Discussiones Mathematicae
Probability and Statistics 33 (2013) 79–97
doi:10.7151/dmps.1151

META-ANALYSIS TECHNIQUES APPLIED IN PREVALENCE RATE ESTIMATION

João Paulo Martins

School of Technology and Management, Polytechnic Institute of Leiria
CEAUL – Center of Statistics and Applications of University of Lisbon

e-mail: jpmartins@ipleiria.pt

Miguel Felgueiras

School of Technology and Management, Polytechnic Institute of Leiria
CEAUL – Center of Statistics and Applications of University of Lisbon
CIIC – Computer Science and Communications Research Centre of Polytechnic Institute of Leiria

e-mail: mfelg@ipleiria.pt

AND

Rui Santos

School of Technology and Management, Polytechnic Institute of Leiria
CEAUL – Center of Statistics and Applications of University of Lisbon

e-mail: rui.santos@ipleiria.pt

Abstract

In some cases, the estimators obtained in compound tests have better features than the traditional ones, obtained from individual tests, cf. Sobel and Elashoff (1975), Garner *et al.* (1989) and Loyer (1983). The bias, the efficiency and the robustness of these estimators are investigated in several papers, e.g. Chen and Swallow (1990), Hung and Swallow (1999) and Lancaster and Keller-McNulty (1998). Thus, the use of estimators based on compound tests not only allows a substantial saving of costs, but they also can (in some situations) be more accurate than the estimators based on the individual tests.

Nevertheless, each laboratory produces estimates for the prevalence rate of a given infection using different methodologies, such as halving nested

procedures (Sobel and Elashoff, 1975) and square array testing (Kim *et al.*, 2007). The logistic regression or the weighted least squares regression can be used in order to combine different prevalence rate estimates (Chen and Swallow, 1990). In this work some meta-analytical techniques are proposed as an alternative approach. This methodology has the advantage of being quite simple and flexible to account for the error source.

Keywords: compound tests, estimation of prevalence, meta-analysis, sensitivity, specificity.

2010 Mathematics Subject Classification: 62F10, 62P10.

1. Introduction

Dorfman (1943) has been the first to use group testing. Individuals were gathered into pools to screen for a binary characteristic (presence or absence of the syphilis antigen) in order to reduce the costs. A negative result on a pooled mixture of blood from n people 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 how many or which ones. In this case, performing individual tests is advised to identify the individual positives in the sample from the individual negatives. The optimal batch size minimizes the expected number of tests as the cost of mixing samples is usually negligible (cf. Liu *et al.* (2011)).

Since Dorfman's seminal work, the research on methodologies involving pooled sample tests has been quite active (Hughes-Oliver, 2006). Moreover, the use of pooled samples does not refer only to the classification problem (identifying all the infected individuals in a sample), since it may also be useful in estimating the prevalence rate p, as Sobel and Elashoff (1975) stated. When the main issue is the estimation problem, the performing of individual tests is only optional, since the goal is no longer to dentify the infected individuals. The use only of 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. Sobel and Elashoff (1975), Loyer (1983) and Garner *et al.* (1989). The bias, the efficiency and the robustness of these estimators have been reviewed in several works, such as those from Chen and Swallow (1990), Lancaster and Keller-McNulty (1998) or Hung and Swallow (1999). Bilder *et al.* (2010) propose 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 (Kim *et al.*, 2007). 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 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 depend on the amount of samples available and the sensitivity, the specificity and the monetary costs of the process (Liu *et al.*, 2011). For an overview about this problem, known as the dilution problem, see Hwang (1976), Wein and Zenios (1996), Zenios and Wein (1998) and Santos *et al.* (2012).

Thus, when estimating an infection prevalence rate, each laboratory may use a different procedure even if the sample size is equal. Moreover, the variable measure for screening the infected individuals could be either qualitative (presence or absence of the infection) or quantitative (an individual is declared infected if the amount of substance detected exceeds some threshold l). The aim of this work is to develop meta-analysis techniques that could allow a researcher to combine different prevalence rate estimates obtained from possible different experimental designs and different estimators. This is quite relevant since the meta-analysis techniques addresses the problem of combining different estimates obtained from similar processes. If there is any differences between the ways studies are performed, this is usually accounted by the use of some covariate(s). The proposed method to address this problem involves the process sensitivity and specificity. Hence, some calculations of these measures are done to some of the most common methods that resort to group testing: hierarchical algorithms and square array testing.

The outline of this work is as follows. Section 2 introduces the binomial model assumption. Section 3 describes some group testing procedures and its error measures. Some new expressions for the sensitivity and specificity of some processes are obtained. In the last section, the two paradigms of meta-analysis are briefly discussed. Subsection 3.3 is the core of this work: it develops an algorithm for combining different estimates with underlying different group testing procedures. An example is also given and some simulation results are presented to acknowledge the importance of knowing the error rates of the different procedures in producing an accurate global estimate.

