Author/s:
Balding, DJ;Torney, DC

Title:
Optimal pooling designs with error detection

Date:
1996-01-01

Citation:
Balding, D. J. & Torney, D. C. (1996). Optimal pooling designs with error detection. Journal of Combinatorial Theory Series A, 74 (1), pp.131-140. https://doi.org/10.1006/jcta.1996.0041.

Persistent Link:
https://hdl.handle.net/11343/52565

Optimal Pooling Designs with Error Detection

David J. Balding
School of Mathematical Sciences
Queen Mary & Westfield College
University of London
Mile End Road, London E1 4NS, UK

David C. Torney
Theoretical Biology and Biophysics
T-10 Mail Stop K-710
Los Alamos National Laboratory
Los Alamos NM 87545 USA

February 1, 2008

Abstract

Consider a collection of objects, some of which may be 'bad', and a test which determines whether or not a given sub-collection contains no bad objects. The non-adaptive pooling (or group testing) problem involves identifying the bad objects using the least number of tests applied in parallel. The 'hypergeometric' case occurs when an upper bound on the number of bad objects is known *a priori*. Here, practical considerations lead us to impose the additional requirement of *a posteriori* confirmation that the bound is satisfied. A generalization of the problem in which occasional errors in the test outcomes can occur is also considered. Optimal solutions to the general problem are shown to be equivalent to maximum-size collections of subsets of a finite set satisfying a union condition which generalizes that considered by Erdös *et al.* [3]. Lower bounds on the number of tests required are derived when the number of bad objects is believed to be either 1 or 2. Steiner systems are shown to be optimal solutions in some cases.

1 Introduction

Each of n objects has an unknown binary status, 'good' or 'bad'. A test is available which, except for occasional failures or errors, establishes whether or not all the objects in a given collection are good. The problem is to resolve the status of each object using the minimum number of tests applied in parallel. The corresponding adaptive problem, in which the choice of test at any stage can depend on the outcomes of previous tests, is sometimes known as 'group testing' (Wolf [8]).

The objects may, for example, be electronic devices which can be tested in series. Another example involves items in a database which are categorized by a sequence of binary classifications and the task is to partition the objects according to the ith classification. The problem is formally similar to that of devising optimal error-correcting codes using parity checks, except that here the test result is 'at least one bad object' rather than 'an odd number of bad objects'. Our work is motivated by an optimal design problem for large-scale experiments aimed at constructing physical maps of human chromosomes: the objects are chromosome fragments which are 'bad' if they contain a certain DNA sequence. An experimental test known as the Polymerase Chain Reaction can determine whether or not a collection of chromosome fragments are all good. In order to facilitate automation, it is desirable that the experiments be applied in parallel so that the experimental design is non-adaptive, or one-stage. Here, we derive experimental designs which, with high probability, are one-stage solutions to an appropriate formalization of the problem. These designs may form stages in solutions to more general problems, for example adaptive (multi-stage) designs which are optimal subject to a cost function which penalizes additional stages.

A *pool* is a set of objects and a *design* is a set of pools. Given a design $\mathcal{D}$, let v denote the number of pools, so that $v \equiv |\mathcal{D}|$. We will say that a pool is *good* if all the objects in it are good, otherwise it is *bad*. Let P denote the total number of bad objects. The test usually distinguishes good pools from bad, but we will also allow the possibility that for some pools the test fails to produce a result and write Q for the number of pools in $\mathcal{D}$ which fail. Before applying the tests P and Q are unknown, but we may have some prior information about them. One simple design consists of testing each object individually a fixed number of times. However if both $P \ll n$ and $Q \ll v$ then 'better' designs are possible.

