A Maximum Likelihood Estimator for the Prevalence Rate Using Pooled Sample Tests

João Paulo Martins, Rui Santos and Miguel Felgueiras

Abstract Since Dorfman's seminal work, the research on methodologies involving pooled sample tests has been quite active (cf. [8]). Moreover, the use of pooled samples does not refer only to the classification problem (identifying all the infected individuals in a population), since it may also be useful in estimating the prevalence rate p, as [18] stated. The use of compound tests is not restricted to hierarchical algorithms whose most common example is Dorfman's two-stage procedure (cf. [4]). Matrix schemes as the square array algorithm (cf. [16]) or multidimensional matrices schemes (cf. [1]) in certain cases outperform Dorfman's procedure (cf. [10]). Maximum likelihood estimates are quite difficult to compute when a procedure does not classify all individuals. This paper presents two innovator methods to compute maximum likelihood estimates in both type of procedures.

Key words: maximum likelihood estimator, prevalence rate, compound tests.

João Paulo Martins
School of Technology and Management, Polytechnic Institute of Leiria, Apartado 4163, 2411-901 Leiria, Portugal, CEAUL – Center of Statistics and Applications of the University of Lisbon
e-mail: jpmartins@ipleiria.pt

Rui Santos
School of Technology and Management, Polytechnic Institute of Leiria, Apartado 4163, 2411-901 Leiria, Portugal, CEAUL – Center of Statistics and Applications of the University of Lisbon
e-mail: rui.santos@ipleiria.pt

Miguel Felgueiras
School of Technology and Management, Polytechnic Institute of Leiria, Apartado 4163, 2411-901 Leiria, Portugal, CEAUL – Center of Statistics and Applications of the University of Lisbon
e-mail: mfelg@ipleiria.pt

1 Introduction

The use of group testing procedures to screen for a binary characteristic is usually set to have started from Dorfman's (cf. [4]) seminal work. His purposed procedure proved to be less expensive than applying only individuals tests in the detection of the syphilis infected soldiers involved in the World War II. The new strategy was to gather groups of n individuals into pools and then performing a pooled sample test. A negative result of the pooled mixture indicates that all of them are free of the disease. A positive result indicates that at least one of the n individuals has the disease, but we do not know who or how many. In this case, performing individual tests is advised to identify the infected individuals in the sample from the non-infected ones. The main issue is to determine the optimal batch size which minimizes the expected number of tests as it is a good measure of the monetary cost, since the cost of mixing samples is usually negligible (cf. [13]).

Pooled samples may be used in two types of problems: a classification problem or an estimation problem. Identifying all the subjects that are infected or have a high level of sugar in blood are examples of classification problems. In both examples it is required to ascertain for each individual if it verifies the condition of interest. Estimating the prevalence rate of a disease or of a gene in some population are examples of estimation problems. In this case, the performance of individual tests is only optional, since the goal is no longer to identify the infected individuals (cf. [3]). The use of only pooled samples has also the advantage of anonymity of the infected members, given that they are not identified. Furthermore, the estimators obtained by applying compound tests have, under certain conditions, better performance than the traditional estimators based on individual tests, cf. [6, 14, 18]. The bias, the efficiency and the robustness of these estimators have been reviewed in several works, such as those from [3, 7, 12]. [2] purposes the use of the package *binGroup* for the *R* software, which includes applications of several compound testing estimators. Thus, the estimators based on group testing not only allow to obtain monetary gains (by decreasing the number of performed tests), but also allow to achieve more accurate estimates, compared to those obtained on the basis of individual tests.

Group testing application can be done in several ways (cf. [10]). The main reason for having different procedures is related to the misclassification problem, as an individual can be wrongly classified. The sensitivity and the specificity of the test (see Definition 1) may be used for measuring the accuracy of the test results. In particular, the sensitivity of a test generally decreases as the pooled sample size increases. The choice for a particular group testing procedure depends on the amount of samples available and the sensitivity, the specificity and the monetary costs of the process (cf. [13]). For an overview about this problem, known as the dilution problem, see [9], [17], [20] and [21].