2. Binomial model

Let $X \frown \text{Binomial}(n, p)$ where $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ and p is unknown and consider the estimator $\widehat{p}_n = \frac{X_n}{n}$. As the estimator mean value is $\mathrm{E}[\widehat{p}_n] = p$ and its variance is $\mathrm{Var}[\widehat{p}_n] = \frac{p(1-p)}{n}$, then $\widehat{p}_n$ weakly converges, as $n \to \infty$, for a normal

random variable distribution

$$\sqrt{n}\left(\widehat{p}_n - p\right) \xrightarrow[n\to\infty]{\mathrm{d}} Z \frown \mathcal{N}\left(0, \sqrt{p\left(1-p\right)}\right). \tag{1}$$

The normal distribution in (1) depends on the unknown parameter p. To overcome this situation it may be used a variance stabilizing transformation, for instance (cf. Johnson *et al.*, 1993),

$$h\left(x\right) = 2\sqrt{n}\arcsin\left(\sqrt{x}\right). \tag{2}$$

Next, a variance stabilizing transformation definition is given.

Definition. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables verifying

$$\sqrt{n}\left(X_n - \theta\right) \xrightarrow[n\to\infty]{d} Z \frown \mathcal{N}\left(0, \sigma\right),$$

then $g : \mathbb{R} \to \mathbb{R}$ is called a variance stabilizing transformation of $\{X_n\}_{n\in\mathbb{N}}$ if $a_n(g(X_n) - g(\theta))$ has an asymptotic normal distribution $\mathcal{N}(0, c)$ where $\{a_n\}_{n\in\mathbb{N}}$ is a sequence of real numbers and $c > 0$ does not depend on θ.

Holland (1973) defines this type of transformations as asymptotically stabilizing variance transformations.

Anscombe (1948) has shown that $2\sqrt{n}\arcsin\left(\sqrt{\widetilde{p}}\right)$, where $\widetilde{p} = \frac{X+3/8}{n+3/4}$, increases the convergence rate to a normal random variable with unit variance and mean $2\sqrt{n}\arcsin\left(\sqrt{p}\right)$.

From (1), it is readily established a $(1-\alpha)\times 100\%$ confidence interval for p

$$\left]\left\{\sin\left(\arcsin\left(\sqrt{\widetilde{p}}\right) - \frac{z_{1-\alpha}}{2\sqrt{n}}\right)\right\}^2, \left\{\sin\left(\arcsin\left(\sqrt{\widetilde{p}}\right) + \frac{z_{1-\alpha}}{2\sqrt{n}}\right)\right\}^2\right[, \tag{3}$$

where $z_{1-\alpha}$ is the quantile $1-\alpha$ of the standard normal distribution. This interval suffers from overconservatism for p near 0 and 1 (cf. Kulinskaya *et al.*, 2008).

One alternative is to use the normal approximation $\mathcal{N}\left(\ln\left(p\right), \left(1-p\right)/np\right)$ to the distribution of $\ln\left(\widetilde{p}\right)$. Hence, other nominal $(1-\alpha)\times 100\%$ confidence interval for p is

$$\left]\exp\left(\ln\left(\widetilde{p}\right) - z_{1-\alpha}\left(\left(1-\widetilde{p}\right)/n\widetilde{p}\right)^{1/2}\right), \exp\left(\ln\left(\widetilde{p}\right) + z_{1-\alpha}\left(\left(1-\widetilde{p}\right)/n\widetilde{p}\right)^{1/2}\right)\right[. \tag{4}$$

Both intervals suffer from high variations in its effective confidence as p decreases to zero. This results from the highly skewed nature of the binomial distribution when p is not close to $1/2$. The log-transformation is more accurate than the defined by (1) when p is close to zero (cf. Kulinskaya *et al.*, 2008).

3. GROUP TESTING PROCEDURES

The accuracy of an estimate of a prevalence rate p is strongly related to the size of the sample which has been used. However, when group testing procedures are used it is necessary to assess the quality of the estimate, which is related to the quality of the results of the experimental test. Thus, consider the problem of estimating the prevalence rate of some disease and let $X_i = 1$ denote an infected individual and $X_i = 0$ denote a non-infected individual. Let $\mathcal{M}$ denote the chosen procedure. Hence, the test sensitivity is equal to $\varphi_s = P\left(X_i^+ | X_i = 1\right)$ where X_i^+ stands for a positive test result. The test specificity is given by $\varphi_e = P\left(X_i^- | X_i = 0\right)$ where X_i^- stands for a negative test result.

Finally, for evaluating the quality of the procedure $\mathcal{M}$, the pooling sensitivity and the pooling specificity as defined by Kim *et al.* (2007) will be used. The pooling sensitivity or the sensitivity of the process measure the probability of an infected individual be correctly identified by the methodology $\mathcal{M}$, that is, $\varphi_s^{\mathcal{M}} = P_{\mathcal{M}}(+|X_i = 1)$. The pooling specificity or the specificity of the process stands for the probability of a non-infected individual be correctly classified by the methodology $\mathcal{M}$, that is, $\varphi_e^{\mathcal{M}} = P_{\mathcal{M}}(-|X_i = 0)$. For an individual testing procedure the sensitivity (specificity) of the process is equal to the test sensitivity (specificity).

Dorfman's procedure is just the first procedure presented from a wider family called hierarchical algorithms. The extensions of this work (cf. Finuncan (1964), Sterret (1957), Wein and Zenios (1996)) suggest 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. Johnson *et al.* (1991) computed expressions for the error rates of these processes: the sensitivity and the specificity of the process (the authors named these quantities as pooling sensitivity and pooling specificity, respectively, but only considered the classification problem). Besides, the authors do not take into account the dilution effect as the sensitivity and specificity of a compound test must depend on the number of infected individuals in the group. This is no surprise, since in the literature this effect is usually disregarded. Moreover, when the dilution effect is considered (e.g. Wein and Zenios (1996) and Zenios and Wein (1998)) the construction of hierarchical models to capture the dilution effect does not measure the influence of dilution and rarefaction (decreasing of the amount of substance per volume unit when infected and non-infected samples are mixed for batched testing) on the sensitivity and

specificity of the chosen process. We will consider both problems and the dilution effect to establish the sensitivity and specificity of the processes.