There are several reasonable optimality criteria for $\mathcal{D}$. An appropriate choice will depend in part on the prior knowledge of P and Q. Bush *et al.* [2] and Hwang & Sós [4] discuss non-adaptive group testing in the 'hypergeometric' case, in which $Q = 0$ and P is bounded above by a known constant p. They define $\mathcal{D}$ to be an optimal solution if it maximizes n for fixed v among designs such that the status of each object can be inferred from the pool outcomes. The

hypergeometric formulation has the drawback that it assumes that the event $P > p$ is excluded *a priori*. It is not in general possible to confirm *a posteriori* that $P \leq p$ and hence false conclusions may be drawn if, unexpectedly, $P > p$. In practice, a large value of p must be chosen to exclude this possibility. Here, we modify the hypergeometric case by imposing the additional requirement that the event $P > p$ can be distinguished *a posteriori*. Consequently, it will be reasonable in practice to allow a small prior probability that $P > p$. Typically, lower values of p can be chosen than under the hypergeometric formulation and hence more efficient designs constructed. The price for these advantages is that the designs are not strictly non-adaptive: with small probability a second stage will be required.

Allowing also for up to q failures, we define $\mathcal{D}$ to be an optimal solution if it maximizes n for fixed v subject to the requirement that whenever $Q \leq q$ we can infer from the pool outcomes either the status of each object or that $P > p$. Proposition 1 of section 2 establishes that optimal solutions $\mathcal{D}$ are equivalent to maximum-size collections of subsets of a v-set such that every subset in the collection has more than q elements distinct from any union of up to p others. This condition is equivalent to that of q-error detection and hence optimal q-failure designs are also optimal q-error-detecting designs. In the $q = 0$ case, we require that no subset in the collection is contained in the union of p others. Hwang & Sós [4] showed that this requirement characterizes the p-complete designs defined by Bush *et al.* [2].

In Theorems 1 and 2 we establish lower bounds on v as a function of n for $p = 1$ and 2 and all $q \geq 0$. The bounds coincide in some cases with the sizes of certain Steiner system solutions which hence are optimal. These results extend the results of Erdös *et al.* [3] who considered the case $p = 2$ and $q = 0$. These authors initially constrained the designs to be uniform, that is each object occurs in the same number of pools. They subsequently derived an asymptotic bound in the unconstrained case. Here, we do not require uniformity but we note that the bounds given in Theorems 1 and 2 can only be achieved by uniform designs. Ruszinkó [7] derives asymptotic bounds for $q = 0$ and arbitrary p, but in the case $p = 2$ the bound obtained by Erdös *et al.* [3] is tighter.

2 Definitions and statement of results

For positive integers $0 \leq i \leq j$, let $\mathcal{X}_j$ denote the set of subsets of $\{1, 2, \ldots, j\}$ and define

$$\mathcal{X}_j^i \equiv \{C \in \mathcal{X}_j : |C| = i\}. \tag{1}$$

Pools are elements of $\mathcal{X}_n$ and designs are subsets of $\mathcal{X}_n$. Given a design $\mathcal{D} \equiv \{A_1, \ldots, A_v\}$ we will write $\hat{\mathcal{D}} \equiv \{B_1, \ldots, B_n\}$ for the *dual* of $\mathcal{D}$ defined by

$i \in A_j$ if and only if $j \in B_i$. Thus A_j indexes the objects in the jth pool whereas B_i indexes the pools which contain the ith object. Let $\phi(A)$ denote the set of indices of bad pools in $\mathcal{D}$ when the objects indexed by A are bad and no failures occur, that is

$$\phi(A) = \cup_{i \in A} B_i. \tag{2}$$

We say that $\mathcal{D}$ is a p-bad, 0-failure solution, or $(p, 0)$-solution, if from $\phi(A)$ we can infer either A or that $|A| > p$, assuming that no failures occur. This occurs if and only if

$$\phi(A) \neq \phi(A') \quad \text{for all } A, A' \in \mathcal{X}_n \text{ such that } A \neq A' \text{ and } |A| \leq p. \tag{3}$$

Note that in the hypergeometric case (Hwang & Sós, [4]), $\phi(A) \neq \phi(A')$ is required only when both $|A| \leq p$ and $|A'| \leq p$.