The outline of this work is as follows. Section 2 introduces the binomial model assumption and discusses some considerations about the prevalence rate maximum likelihood (ML) estimator when pooled samples are used. Section 3 describes the two main types of group testing procedures and is the core of this work as new ways of computing ML estimates are provided. For the hierarchical algorithms dealt in

subsection 3.1 it is purposed a classification of an individual for estimation purposes that does not need any individual tests. This allows the application of the traditional ML estimators even when in the algorithm last stage no individual tests are performed. The array-based group testing procedures are presented in subsection 3.2. In these kind of procedures, the ML estimates computation is very difficult to perform. Hence, an iterative method is purposed in order to obtain reasonable estimates that can be easily obtained. In section 4 some final remarks are discussed.

2 The Binomial Model

Let p denote the probability that an individual is infected, n be the pool sample size and t the number of performed tests. The total number of individuals is $N = n \times t$. Let us also assume that the individuals status (infected/not infected) within a pool sample are independent. The probability of having an infected pooled sample is $\pi_n = 1-(1-p)^n$. Hence the total number of infected samples is described by a binomial random variable $T \frown \text{Bin}(t, \pi_n)$. The ML estimator of π_n is

$$\widehat{\pi_n} = \frac{T}{t}. \tag{1}$$

As p is given by a simple transformation of π_n, it is straightforward to prove, applying the proprieties of the ML estimators, that the ML estimator of p is

$$\widehat{p} = 1 - \left(1 - \frac{T}{t}\right)^{1/n}. \tag{2}$$

For $n = 1$, $\hat{p} = 1 - \left(1 - \frac{T}{t}\right) = \frac{T}{t}$ is an unbiased estimator of p. For $n > 1$, the estimator is positively biased . Expressions for the estimator expected value and variance may be found in [7].

As screening errors may occur, that model is, in practice, unrealistic. Thus, consider the problem of estimating the prevalence rate of some disease. Let $X_i = 1$ denote an infected individual and $X_i = 0$ denote a non-infected individual. In addition, let X_i^+ stand for a positive test result and X_i^- stand for a negative test result. In order to assess the sources of error two measures will be considered.

Definition 1. Consider an individual X_i that is tested individually. The probability $\varphi_s = \text{P}\left(X_i^+ | X_i = 1\right)$ is called the **test sensitivity** and $\varphi_e = P\left(X_i^- | X_i = 0\right)$ is called the **test specificity**.

When a pooled sample test is performed the probability of having a positive result from an infected sample may decrease. As the amount of substance per unit of volume is less or equal to the amount of substance found in a unit of volume collected from an infected individual it may be difficult to screen the infected pool sample as positive. However, the probability of getting a negative outcome on a non-infected sample is equal to φ_e as there is no dilution problems. Thus, [17] define the

concepts of specificity and sensitivity of some specific methodology of classification or estimation $\mathscr{M}$ (these concepts are closely related to the pooling sensitivity and pooling specificity concepts defined in [10]). These measures assess the quality of an outcome provided by some methodology $\mathscr{M}$.

Definition 2. The **methodology sensitivity** or the procedure sensitivity is the probability of an infected individual being correctly identified by the methodology $\mathscr{M}$, that is, $\varphi_s^{\mathscr{M}} = P_{\mathscr{M}}\left(X_i^+ | X_i = 1\right)$. The **methodology specificity** or the procedure specificity stands for the probability of a non-infected individual being correctly classified by the methodology $\mathscr{M}$, that is, $\varphi_e^{\mathscr{M}} = P_{\mathscr{M}}\left(X_i^- | X_i = 0\right)$.

For an individual testing procedure the sensitivity (specificity) methodology is equal to the test sensitivity (specificity). For instance, in the Dorfman's procedure and, admitting no dilution effect, the probability of an infected individual being screened as positive is

$$\varphi_s^{\mathscr{M}} = \varphi_s^2, \tag{3}$$

as it is required that both pooled and subsequent individual test outcomes to be positive. Note that the methodology sensitivity is less than the test sensitivity, ie.,

$$\varphi_s^{\mathscr{M}} \leq \varphi_s, \tag{4}$$

for $n > 1$.