For practical reasons, only two or three stages are usually performed. Hence, next we just generalize the formula for the sensitivity and the specificity of the two-stages procedure discussed in Santos *et al.* (2012) and compute those quantities for the three stages case.

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, when we just want to estimate the prevalence rate, this might not be the case and the condition verified is just $n_1 > \cdots > n_s \geq 1$ (cf. Chen and Swallow (1990), Hung and Swallow (1999), Lancaster and Keller-McNulty (1998)). For low prevalence rates, the use of $n_s > 1$ 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. Santos *et al.*, 2012). 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

In an infected sample at the s-th stage, if the test outcome is positive 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 well classified.

Suppose that the tests results are independent. Let $I^{[n]} = \sum_{i=1}^{n} X_i$ represent the number of infected elements in a sample of size n and $I^{[i,n]}$ denote the probability $\mathrm{P}(I^{[n]} = i) = \binom{n}{i} p^i q^{n-i}, i = 0, \ldots, n$. Let $X^{[+,n]}$ [resp. $X^{[-,n]}$] represent a positive [resp. negative] result on the compound test with n individuals and denote $\varphi_s^{[m,n]} = \mathrm{P}(X^{[+,n]} | I^{[n]} = m)$. For $s = 2$, the sensitivity of the process is given by the probability of an individual being correctly identified as infected. This happens when in both stages the pooled or individual sample is classified as positive

$$\begin{aligned}
\varphi^{H}_{s_{n_1,n_2}} &= \mathrm{P}\left(X_1^{+}|X_1=1\right)\\
&= \sum_{i=0}^{n_1-1}\mathrm{P}\left(X^{[+,n_2]}|X_1=1,X^{[+,n_1]}\right)\mathrm{P}\left(X^{[+,n_1]}|X_1=1,I^{[n_1-1]}=i\right)\mathrm{P}\left(I^{[n_1-1]}=i\right)\\
&= \sum_{i=0}^{n_1-1}\sum_{j=\max(0,n_2-i-1)}^{\min(i,n_2-1)}\frac{\binom{i}{j}\binom{n_1-i-1}{n_2-j-1}}{\binom{n_1-1}{n_2-1}}\mathrm{P}\left(X^{[+,n_2]}|I^{[n_2]}=j+1\right)\mathrm{P}\left(X^{[+,n_1]}|I^{[n_1]}=i+1\right)I^{[i,n_1-1]}\\
&= \sum_{i=0}^{n_1-1}\sum_{j=\max(0,n_2-i-1)}^{\min(i,n_2-1)}\frac{\binom{i}{j}\binom{n_1-i-1}{n_2-j-1}}{\binom{n_1-1}{n_2-1}}\varphi_s^{[j+1,n_2]}\varphi_s^{[i+1,n_1]}I^{[i,n_1-1]}.
\end{aligned}$$

Note that, disregarding the subject X_1, the number of infected individuals distribution at the second stage given i infected individuals at the first stage is an hypergeometric distribution $\mathrm{H}(n_1-1,n_2-1,i)$ where n_1-1 is the population size, i is the number of successes within the population and n_2-1 is the number of draws. The specificity of the process is given by the probability of a non-infected individual being correctly identified $\left(X_1^{-}\right)$. This is the case when the test outcome of the pooled sample at the first stage is negative and when it is positive but, at the second stage, the pooled sample is whether infected or screened as negative. Hence, four cases have to be considered.

$$\begin{aligned}
\varphi^{H}_{e_{n_1,n_2}} &= \mathrm{P}\left(X_1^{-}|X_1=0\right)\\
&= \sum_{i=0}^{a_2}\left(\mathrm{P}\left(X^{[-,n_1]}|I^{[n_1-1]}=i\right)+\mathrm{P}\left(X^{[-,n_2]}|X^{[+,n_1]}\right)\mathrm{P}\left(X^{[+,n_1]}|I^{[n_1-1]}=i\right)\right)\mathrm{P}\left(I^{[n_1-1]}=i\right)\\
&= \left[\varphi_e+\varphi_e\left(1-\varphi_e\right)\right]q^{n_1-1}+\sum_{i=1}^{a_2}\left[\left(1-\varphi_s^{[i,n_1]}\right)+\mathrm{P}\left(X^{[-,n_2]}|X^{[+,n_1]}\right)\varphi_s^{[i,n_1]}\right]I^{[i,n_1-1]}\\
&= \alpha q^{a_2}+\sum_{i=1}^{a_2}\left[\varphi_e\frac{\binom{a_2-i}{n_2-1}}{\binom{a_2}{n_2-1}}\varphi_s^{[i,n_1]}+\sum_{j=b_1}^{b_2}\frac{\binom{i}{j}\binom{a_2-i}{n_2-j-1}}{\binom{a_2}{n_2-1}}\left(1-\varphi_s^{[j,n_2]}\right)\varphi_s^{[i,n_1]}+\left(1-\varphi_s^{[i,n_1]}\right)\right]I^{[i,a_2]},
\end{aligned}$$

where $\alpha=2\varphi_e-\varphi_e^2$, $q=1-p$, $a_2=n_1-1$, $b_1=\max\left(1,n_2-i-1\right)$, $b_2=\min\left(i,n_2-1\right)$ and $1-\varphi_s^{[0,n_2]}$ stands for φ_e. Observe that n_1 and n_2 do not affect the probability of getting a negative outcome when testing a non-infected pooled sample since the dilution effect occurs when at least one infected individual is part of the mixed sample. These results are just a straightforward generalization of Santos *et al.* (2012) result for $n_2=1$.