We define $\mathcal{D}$ to be a p-bad, q-failure solution, or (p, q)-solution, if from $\phi(A)$ we can infer either A or that $|A| > p$, even in the presence of up to q failures. This occurs if and only if each $(v-q)$-subset of $\mathcal{D}$ is a $(p, 0)$-solution. We write $\sigma^v_{p,q}$ for the set of duals of (p, q)-solutions and say that $\mathcal{D}$ is optimal if $\hat{\mathcal{D}}$ has maximum cardinality in $\sigma^v_{p,q}$. From (3) it follows that $\mathcal{D}$ is a (p, q)-solution if and only if

$$|\phi(A) \Delta \phi(A')| > q \quad \text{for all } A, A' \in \mathcal{X}_n \text{ such that } A \neq A' \text{ and } |A| \leq p, \tag{4}$$

in which $B \Delta C \equiv (B \backslash C) \cup (C \backslash B)$. Note that (4) can be regarded as the definition of a solution in the case that test failures do not occur but up to q wrong outcomes may be recorded and the detection of any such error is required. Hence optimal p-bad, q-failure solutions are also optimal p-bad, q-error-detecting solutions.

Proposition 1 *A design $\mathcal{D}$ is a (p, q)-solution, that is $\hat{\mathcal{D}} \in \sigma^v_{p,q}$, if and only if*

$$|B_i \setminus \phi(A)| > q \quad \textit{for every } A \in \mathcal{X}_n \textit{ with } |A| \leq p \textit{ and all } i \in \{1, 2, \ldots, n\} \backslash A. \tag{5}$$

Corollary 1 *A design D satisfies $\hat{\mathcal{D}} \in \sigma^v_{1,q}$ if and only if $|B \backslash B'| > q$ for all distinct $B, B' \in \hat{\mathcal{D}}$.*

Proof By considering the case that $A \in \mathcal{X}^p_n$ and $A' = A \cup \{i\}$ for some $i \notin A$, we see that (5) is necessary for (4). Suppose that $A, A' \in \mathcal{X}_n$ with $A \neq A'$ and $|A| \leq p$. If $A' \backslash A \neq \emptyset$ then it follows from (5) that $|\phi(A') \backslash \phi(A)| > q$. Alternatively, if $A' \backslash A = \emptyset$ then both $|A'| \leq p$ and $A \backslash A' \neq \emptyset$ and hence (5) implies that $|\phi(A) \backslash \phi(A')| > q$. In either case we have $|\phi(A) \Delta \phi(A')| > q$ and hence (5) is sufficient for (4).

Let $0 \leq t < k \leq v$. A (t,k,v)-packing is a set $\mathcal{P} \subseteq \mathcal{X}_v^k$ such that the intersection of any two elements of $\mathcal{P}$ has cardinality at most t. A direct corollary of Proposition 1 is that if $\hat{\mathcal{D}}$ is a $(t, pt{+}q{+}1, v)$-packing then $\hat{\mathcal{D}} \in \sigma_{p,q}^v$. If $|\mathcal{P}| = \binom{v}{t+1}\binom{k}{t+1}^{-1}$ then each element of $\mathcal{X}_v^{t+1}$ is contained in precisely one element of $\mathcal{P}$ and $\mathcal{P}$ is also called a Steiner system, denoted $S(t{+}1, k, v)$. For further details including a list of small Steiner systems known to exist, we refer to Beth *et al.* [1].

Theorem 1 *If a design $\mathcal{D}$ satisfies $\hat{\mathcal{D}} \in \sigma_{1,q}^v$ then $n \equiv |\hat{\mathcal{D}}|$ satisfies*

$$n \leq \frac{1}{K_q}\binom{v}{\lfloor v/2 \rfloor}, \tag{6}$$

in which $K_0 = 1$ and, for q even,

$$K_q = \sum_{s=0}^{q/2} \binom{\lfloor v/2 \rfloor}{s}\binom{\lceil v/2 \rceil}{s}, \tag{7}$$

while for q odd,

$$K_q = K_{q-1} + \frac{1}{T}\binom{\lfloor v/2 \rfloor}{(q{+}1)/2}\binom{\lceil v/2 \rceil}{(q{+}1)/2} \tag{8}$$

where $T \equiv \lfloor 2\lfloor v/2 \rfloor/(q{+}1) \rfloor$.