For computing the probability of a non-infected individual being correctly classified it is necessary to account for three possible situations:

- the pooled sample is not infected and the pooled test outcome is negative;
- the pooled sample is not infected but the pooled test outcome is positive and the subsequent individual test outcome is negative;
- the pooled sample is infected and the pooled test outcome is positive but in the subsequent individual test the subject is correctly classified as non-infected.

Hence,

$$\varphi_e^{\mathscr{M}} = \varphi_e (1-p)^{(n-1)} + (1-\varphi_e)\varphi_e (1-p)^{(n-1)} + \varphi_s \varphi_e \left(1 - (1-p)^{(n-1)}\right). \tag{5}$$

The exponent $n-1$ in equation (5) is due to the fact that we are computing a conditional probability. These probabilities allows us to compute the real bias of

$$\widehat{p} = \frac{T}{N} \tag{6}$$

when the Dorfman's procedure is applied. Now, T stands for the number of specimens classified as positive. T is a binomial random variable described by $T \frown \text{Bin}(N, p^*)$. It depends on the methodology specificity $\varphi_e^{\mathscr{M}}$ and on the methodology sensitivity $\varphi_s^{\mathscr{M}}$, therefore, $p^* = \psi(\varphi_s, \varphi_e, p)$. [17] computed the value of p^* by

$$\begin{aligned} p^* &= P\left(X_i^+|D\right)P(D) + P\left(X_i^+|\overline{D}\right)P\left(\overline{D}\right) \\ &= \varphi_s^{\mathscr{M}}\, p + \left(1-\varphi_e^{\mathscr{M}}\right)(1-p) \\ &= 1-\varphi_e^{\mathscr{M}} + \left(\varphi_s^{\mathscr{M}} + \varphi_e^{\mathscr{M}} - 1\right) p. \end{aligned} \tag{7}$$

Hence, the estimator is, in general, biased. The bias is equal to

$$\text{Bias}\left(\widehat{p}\right) = p^* - p \tag{8}$$

and the estimator variance is

$$\text{Var}\left(\widehat{p}\right) = \frac{p^*\left(1-p^*\right)}{N}. \tag{9}$$

The mean square error (MSE) of the estimator is, by definition,

$$\text{MSE}\left(\widehat{p}\right) = \left[\text{Bias}\left(\widehat{p}\right)\right]^2 + \text{Var}\left(\widehat{p}\right). \tag{10}$$

Note that, for instance, if $\varphi_e^{\mathscr{M}} = \varphi_s^{\mathscr{M}} = p = 0.5$ the estimator is actually unbiased. The mean square error is a possible measure for assessing the quality of the estimates in each procedure. This measure may be used to combine different prevalence rate estimates. [3] uses a logistic regression whose parameters are computed iteratively but measures the quality of each estimate just using the pooled sample size. [15] provides an iterative meta-analysis-based procedure that uses the mean square error as weights for achieving a single estimate. The content in subsection 3.2 enhances [15] work as it provides a computational method for estimating the prevalence rate from an array-based group testing algorithm and, even more important to that meta-analysis technique, it provides a way to estimate the MSE of the estimator.

3 ML Estimators in Several Group Testing Procedures

On a pooled sample-based procedure there are two goals: minimizing the sources of error and providing a less expensive method than individual testing for achieving the investigation goal. To assess the savings of some procedure $\mathscr{M}$, the **relative cost** will be used as a measure of the **methodology efficiency**, $\text{RC}(\mathscr{M})$, that is, the expected number of tests per specimen since the cost of mixing samples is usually negligible. When only individual tests are performed the methodology efficiency is equal to one. In general, the methodology efficiency is high for low prevalence rates as the pooled samples sizes tend to decrease with p. For instance, in the traditional Dorfman's procedure the maximum efficiency for a prevalence rate equal to 0.1, 0.01 and to 0.001 is obtained by using a pooled sample size equal to 4, 11 and 32, respectively (cf. [4]).

