R. (2013). Regression models for group testing data with pool dilution effects. *Biostatistics* **14**, 284–298.

Pilcher, C., Fiscus, S., Nguyen, T., Foust, E., Wolf, L., Williams, D., Ashby, R., O'Dowd, J., McPherson, J., Stalzer, B., Hightow, L., Miller, W., Eron, J., Cohen, M. and Leone, P. (2005). Detection of acute infections during HIV testing in North Carolina. *N. Engl. J. Med.* **352**, 1873–1883.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Turner, D. W., Stamey, J. D. and Young, D. M. (2009). Classic group testing with cost for grouping and testing. *Comput. Math. Appl.* **58**, 1930–1935.

Tu, X. M., Litvak, E. and 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.

Xie, M., Tatsuoka, K., Sacks, J. and Young, S. S. (2001). Group testing with blockers and synergism. *J. Amer. Statist. Assoc.* **96**, 92–102.

Zenios, S. A. and Wein, L. M. (1998). Pooled testing for HIV prevalence estimation: Exploiting the dilution effect. *Stat. Med.* **17**, 1447–1467.

Zhang, Z., Liu, C., Kim, S. and Liu, A. (2014). Prevalence estimation subject to misclassification: the mis-substitution bias and some remedies. *Stat. Med.* **33**, 4482–4500.

Department of Mathematics, National Central University, Taoyuan 32001, Taiwan.
E-mail: shhuang@math.ncu.edu.tw
Department of Applied Math., National Sun Yat-sen University, Kaohsiung 80424, Taiwan.
E-mail: lomn@math.nsysu.edu.tw
Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA.
E-mail: kshedden@umich.edu

(Received September 2017; accepted March 2018)