Corollary 2 (Sperner, 1928) *The set $\mathcal{X}_v^{\lfloor v/2 \rfloor}$ is optimal in $\sigma_{1,0}^v$.*

Corollary 3 *If $S(\lfloor v/2 \rfloor{-}1, \lfloor v/2 \rfloor, v)$ exists then it is optimal in $\sigma_{1,1}^v$.*

Theorem 2 *If a design $\mathcal{D}$ satisfies $\hat{\mathcal{D}} \in \sigma_{2,q}^v$ then*

$$n \leq \binom{v}{t^*}\binom{2t^* + q - 1}{t^*}^{-1}, \tag{9}$$

in which t^* is the least integer value of t such that

$$v \leq 5t + 2 + \frac{q(q{-}1)}{t{+}q}, \tag{10}$$

so that $t^* = \lceil (v{-}2)/5 \rceil$ if $q = 0$ or 1.

Corollary 4 *If $S(t^*, 2t^*+q-1, v)$ exists then it is optimal in $\sigma^v_{2,q}$.*

In the $p = 1$ case, Stirling's formula and (6) give

$$n < \lfloor q/2 \rfloor ! \lceil q/2 \rceil ! \frac{2^{v+q+1}}{v^q \sqrt{2\pi v}}, \tag{11}$$

and hence asymptotically

$$v > \frac{\log(n)}{\log(2)} \left(1+o(1)\right). \tag{12}$$

For $p = 2$, Stirling's formula and (9) give

$$n < \frac{5^{v+1/2}}{2^{2v+q+1/2}}, \tag{13}$$

and hence

$$v > \frac{\log(n)}{\log(5/4)} \left(1+o(1)\right). \tag{14}$$

The asymptotic bound (14) was obtained by Erdös *et al.* [3].

3 Proof of Theorem 1

By Corollary 1, if $q = 0$ then each $\hat{\mathcal{D}} \in \sigma^v_{1,0}$ satisfies the requirement of no pairwise containment and the theorem was proved in this case by Sperner [6]. A simple proof of Sperner's result is given by Lubell [5]. Here, and in the sequel, 'chain' will always mean a maximal chain of $\mathcal{X}_v$, ordered by inclusion. Each chain contains at most one element of $\hat{\mathcal{D}}$ and hence we can associate with each $B \in \hat{\mathcal{D}}$ a 'cost' which is the proportion of all chains which contain B and hence no further element of $\hat{\mathcal{D}}$. For any k, the set of chains can be partitioned by the k-sets into $\binom{v}{k}$ equal parts. Therefore the cost of B is $1/\binom{v}{k}$, where $k \equiv |B|$, which is minimized at $k = \lfloor v/2 \rfloor$. Since $\mathcal{X}_v^{\lfloor v/2 \rfloor}$ consists only of minimal cost elements but achieves the maximum total cost of one, it is optimal in $\sigma^v_{1,0}$.

This argument can be extended to the $p = 1$, $q > 0$ case. Suppose first that q is even and consider $B \in \hat{\mathcal{D}}$ with $|B| = k$. Define the s-neighbours of B to be the sets $C \in \mathcal{X}_v^k$ such that $|B \backslash C| = |C \backslash B| = s$. We will say that B 'blocks' the chains which contain one of its s-neighbours for some $s \leq q/2$. Let B' denote an element of $\hat{\mathcal{D}}$ distinct from B. If a chain contains both an s-neighbour of B and an s'-neighbour of B' then either $|B \backslash B'| \leq s+s'$ or $|B' \backslash B| \leq s+s'$. It follows from Corollary 1 that we cannot have both $s \leq q/2$ and $s' \leq q/2$. Therefore a chain cannot be blocked by more than one element of $\hat{\mathcal{D}}$. Each chain blocked by

B can contain no element of $\hat{\mathcal{D}}$ other than B and hence we can associate with B the cost $h(B)$ which is the proportion of chains blocked by B. The value of $h(B)$ is

$$h(B) = K_{q,k}\binom{v}{k}^{-1}, \tag{15}$$

where $K_{q,k}$ denotes the total number of s-neighbours of B with $s \leq q/2$, which is given by

$$K_{q,k} = \sum_{s=0}^{q/2} \binom{k}{s}\binom{v-k}{s}. \tag{16}$$

It is readily verified that $h(B)$ is minimized when $k = \lfloor v/2 \rfloor$ or $\lceil v/2 \rceil$ and, in either case, the value of $K_{q,k}$ is K_q, defined at (7). Since chains cannot be multiply blocked, the sum of the costs of the elements of $\hat{\mathcal{D}}$ cannot exceed one. Therefore n is bounded above by the inverse of the minimal cost which establishes (6) in the case q even.