The most used pooled sample procedures can be binned in the following two groups:

- **Hierarchical algorithms** – a pooled sample is tested and if the test outcome is positive it is divided into smaller nonoverlapping groups until eventually all individuals have been tested;
- **Array-based group testing algorithms** – in its simplest two-stage version (square array), a sample of size n^2 is placed in a $n \times n$ matrix and then all the individuals within the same row and the same column are gathered for batched testing.

3.1 Hierarchical Algorithms

Dorfman's procedure is just one example of a wider family called hierarchical algorithms. Some improvements of his work have been made (cf. [5, 19, 20]) by dividing positive pools into smaller subpools until eventually all positive specimens are individually tested.

A multistage hierarchical algorithm is an algorithm that generalizes Dorfman's procedure to more than two stages, that is, a sample is divided at each stage into smaller nonoverlapping groups until eventually all positive specimens are individually tested. At each stage, subsamples from the samples that tested positively are retested. For practical reasons, only two or three stages are usually performed. Let us consider an hierarchical algorithm with s stages and let n_i denote the number of individuals at the i-th stage. At the last stage, when the classification problem is considered, we have $n_s = 1$. However, this might not be fulfilled, when we just want to estimate the prevalence rate, and the condition verified is just $n_1 > \cdots > n_s \geq 1$ (cf. [3, 7, 12]). For low prevalence rates, the use of $n_s > 1$ for achieving a greater efficiency may be justified if a positive outcome when testing a pooled sample of size n_s at the last stage means (almost surely) that only one of the individuals is infected (cf. [17]). Hence, when $n_s > 1$ we will consider that, for estimation purposes, an individual X_i is correctly/wrongly classified ($\checkmark/\times$) according to the next table

Table 1 Correct and wrong decisions at the s-th stage

		Pooled sample at the s-th stage	
		Infected	Not infected
$X_i = 0$	Test result +	✓	×
	Test result −	✓	✓
$X_i = 1$	Test result +	✓	Not possible
	Test result −	×	Not possible

One of the less intuitive shown classifications in Table 1 is, for an infected sample at the s-th stage, if the test outcome is positive the decision is correct. This is almost 100% true as it means (almost surely) that only one is infected and that the others individuals are not. Therefore, concerning the estimation problem, all the individuals are (almost surely) well classified as one infected and $n_s - 1$ non-infected. Hence, although we may not be able to set who is infected it is now straightforward to compute the ML estimator (6). However, the given estimate may not be a ML estimate since, although unlikely, it is possible to have two infected individuals at the s stage.

3.2 Array-based Group Testing Algorithms

Array-based group testing is an alternative to hierarchical group testing that uses overlapping pools. In its simplest two-stage version (square array), denoted by $A2\,(n:1)$, a sample of size n^2 is placed in a $n \times n$ matrix in the following way. Each individual is allocated at one and only one matrix position. Then, all the individuals within the same row and the same column are gathered for batched testing. This process involves at least $2n$ tests as subsequent individual tests are performed to the samples lying on a row and/or column that tested positively. A variant of this methodology consists in performing a priori a pooled sample test to all the n^2 individuals (masterpool). If the masterpool test result is negative no further testing is needed as the individuals are all negative classified. This methodology with a master pool will be represented by $MA2\left(n^2:n:1\right)$. The performance of subsequent individual tests is required to avoid ambiguities. For instance, it is possible to have a row that tested positive but all columns tested negative. To obtain a greater efficiency we suggest the dropping of the subsequent individual testing because it is not necessary to determine who are exactly the infected individuals, when dealing with an estimation problem . Let us look for a simple example of a square array procedure with two lines (with or without) a master pool.

Example 1. [11] compares the operating characteristics of two square array procedures with a master pool: $MA2\,(49:7:1)$ and $MA2(100;10;1)$ when screening for a disease in Malawi with prevalence rate 0.045. We computed the operating characteristics of the last procedure without any individual tests: $MA2(100;10)$. In this case, an individual is classified as positive if and only if both "row" and "column" tested positive.