For $s=3$, an infected individual is correctly screened if at every stage the test outcome is positive. Thus, omitting the sum limits

$$\begin{aligned}
\varphi^{H}_{s_{n_1,n_2,n_3}} &= \mathrm{P}\left(X_1^{+}|X_1=1\right)\\
&= \sum_i\sum_j\sum_k\mathrm{P}\left(X^{[+,n_3]}|X^{[+,n_1]},X^{[+,n_2]}\right)\mathrm{P}\left(X^{[+,n_2]}|X^{[+,n_1]}\right)\mathrm{P}\left(X^{[+,n_1]}|I^{[n_1-1]}=i\right)I^{[i,n_1-1]}\\
&= \sum_i\sum_j\sum_k\frac{\binom{j}{k}\binom{n_2-j-1}{n_3-k-1}}{\binom{n_2-1}{n_3-1}}\frac{\binom{i}{j}\binom{n_1-i-1}{n_2-j-1}}{\binom{n_1-1}{n_2-1}}\varphi_s^{[k+1,n_3]}\varphi_s^{[j+1,n_2]}\varphi_s^{[i+1,n_1]}I^{[i,n_1-1]},
\end{aligned}$$

where i,j,k stand for the number of infected individuals at the first, second and third stage respectively.

The specificity of the process is computational demanding since several situations are possible when a non-infected individual is screened correctly. We will consider the cases of having the first negative outcome at the first, second and third stage – $\varphi^{H,1}_{e_{n_1,n_2,n_3}}, \varphi^{H,2}_{e_{n_1,n_2,n_3}}, \varphi^{H,3}_{e_{n_1,n_2,n_3}}$. The sum upper limits are once again omitted. Therefore

$$\begin{aligned}\varphi^{H,1}_{e_{n_1,n_2,n_3}} &= \sum_{i=0}^{a_2} \mathrm{P}\left(X_1^{[-,n_1]} | X_1 = 0, I^{[n_1-1]} = i\right) I^{[i,n_1-1]} \\ &= \varphi_e q^{n_1-1} + \sum_{i=1}^{a_2} \left(1 - \varphi_s^{[i,n_1]}\right) I^{[i,n_1-1]}.\end{aligned}$$

Concerning the process specificity,

$$\begin{aligned}\varphi^{H,2}_{e_{n_1,n_2,n_3}} &= \sum_{i=0}^{a_2} \mathrm{P}\left(X^{[-,n_2]}, X^{[+,n_1]} | X_1 = 0, I^{[n_1-1]} = i\right) I^{[i,n_1-1]} \\ &= \left[\varphi_e\left(1-\varphi_e\right)\right] q^{n_1-1} + \sum_{i=1}^{a_2} \left[\mathrm{P}\left(X^{[-,n_2]} | X_1 = 0, X^{[+,n_1]}\right) \varphi_s^{[i,n_1]}\right] I^{[i,n_1-1]} \\ &= \left(\varphi_e - \varphi_e^2\right) q^{a_2} + \sum_{i=1}^{a_2} \left[\varphi_e \frac{\binom{n_1-i-1}{n_2-1}}{\binom{n_1-1}{n_2-1}} \varphi_s^{[i,n_1]} + \sum_{j=b_1}^{b_2} \left(1-\varphi_s^{[j,n_2]}\right) \frac{\binom{i}{j}\binom{n_1-i-1}{n_2-j-1}}{\binom{n_1-1}{n_2-1}} \varphi_s^{[i,n_1]}\right] I^{[i,n_1-1]}\end{aligned}$$

and

$$\begin{aligned}\varphi^{H,3}_{e_{n_1,n_2,n_3}} &= \sum_{i=0}^{a_2} \mathrm{P}\left(X^{[-,n_3]}, X^{[+,n_1]}, X^{[+,n_2]} | X_1 = 0, I^{[n_1-1]} = i\right) I^{[i,n_1-1]} \\ &= \left[\varphi_e\left(1-\varphi_e\right)^2\right] q^{n_1-1} + \sum_{i=1}^{a_2} \left[\mathrm{P}\left(X^{[-,n_3]}, X^{[+,n_2]} | X_1 = 0, X^{[+,n_1]}\right) \varphi_s^{[i,n_1]}\right] I^{[i,n_1-1]} \\ &= \left[\varphi_e\left(1-\varphi_e\right)^2\right] q^{n_1-1} + \sum_{i=1}^{a_2} \frac{\binom{n_1-i-1}{n_2-1}}{\binom{n_1-1}{n_2-1}} \varphi_e\left(1-\varphi_e\right) \varphi_s^{[i,n_1]} \\ &\quad + \sum_{i=1}^{a_2} \sum_{j=b_1}^{b_2} \varphi_e \frac{\binom{n_2-j-1}{n_3-1}}{\binom{n_2-1}{n_3-1}} \varphi_s^{[j,n_2]} \varphi_s^{[i,n_1]} I^{[i,a_2]} \\ &\quad + \sum_{i=1}^{a_2} \sum_{j=b_1}^{b_2} \sum_{k=1}^{c_2} \left(1-\varphi_s^{[k,n_3]}\right) \frac{\binom{j}{k}\binom{n_2-j-1}{n_3-k-1}}{\binom{n_2-1}{n_3-1}} \varphi_s^{[j,n_2]} \frac{\binom{i}{j}\binom{n_1-i-1}{n_2-j-1}}{\binom{n_1-1}{n_2-1}} \varphi_s^{[i,n_1]} I^{[i,a_2]},\end{aligned}$$