Now suppose that q is odd. Each chain which contains an s-neighbour of B, for some $s \leq (q+1)/2$, can contain no element of $\hat{\mathcal{D}}$ other than B and we say that these chains are blocked by B. If B' is an element of $\hat{\mathcal{D}}$ distinct from B then no chain can be both an s-neighbour of B and an s'-neighbour of B' when $s+s' \leq q$. However, if either $|B \backslash B'| = q+1$ or $|B' \backslash B| = q+1$ then there exist chains which contain both a $((q+1)/2)$-neighbour of B and a $((q+1)/2)$-neighbour of B' and hence such chains are multiply blocked.

Let $\mathcal{C}$ denote the set of elements of $\hat{\mathcal{D}}$ which have a $((q+1)/2)$-neighbour in the chain $\{\emptyset, \{1\}, \{1,2\}, \ldots, \{1,2,\ldots,v\}\}$. If $B \in \mathcal{C}$ with $|B| = k$ then

$$|B \cap \{k+1, k+2, \ldots, v\}| = \frac{q+1}{2} = |\{1,2,\ldots,k\} \setminus B|. \tag{17}$$

Further, if $B' \in \mathcal{C}$ and $B' \neq B$ then, since both $|B \backslash B'| > q$ and $|B' \backslash B| > q$ must be satisfied, B' contains $\{1,2,\ldots,k\} \backslash B$ and is disjoint from $\{k+1, k+2, \ldots, v\} \cap B$. Hence $|\mathcal{C}|$ cannot exceed $\min\{\lfloor 2k/(q+1) \rfloor, \lfloor 2(v-k)/(q+1) \rfloor\}$, which takes maximum value T when $k = \lfloor v/2 \rfloor$. Therefore the number of $((q+1)/2)$-neighbours in $\hat{\mathcal{D}}$ of a given chain is at most T.

Associate with B a cost

$$h'(B) = K'_{q,k}\binom{v}{k}^{-1}, \tag{18}$$

where $K'_{q,k}$ is $K_{q-1,k}$ plus $1/T$ times the number of $((q+1)/2)$-neighbours of B. This cost is minimized at $k = \lfloor v/2 \rfloor$ or $\lceil v/2 \rceil$ and in either case $K'_{q,k}$ is K_q defined at (8). The sum of the costs of the elements of $\hat{\mathcal{D}}$ cannot exceed one and hence the theorem.

4 Proof of Theorem 2

Definition 1 *For any $\hat{\mathcal{D}} \in \sigma^v_{2,q}$, we follow Erdös* et al. [3] *and say that $b \in \mathcal{X}_v$ is* private in $\hat{\mathcal{D}}$ *if there exists a unique $B \in \hat{\mathcal{D}}$ such that $b \subseteq B$.*

Definition 2 *If $B \in \mathcal{X}_v$ with $|B| > q$ then $\mathcal{F} \subset \mathcal{X}_v$ is a* (2,q)-cover *of B precisely if both*

1. *if $b \in \mathcal{F}$ and $b \subset b' \subseteq B$ then $b' \in \mathcal{F}$; and*
2. *for every $b \subseteq B$ with $|B \backslash b| \leq q$ at least one part of every two-partition of b is in $\mathcal{F}$.*

Lemma 1 *A design $\mathcal{D}$ satisfies $\hat{\mathcal{D}} \in \sigma^v_{2,q}$ if and only if for each $B \in \hat{\mathcal{D}}$, the sets which are private in $\hat{\mathcal{D}}$ form a $(2,q)$-cover of B.*

Proof Let $b \subseteq B$ with $|B \backslash b| \leq q$. It follows from Proposition 1 that $|B| > q$ and b is private in $\hat{\mathcal{D}}$. If there exists a partition of b into two non-private parts then there must be C and C' in $\hat{\mathcal{D}} \backslash \{B\}$ such that $b \subseteq (C \cup C')$ and hence $|B \setminus (C \cup C')| \leq q$, which contradicts Proposition 1. Therefore a necessary condition for $\mathcal{D} \in \sigma^v_{2,q}$ is that for each $B \in \hat{\mathcal{D}}$ the private subsets of B in $\hat{\mathcal{D}}$ form a $(2,q)$-cover. Sufficiency is immediate from Proposition 1.