Hence, if all the columns (rows) tested negative and a row (column) tested positive, all the individuals are classified as negative. This approach although is much more efficient than the others has a great drawback. It almost surely underestimate the prevalence rate!

As it is not possible to use the proportion of defective individuals without avoiding an underestimation of the prevalence rate, we propose the computation of a ML

Table 2 Comparing the efficiency of the different methodologies

Methodology	RC($\mathscr{M}$)	$\varphi_e^{\mathscr{M}}$	$\varphi_s^{\mathscr{M}}$
Indiviual test	1	0.9900	0.9000
$MAS(49;7;1)$	0.34	0.9995	0.6810
$MAS(100;10;1)$	0.31	0.9991	0.6596
$MAS(100;10)$	0.22	0.9990	0.7290

estimate, using a proper script. This will combine a greater efficiency with the computation of an accurate estimate.

When the number of rows and columns of the two-dimensional array is low it is possible to compute the exact value of the likelihood function for a given prevalence rate p_0. For an array with two rows and two columns it is easy although hardworking to write a script to compute the ML function for any value. Hence, a proper iterative process gives the ML estimate.

The inputs of the script must be the test sensitivity φ_s, the test specificity φ_e and the number of arrays that have $i-1$ positive rows and $j-1$ positive columns for $i = 1,2,3$ and $j = 1,2,3$. These values may be inserted on a 3×3 matrix O.

For computing the ML function at p_0 it is also required to compute the probability of observing $i-1$ positive rows and $j-1$ positive columns, where $i = 1,2,3$ and $j = 1,2,3$, given p_0 and taking into account the test sensitivity φ_s and the test specificity φ_e. Suppose these values are recorded in a matrix P. For instance, if $\varphi_s = \varphi_e = 0.95$ (consider that the individual test sensitivity is equal to the pooled sample sensitivity) and $p_k = 0.1$. The matrix P_0 is

$$P_0 = \begin{pmatrix} 0.5351 & 0.0689 & 0.0029 \\ 0.0689 & 0.2477 & 0.0277 \\ 0.0029 & 0.0277 & 0.0183 \end{pmatrix}$$

As the matrix of a square array is always symmetric it can be written as an upper triangular matrix

$$P = \begin{pmatrix} 0.5351 & 0.1378 & 0.0058 \\ 0 & 0.2477 & 0.0554 \\ 0 & 0 & 0.0183 \end{pmatrix}$$

In this case, the matrix O must be also rewritten as $O(i,j) = O(i,j) + O(j,i)$ for $j > i$. Note that it is expected to have near of 14% of the arrays with only one positive row (column) and no positive columns (rows). In the traditional application of a square array methodology, this would require the performance of individual tests. The ML function for p_0 is given by

$$ML(p_0) = \prod_{i=1}^{3} \prod_{j=i}^{3} P(i,j)^{O(i,j)}. \tag{11}$$

Example 2. To assess the MSE of the estimator, for a prevalence rate $p = 0.1$, 100 replicas of a 2×2 array ($A2(2:1)$) were simulated in software MatLab R2011 and the ML estimate was computed. This procedure was repeated 1000 times to produce 1000 prevalence rate estimates. The matrix O was set to be equal to the matrix P. Although, in practice, the matrix O only admits integers values, that is not important for our intents.

The estimates mean value was 0.1189 with standard error 0.0120. The 5% and 95% percentiles are, respectively, 0.1029 and 0.1408. Thus, an estimate for the estimator mean square error is

$$\mathrm{MSE}(ML) = 5.01 \times 10^{-4}. \tag{12}$$

To evaluate the estimator MSE we will compare these results with the ones obtained using the Dorfman's procedure. The optimal batch size for $p = 0.1$ is $n = 4$. By (8), (9) and (10), the mean square error is given by

$$\mathrm{MSE}(\widehat{p}) = 0.015671^2 + \frac{0.102291}{400} = 5.01 \times 10^{-4}. \tag{13}$$

The MSE is the same for both methods. Moreover both present a problem of overestimation due to the test sensitivity and specificity.