where $c_1 = \max\left(1, n_3 - j - 1\right)$ and $c_2 = \min\left(j, n_3 - 1\right)$ Finally,

$$\varphi^{H}_{e_{n_1,n_2,n_3}} = \varphi^{H,1}_{e_{n_1,n_2,n_3}} + \varphi^{H,2}_{e_{n_1,n_2,n_3}} + \varphi^{H,3}_{e_{n_1,n_2,n_3}}.$$

Santos *et al.* (2012) have shown that, for low prevalence rates, in order to assess the dilution effect in the misclassification of an individual, knowing what happens

when just one (or two) infected elements are present in the pooled sample is sufficient. In that case, the previous formulas can be easily simplified.

Array-based specimen pooling is an alternative to hierarchical group testing that uses overlapping pools. In its simplest version (square array), a sample of size n^2 is placed in a $n \times n$ matrix. Then, all the samples within the same row and the same column are gathered for batched testing. So, this process involves at least $2n$ tests as subsequent individual tests may be performed. In a square array procedure without further testing, an individual is declared infected if both experimental tests to its "row" and "column" samples have a positive result. If individual testing is performed in the suspected infected samples (individuals whose pooled samples tested positive), a sample is screened as infected if and only if the row, column and individual test outcomes are all positive. A variant of this method, the so called square array with master pool testing, involves a initial test to a mixture with all the n samples. For the simplest case of a $n \times n$ square array testing with no subsequent individual testing, let $X_{i,j} = 1\,(X_{i,j} = 0)$ denote an infected (non-infected) individual at the i-th row and j-th column of the $n \times n$ matrix. Denote a positive (negative) result in the i-th row, j-th column and i-th row and j-th column cell by $X^+_{i,:}(X^-_{i,:}), X^+_{:,j}\left(X^-_{:,j}\right)$ and $X^+_{i,j}\left(X^-_{i,j}\right)$, respectively. The process sensitivity is equal to

$$\begin{aligned}\varphi^A_{s_{n^2}} &= \mathrm{P}\left(X^+_{i,j}|X_{i,j}=1\right)\\ &= \mathrm{P}\left(X^+_{i,:}, X^+_{:,j}|X_{i,j}=1\right).\end{aligned}$$

Assuming that given the true status of the i-th row and j-th column, the row and column tests are conditionally independent of each other,

$$\begin{aligned}\varphi^A_{s_{n^2}} &= \sum_{i=0}^{n-1}\mathrm{P}\left(X^+_{i,:}|I^{[i+1,n-1]}\right)I^{[i,n-1]}\sum_{j=0}^{n-1}\mathrm{P}\left(X^+_{:,j}|I^{[j+1,n-1]}\right)I^{[j,n-1]}\\ &= \sum_{i=1}^{n}\varphi_s^{[i,n]}I^{[i-1,n-1]}\sum_{j=1}^{n-1}\varphi_s^{[j,n]}I^{[j-1,n-1]}\\ &= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\varphi_s^{[i,n]}\varphi_s^{[j,n]}I^{[i-1,n-1]}I^{[j-1,n-1]}.\end{aligned}$$

Concerning the specificity, an individual is screened as non-infected if at least the i-th or j-th columns tests result are negative. Thus,

$$\begin{aligned}\varphi^{A}_{e_{n^2}} &= 1-\sum_{i=0}^{n-1}\mathrm{P}\left(X^{+}_{i,:}|X_{i,j}=0,I^{[i,n-1]}\right)I^{[i,n-1]}\sum_{j=0}^{n-1}\mathrm{P}\left(X^{+}_{:,j}|X_{i,j}=0,I^{[j,n-1]}\right)I^{[j,n-1]}\\ &= 1-\left((1-\varphi_e)\,q^{n-1}+\sum_{i=1}^{n-1}\varphi_s^{[i,n]}I^{[i,n-1]}\right)\left((1-\varphi_e)\,q^{n-1}+\sum_{j=1}^{n-1}\varphi_s^{[j,n]}I^{[j,n-1]}\right).\end{aligned}$$

For the others cases, Kim *et al.* (2007) give expressions for both measures. This square array design is the common example of a two dimensional procedure. Others two dimensional designs are possible as well extensions to higher dimensions (cf. Berger *et al.* (2000)) although its application in practice is very rare.

4. Meta-analysis with k studies

For the application of the meta-analysis techniques (combination of the information provided from different studies) it is necessary to decide on the nature of the parameter of interest. If there is evidence to assume that the parameter is the same across the different studies, a fixed effects model (subsection 4.1) is recommended. But, if the parameter is in fact a random variable with possible different values in each study, the use of a random effects model is advised (subsection 4.2). In this case, the use of covariates in a meta-regression study is an useful tool for assessing the variance between studies. There are several ways for deciding the kind of model that best suits the meta-analysis (Hartung *et al.*, 2008).