Definition 3 *For $\mathcal{F}$ a $(2,q)$-cover of $B \in \mathcal{X}_v$, define $h(B, \mathcal{F})$ to be the proportion of all chains which intersect $\mathcal{F}$.*

Proposition 2 *For any $B \in \mathcal{X}_v$ with $|B| = k$ such that $q < k < v-1$ and $\mathcal{F}$ a $(2,q)$-cover of B,*

$$h(B, \mathcal{F}) \geq \binom{2t+q-1}{t} \binom{v}{t}^{-1}, \tag{19}$$

in which $t = \lfloor (k-q+1)/2 \rfloor$. Equality is achieved in (19) if and only if $k-q$ is odd and $\mathcal{F} = \mathcal{F}^*$ where $\mathcal{F}^* \equiv \{b \subset B : |b| \geq t\}$.

Proof If $k = 2t+q-1$ then every two-partition of any $(k-q)$-subset of B contains a part b such that $|b| \geq t$ and hence $\mathcal{F}^*$ is a $(2,q)$-cover of B. Further, $h(B, \mathcal{F}^*)$ is precisely the proportion of chains which contain a t-subset of B and hence $h(B, \mathcal{F}^*)$ achieves equality in (19).

Suppose that $k = 2t+q-1$ and let s denote the largest integer such there exists some $b \subset B$ with $|b| = t+s$ and $b \notin \mathcal{F}$. It follows from Definition 2 that $s < t-1$. If $s < 0$ then either $\mathcal{F} = \mathcal{F}^*$ or $\mathcal{F}^* \subset \mathcal{F}$ and $h(B, \mathcal{F}^*) < h(B, \mathcal{F})$ and hence we may assume that $s \geq 0$. If there exists some $b \in \mathcal{F}$ such that $|b| < t-s-1$ then $\mathcal{F} \backslash \{b\}$ is also a $(2, q)$-cover of B and $h(B, \mathcal{F} \backslash \{b\}) < h(B, \mathcal{F})$. Hence we may assume that there is no such b. From Definition 2, if b is a $(k-q)$-subset of B then the number of $(t+s)$-subsets of b not in $\mathcal{F}$ is not greater than the number of $(t-s-1)$-subsets of b in $\mathcal{F}$. Summing over all such b, each r-set occurs in $\binom{k-r}{q}$ terms of the sum and hence

$$\binom{k-t-s}{q} \bar{f}_{t+s} \leq \binom{k-t+s+1}{q} f_{t-s-1}, \tag{20}$$

in which $\bar{f}_r$ denotes the number of r-subsets of B not in $\mathcal{F}$ while f_r denotes the number of such subsets in $\mathcal{F}$. Inequality (20) is equivalent to

$$\bar{f}_{t+s} \leq f_{t-s-1} \frac{(k-t+s+1)!(t-s-1)!}{(k-t-s)!(t+s)!}. \tag{21}$$

Construct $\mathcal{F}'$ from $\mathcal{F}$ by removing all $(t-s-1)$-sets and adding any missing $(t+s)$-subsets of B, so that

$$\mathcal{F}' \equiv \{b \subset B : |b| = t+s\} \cup \mathcal{F} \setminus \{b \subset B : |b| = t-s-1\}. \tag{22}$$