However, when the number of rows and columns is just as high as 3 or more it is not easy to use the previous method to compute a value of the ML function. In this case, we suggest the computation of an estimate for the ML function value for a given prevalence rate in the following way.

1. Record in a matrix O of size $r \times c$ the number of two-dimensional arrays with $i-1$ positive rows and $j-1$ positive columns where $i = 1, \cdots, r$ and $j = 1, \cdots, c$.
2. For some possible prevalence rate values p, chosen in some logical sequence (for instance, $0, 0.1, 0.2, \cdots, 1$) simulate a reasonable number of replicas rep of the possible matrices.
3. Compute the probability of observing $i-1$ positive rows and $j-1$ positive columns for each replica (taking into account the test sensitivity φ_s and the test specificity φ_e, and store that value in the position $(i-1, j-1)$ of the matrix P). Add the probabilities computed for all the replicas and multiply P by $1/rep$.
4. Compute the ML function for the matrix O using the values of P.

$$ML(p_0) = \prod_{i=1}^{r} \prod_{j=1}^{c} \left(P(i,j)\right)^{O(i,j)}.$$

5. Compare the ML function for each prevalence rate estimates and chose the two estimates with the highest ML function value.
6. Repeat the process from step two until the difference between the ML function at the two points chosen in step five be lower than some prefixed tolerance.
7. The estimate is the weighted mean value between those two points, say p_1 and p_2, using as weights $ML(p_1)$ and $ML(p_2)$, i.e.,

$$\widehat{p} = \frac{ML(p_1) \times p_1 + ML(p_2) \times p_2}{ML(p_1) + ML(p_2)}.$$

$P(i, j)$ is an estimate of the probability of having $i-1$ positive rows and $j-1$ positive columns in an array. In practice, the values for p in step two don't have to span all the interval $[0,1]$ as the use of pooled samples is advised only to prevalence rates lower than about $1/3$.

Let us look at the following example.

Example 3. Consider a matrix O generated by simulating 1000 replicas of a square array $A2(4:1)$ for a prevalence rate $p = 0.01$ and $\varphi_s = \varphi_e = 0.99$ using software MatLab R2011.

$$O = \begin{pmatrix} 76 & 4 & 0 & 0 & 0 \\ 4 & 14 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The total proportion of infected individuals of this simulation was, by chance, equal to the prevalence rate $p = 0.01$.

For computing a ML estimate for the prevalence rate given this matrix O, 100 square arrays were simulated to compute each matrix P. Hence, the sample size is $4^2 \times 100 = 1600$. Then, 100 estimates were, independently, computed. The estimates mean value was 0.0148 with standard error 0.0027. The 5% and 95% percentiles are, respectively, 0.0113 and 0.0183. Thus, an estimate for the estimator mean square error is

$$\mathrm{MSE}(ML) = 3.02 \times 10^{-5}. \tag{14}$$

Once again, in order to evaluate the estimator MSE we will compare these results with the ones obtained using the Dorfman's procedure. The optimal batch size for $p = 0.01$ is $n = 11$. The mean square error is given by

$$\mathrm{MSE}(\widehat{p}) = 0.00177^2 + \frac{0.011636}{1600} = 1.04 \times 10^{-5}. \tag{15}$$

The MSE of both estimators are similar. Thus, the estimates given by this algorithm seem to be reliable. However, we are not performing a formal comparison between the two methods as there isn't no way, at least to our knowledge, to find the optimal array-based group testing design for a given estimation problem (unless one supposes there is no test errors, cf. [11]).

4 Final Remarks

The main achievement of this work is the dropping of the individual tests when we just want to determine a prevalence rate estimate.

When one is dealing with an hierarchical method, Table 1 shows how to interpret a result at the last stage of the procedure in order to produce an estimate. It will be at least very close to the proportion of defective subjects computed using individual tests at that stage.