4.1. Fixed effects model

Suppose that k studies concerning the estimation of some prevalence rate $p_1,\ldots,p_k$ are available. In a fixed effects model it is assumed that $p_1=\cdots=p_k=\theta$.

Let E_i be the estimator used in the i-th study. An overall estimator is given by

$$\widehat{\theta}=\frac{\sum_{i=1}^{k}w_iE_i}{\sum_{i=1}^{k}w_i}, \tag{5}$$

where $\widehat{\theta}$ is a convex linear combination of the E_i estimators. Since the within study variances are usually unknown, the weights w_i are chosen as the inverse of the estimated effect size variance

$$w_i=\frac{1}{\widehat{\mathrm{Var}\,(E_i)}}, \tag{6}$$

in order to reduce the overall estimator $\widehat{\theta}$ variance.

As Hartung *et al.* (2008) point out, it is also useful to attach some quality index q_i to the i-th study along with the nonnegative weights w_i's. Thus, yielding the following estimator of θ

$$\widehat{\theta}^* = \frac{\sum_{i=1}^{k} q_i w_i E_i}{\sum_{i=1}^{k} q_i w_i} \tag{7}$$

with estimated asymptotic variance

$$\widehat{\sigma}^2 \left[\widehat{\theta}^*\right] = \widehat{\mathrm{Var}} \left[\widehat{\theta}^*\right] = \frac{\sum_{i=1}^{k} q_i^2 w_i^2 \widehat{\mathrm{Var}} \left[E_i\right]}{\left(\sum_{i=1}^{k} q_i w_i\right)^2}. \tag{8}$$

A $(1-\alpha)\times 100\%$ confidence interval for θ is given by

$$\left]\widehat{\theta}^* - z_{1-\alpha/2}\widehat{\sigma}_{\widehat{\theta}^*}, \widehat{\theta}^* + z_{1-\alpha/2}\widehat{\sigma}_{\widehat{\theta}^*}\right[. \tag{9}$$

4.2. Random effects model

In a random effects model the parameter p is described by some distribution F. Let μ stand for the random variable mean value and τ^2 stand for the random variable variance (between study variance). The estimator $\widehat{\theta}^*$ of μ may be updated to (cf. Kulinskaya *et al.*, 2008)

$$\widehat{\theta}^*(\tau) = \frac{\sum_{i=1}^{k} q_i w_i(\tau) E_i}{\sum_{i=1}^{k} q_i w_i(\tau)}, \tag{10}$$

where

$$w_i(\tau) = 1/\left(\tau^2 + \widehat{\sigma}_i^2(\theta_i)\right) \tag{11}$$

and $\widehat{\sigma}_i^2$ represents the within study variance.

For estimating τ^2, the DerSimonian and Laird (1986) estimator defined by

$$\widehat{\tau}_{DL}^2 = \max\left(0, \frac{Q-(k-1)}{\sum_{i=1}^{k}\widehat{w}_i - \sum_{i=1}^{k}\widehat{w}_i^2 / \sum_{i=1}^{k}\widehat{w}_i}\right), \tag{12}$$

where Q is the commonly used Cochran's statistic (cf. Cochran, 1954). Rukhin, Biggerstaff and Vagel (2000) provide the equations for the maximum likelihood estimator and for the restricted maximum likelihhod estimator. Biggerstaff and Tweedie (1997) provide confidence intervals on τ.

4.3. Prevalence rate estimation

There are few papers on the issue of combining prevalence rate estimates from different studies. One of the reasons is that we always have the obvious solution of presenting an overall estimate as defined in (7) or (10). However, as far as we known, none performing quality index has been defined for the estimators. However, two alternatives were presented by Chen and Swallow (1990), but they do not use an overall estimator. These author's methods are based in the computation of the slope of linearized logistic regression model. This process also does not consider possible different experimental designs.

Our suggestion is to use the probability of denoting a specimen as positive $\left(X_1^+\right)$ when a methodology $\mathcal{M}$ is chosen (cf. Santos *et al.*, 2012)

$$\begin{aligned} p^{\mathcal{M}} &= P_{\mathcal{M}}\left(X_1^+|X_1=1\right)P\left(X_1=1\right)+P_{\mathcal{M}}\left(X_1^+|X_1=0\right)P\left(X_1=0\right) \\ &= \varphi_s^{\mathcal{M}}p+\left(1-\varphi_e^{\mathcal{M}}\right)(1-p) \\ &= 1-\varphi_e^{\mathcal{M}}+\left(\varphi_s^{\mathcal{M}}+\varphi_e^{\mathcal{M}}-1\right)p, \end{aligned}$$

where $\varphi_s^{\mathcal{M}}$ and $\varphi_e^{\mathcal{M}}$ stand for the process sensitivity and the process specificity, as previously defined. Thus, the number of specimens screened as positive follows a binomial distribution $B\left(N,p^{\mathcal{M}}\right)$. Hence, the expected number of specimens denoted as defective is equal to $Np^{\mathcal{M}}$. Solving the previous equation in order to p,