Now, $\mathcal{F}'$ is also a $(2, q)$-cover of B and $h(B, \mathcal{F}') - h(B, \mathcal{F})$ is precisely the proportion of chains which contain a $(t+s)$-subset of B but no element of $\mathcal{F}$ minus the proportion of chains which contain a $(t-s-1)$-set in $\mathcal{F}$ but no other element of $\mathcal{F}$. Therefore

$$h(B, \mathcal{F}') - h(B, \mathcal{F}) = \frac{v-k}{v-t-s} \bar{f}_{t+s} \binom{v}{t+s}^{-1} - \frac{v-k}{v-t+s+1} f_{t-s-1} \binom{v}{t-s-1}^{-1}, \tag{23}$$

and hence $h(B, \mathcal{F}') \geq h(B, \mathcal{F})$ if and only if

$$\bar{f}_{t+s} \geq f_{t-s-1} \frac{(v-t+s)!(t-s-1)!}{(v-t-s-1)!(t+s)!}. \tag{24}$$

Since $k+1 < v$, inequality (24) contradicts (21) and hence $h(B, \mathcal{F}') < h(B, \mathcal{F})$. Therefore $h(B, \mathcal{F})$ is not minimal and the proposition is established in this case.

When $k = 2t$ and $q = 0$, for every partition of B into two t-sets, one of the parts is in $\mathcal{F}$. Hence $\mathcal{F}$ contains at least $\binom{2t-1}{t}$ sets of size t. From an argument similar to that above, it is readily shown that $h(B, \mathcal{F})$ is minimized when $\mathcal{F}$ contains every $(t+1)$-subset of B but no $(t-1)$-subset. Hence the bound (19)

follows with strict inequality. If $k = 2t+q$, $q > 0$, then for any $x \in B$ the set $\{b \in \mathcal{F} : x \notin b\}$ is a $(2, q-1)$-cover of $B\backslash\{x\}$ and the proposition follows from the case $k = 2t+q-1$.

Proof of Theorem 2 Let x_t denote the RHS of (19). Then

$$\frac{x_{t+1}}{x_t} = \frac{(2t+q+1)(2t+q)}{(v-t)(t+q)} \tag{25}$$

which exceeds one if and only if inequality (10) is not satisfied. Hence x_t is minimized at $t \equiv t^*$. From Definition 1, if $B \in \hat{\mathcal{D}}$ and $\hat{\mathcal{D}} \in \sigma^v_{2,q}$ then any chain which intersects $\mathcal{F}$ cannot intersect any other set which is private in $\hat{\mathcal{D}}$. Therefore, invoking Lemma 1, $n \equiv |\hat{\mathcal{D}}|$ is bounded above by the inverse of the minimum value of $h(B, \mathcal{F})$ over all $\mathcal{F}$ a $(2, q)$-cover of B, which in turn is bounded above by $1/x_{t^*}$.

Acknowledgements We thank Dr Charles Goldie of QMW for helpful comments on an early draft of the manuscript. Work supported in part by the UK Science and Engineering Research Council under grants GR/F 98727 (DJB) and GR/J 05880 (DCT) and in part through the Center for Human Genome Studies at Los Alamos National Laboratory under grant US DOE/OHER ERWF118.

References

[1] Beth T., Jungnickel D. and Lenz H. (1986) *Design Theory*. Cambridge U.P.

[2] Bush K.A., Federer W.T., Pesotan H. and Raghavarao D. (1984) New combinatorial designs and their applications to group testing. *J. Stat. Plann. Inf.*, **10**, pp 335-343.

[3] Erdös P., Frankl P. and Füredi Z. (1982) Families of finite sets in which no set is covered by the union of two others. *J. Comb. Th.*, **A 33**, pp 158-166.

[4] Hwang F.K. and Sós V.T. (1987) Non-adaptive hypergeometric group testing. *Stud. Sci. Math. Hung.*, **22**, pp 257-263.

[5] Lubell D. (1966) A short proof of Sperner's lemmas. *J. Comb. Th.*, **1**, p 299.

[6] Sperner E. (1928) Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, **27**, pp 544-548.

[7] Ruszinkó M. (1994) On the upper bound of the size of the r-cover-free families. In press *J. Comb. Th.* **A**.

[8] Wolf J.K. (1985) Born again group testing : multiaccess communications. *IEEE Trans. Inf. Th.*, **IT-31**, pp 185-191.