The use of square array methodologies is only possible with the advent of robotic pooling. These methods can be very efficient if no individual tests are performed. The iterative method for computing a ML estimate allows the use of that kind of strategies and the computational cost does not have to be very high in order to obtain accurate estimates (comparing to Dorfman's procedure). One problem that stills unsolved is to find a method to easily identify the best array to use in a given situation. This issue will be dealt in a future work. A generalization of this iterative method to higher dimensional arrays is straightforward. More details on the use of arrays with dimensions higher than two is discussed by [11].

Acknowledgements Research partially sponsored by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal − FCT under the project (PEst-OE/MAT/UI0006/2011).

References

1. Berger, T., Mandell, J.W., Subrahmanya, P.: Maximally efficient two-stage screening. Biometrics **56**, 833–840 (2000)
2. Bilder, C.R., Zang, B., Schaarschmidt, F., Tebbs, J.M.: binGroup: a package for group testing. The R Journal **2**, 56–60 (2010)
3. Chen, C.L., Swallow, W.H.: Using group testing to estimate a proportion, and to test the binomial model. Biometrics **46**, 1035–1046 (1990)
4. Dorfman, R.: The detection of defective members in large populations. Ann. Math. Stat. **14**, 436–440 (1943)
5. Finucan, H.M.: The blood testing problem. App. Stat. **13**, 43–50 (1964)
6. Garner, F.C., Stapanian, M.A., Yfantis, E.A., Williams, L.R.: Probability estimation with sample compositing techniques. J. Off. Stat. **5**, 365–374 (1989)
7. Hung, M., Swallow, W.H.: Robustness of group testing in the estimation of proportions. Biometrics **55**, 231–237 (1999)
8. Hughes-Oliver, J.M.: Pooling experiments for blood screening and drug discovery. In: Dean, A., Lewis, S. (eds.) Screening — Methods for Experimentation in Industry, Drug Discovery, and Genetics, 46-68. Springer, New York
9. Hwang, F.K.: Group testing with a dilution effect. Biometrika **63**, 671–673 (1976)
10. Kim, H., Hudgens, M., Dreyfuss, J., Westreich, D., Pilcher, D.: Comparison of group testing algorithms for case identification in the presence of testing errors. Biometrics **63**, 1152–1163 (2007)
11. Kim, H., Hudgens, M.: Three-Dimensional Array-Based Group Testing Algorithms. Biometrics **65**, 903–910 (2009)
12. Lancaster, V.A., Keller-McNulty, S.: A review of composite sampling methods. J. Am. Stat. Assoc. **93**, 1216–1230 (1998)
13. Liu, S.C., Chiang, K.S., Lin, C.H., Chung, W.C., Lin, S.H., Yang, L.C.: 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 (2011)
14. Loyer, M.W.: Bad probability, good statistics, and group testing for binomial estimation. Am. Stat. **37**, 57–59 (1983)

15. Martins, J.P., Felgueiras, M., Santos, R.: Meta-analysis techniques applied in prevalence rate estimation. Discuss. Math. Probab. Stat. (in press)
16. Phatarfod, C.D., Sudbury, A.: The use of a square array scheme in blood testing. Stat. Med. **13**, 2337–2343 (1994)
17. Santos, R., Pestana, D., Martins, J.P.: Extensions of Dorfman's theory. In: Oliveira, P.E., da Graa Temido, M, Henriques, C., Vichi, M. (eds.) Studies in Theoretical and Applied Statistics, Recent Developments in Modeling and Applications in Statistics, pp. 179-189. Springer, New York (2013)
18. Sobel, M., Elashoff, R.M.: Group testing with a new goal, estimation. Biometrika **62**, 181–193 (1975)
19. Sterret, A.: On the detection of defective members of large populations. Ann. Math. Stat. **28**, 1033–1036 (1957)
20. Wein, L.M., Zenios, S.A.: Pooled testing for HIV screening: capturing the dilution effect. Oper. Res. **44**, 543–569 (1996)
21. Zenios, S., Wein, L.: Pooled testing for HIV prevalence estimation exploiting the dilution effect. Stat. Med. **17**, 1447-1467 (1998)