$$p=\frac{p^{\mathcal{M}}+\varphi_e^{\mathcal{M}}-1}{\varphi_s^{\mathcal{M}}+\varphi_e^{\mathcal{M}}-1}, \tag{13}$$

we find the following estimator for p

$$\widetilde{p}=\frac{\frac{\sum_{i=1}^N Y_i}{N}+\varphi_e^{\mathcal{M}}-1}{\varphi_s^{\mathcal{M}}+\varphi_e^{\mathcal{M}}-1} \tag{14}$$

for $1-\varphi_e<\frac{\sum_{i=1}^N Y_i}{N}<\varphi_s$ and $\varphi_s^{\mathcal{M}}+\varphi_e^{\mathcal{M}}>1$ where Y_i's are independent Bernoulli random variables ($Y_i=1(0)$ stands for a positive (negative) process classification of the i-th individual) and N is the sample size. The restriction $\frac{\sum_{i=1}^N Y_i}{N}<\varphi_s$ could be dropped in practice as the prevalence rate p is low when group testing procedures are applied (cf. Hung and Swallow, 1999). Hence, it is expected that condition to verify at least for a reasonable sample size N. For reasonable process sensitivity and process specificity the inequality $\varphi_s^{\mathcal{M}}+\varphi_e^{\mathcal{M}}>1$ is also verified. The condition $\frac{\sum_{i=1}^N Y_i}{N}>1-\varphi_e$ is very important as the expected number of false positives (given by $(1-\varphi_e)N$) is higher than the expected number of infected individuals (pN). Moreover, the number of individuals screened as positive is also

raised by the number of infected individuals correctly classified. Our estimator can not be applied in this case as it will lead to a negative estimate. For avoiding an overestimation of p we do not recommend the use of any estimator in these conditions.

When using group testing procedures, it is not always possible to get an observed value of $\sum_{i=1}^{n} Y_i$ directly. Loyer (1983) and Hung and Swallow (1999) discuss the estimation of the prevalence rate in hierarchical algorithms procedures and Xie *et al.* (2001) consider the case of square array algorithms.

Let $\widetilde{p}_i$ denote the estimator for p in the i-th study and consider the stabilizing variance transformation $g(x) = 2\sqrt{n}\arcsin(\sqrt{x})$ defined in (2). We advise the use of the estimator of

$$h(\widetilde{p}_i) = \sqrt{n}\arcsin\left(\sqrt{\widetilde{p}_i}\right). \tag{15}$$

The estimator mean $\mathrm{E}[h(\widetilde{p}_i)]$ is asymptotically equal to $h(p)$, i.e., $\mathrm{E}[h(\widetilde{p}_i)] \doteq h(p)$. This estimator mean square error is

$$\mathrm{MSE}[h(\widetilde{p}_i)] = (\mathrm{E}(h(\widetilde{p}_i)) - h(p))^2 + \mathrm{Var}[h(\widetilde{p}_i)]. \tag{16}$$

Thus, we suggest as weights

$$w_i = \frac{1}{\mathrm{Var}[h(\widetilde{p}_i)]}, \tag{17}$$

and as quality index

$$q_i = \frac{1}{(h(\widetilde{p}_i) - h(p))^2}. \tag{18}$$

Finally, an overall estimator is given by

$$h(\overline{p}) = \frac{\sum w_i q_i h(\widetilde{p}_i)}{\sum w_i q_i}. \tag{19}$$

The use of the estimators $h(\widetilde{p}_i)$ reduces the $h(\overline{p})$ estimator variance. Besides, it has the advantage that, even for a small study sample size, an estimate of the within study variance will not be required. However, note that those estimators are not unbiased for $h(p)$, since h is a non-linear transformation. Note that the meaning of an estimate given by $\overline{p}$ depends on wether a fixed effects or a random effects model is assumed.

Example 1. Let $X_n \frown \mathrm{Binomial}(n, p)$ and consider the estimator defined in (15). Since, as $n \to \infty$,

$$\left(\arcsin\sqrt{\widetilde{p}_n} - \arcsin\sqrt{p_n}\right) \xrightarrow[n\to\infty]{\mathrm{d}} T \frown \mathcal{N}\left(0, \frac{1}{2\sqrt{n}}\right), \tag{20}$$

then

$$h\left(\overline{p}\right) = \frac{\sum w_i q_i \arcsin\sqrt{\widetilde{p}_i}}{\sum w_i q_i}. \tag{21}$$

From (17) and (20) we get

$$w_i = 4n \tag{22}$$

and by (18) we have

$$\begin{aligned} q_i &= \frac{1}{\left(h(\widetilde{p}_i) - h\left(p\right)\right)^2} \\ &= \frac{1}{\left(\arcsin\left(\sqrt{\widetilde{p}_i}\right) - \arcsin\left(\sqrt{p}\right)\right)^2}. \end{aligned} \tag{23}$$

However, the estimator $h\left(\overline{p}\right)$ depends on the unknown parameter θ! To overcome this problem we suggest the following algorithm.

Algorithm.

Let $h_0\left(\overline{p}\right)$ be an initial estimate and then compute recursively new estimates according to the relation

$$h_{j+1}\left(\overline{p}\right) = \frac{\sum_{i=1}^{k} w_i \times q_i \times h_j\left(\widetilde{p}_i\right)}{\sum_{i=1}^{k} w_i \times q_i}. \tag{24}$$

The process stops when a new estimate differs from the previous one less than some tolerance ε (previously established). There is a drawback in this process because

$$q_i = \frac{1}{\left(h\left(\widetilde{p}_i\right) - h\left(p\right)\right)^2}$$

and the denominator may assume the zero value. This is easily overcome, replacing q_i by

$$q_i = \frac{1}{\left(\max\left(t, h\left(\widetilde{p}_i\right) - h\left(p\right)\right)\right)^2},$$

where t is some value close to zero. The process convergence is not affected by this minor change.

Example 2. The table below contains the data on 13 trials on the prevention of tuberculosis using BCG vaccination that may be found in Hartung *et al.* (2008).

Table 2. Data from a meta-analysis study on the effect of BCG.

Study	Infected	Non-infected
1	4	119
2	6	300
3	3	228
4	62	13536
5	33	5036
6	180	1361
7	8	2537
8	505	87886
9	29	7470
10	17	1699
11	186	50448
12	5	2493
13	27	16886

In study 6 the proportion of infected individuals is at least 3 times higher than in the other studies. So, suppose that for all studies the process sensitivity is $\varphi_s = 0.95$ and the process specificity is $\varphi_e = 0.995$. In this case, studies 4, 7, 9 and 11 to 13 can not be used since the estimate is less than $1 - \varphi_e = 0.005$. Using the remaining studies, an overall estimate of the prevalence rate is, for a tolerance $\varepsilon = 10^{-6}$ and $t = 10^{-4}$, 0.2837%, (note that study 8, with the biggest sample size, provides a prevalence rate estimate of only 0.07548). If the process sensitivity and the process specificity of the study 6 is reduced to $\varphi_s = \varphi_s = 0.9$ then the prevalence rate overall estimate reduces to only 0.0982%. This shows the importance of giving more weight to the estimates obtained from methodologies with lesser error rates.

The previous example shows the impact of the quality index in the global estimate. However, the example just considered one of the studies different from the others concerning the process sensitivity and the process specificity. To verify the differences, in general, between our estimator and an unweighted mean of the estimates some simulations were performed using the MatLab 6 software. Hence, to assess the effect of our methodology in the accuracy of the global estimate, a 10^4 replicas of meta-analysis application were simulated. In this simulation:

- the prevalence rate was generated by an uniform random variable with values on the interval $(0.0001, 0.3)$;

- the overall estimator defined in (19) was used in each simulated study with $\varepsilon = 10^{-4}$ and $t = 10^{-9}$;
- the number of studies K within each meta-analysis was generated by a discrete uniform random variable varying on the set $\{5, 6, \ldots, 15\}$, i.e., $K \frown \text{UniformDiscrete}\{5, \ldots, 15\}$;
- each study dimension N was dependent on the prevalence rate since a very low prevalence rate will require a larger sample size. Thus, for:

  – $p \in (0.01, 0.3)$ N: was generated by a discrete uniform random variable $N \frown \text{UniformDiscrete}\,(5, \ldots, 1000)$;

  – $p \in (0.001, 0.01)$ N: was generated by a discrete uniform random variable $N \frown \text{UniformDiscrete}\,(50, \ldots, 10000)$;

  – $p \in (0.0001, 0.001)$ N: was generated by a discrete uniform random variable $N \frown \text{UniformDiscrete}\,(500, \ldots, 100000)$;
- each study process sensitivity and specificity was generated by distinct random variables with uniform distribution on the interval $(0.80, 1)$.

The following table summarizes the simulation results. For each estimator (unweighted and weighted mean) it is given the mean, median, percentiles 5 and 95 and the standard deviation of the bias absolute value.

Table 3. Meta-analysis simulation with 10^5 replicas.

		Estimator	
		unweighted mean	weighted mean
Bias absolute value	mean	0.001863	0.001279
	median	0.000800	0.000802
	P_5	0.000106	0.000072
	P_{95}	0.005537	0.003435
	Std. deviation	0.002978	0.002200

In the simulation, the choice of the unweighted mean of the estimates or of the weighted mean without a quality index led to similar results. When a quality index is used, the bias reduces over 20% in average. This shows that although when using the quality index an initial estimate for the prevalence rate is required, this is not important for the convergence of our method. For instance, our initial value was 0.5 although we only consider prevalence rates below 0.1. An alternative to this procedure may be considered if there is any reliable information about

variance of the estimators used in each study. In that case the weights w_i's may be rewritten as (6) or (11) wether a fixed effects model or a random effects model is being considered.

5. Conclusion

Since there are several processes for estimating the prevalence rate of a disease (or its mean value), it is necessary to consider the sensitivity and specificity of the process used for finding each estimate. The use of variance stabilizing transformations avoids the within study variance estimation, thus reducing the possible sources of error. The analytical expressions for those quantities could be implemented in the future in a statistical package in order to favor the extension of meta-analysis techniques to the problems involving different group testing procedures.

Acknowledgement

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] F. Anscombe, *The transformation of poisson, binomial and negative-binomial data*, Biometrika **35** (1948) 246–254.

[2] T. Berger, J.W. Mandell and P. Subrahmanya, *Maximally efficient two-stage screening*, Biometrics **56** (2000) 833–840.

[3] B.J. Biggerstaff and R.L. Tweedie, *Incorporating Variability in Estimates of Heterogeneity in the Random Effects Model in Meta-analysis*, Stat. Med. **16** (1997) 753–768.

[4] C.R. Bilder, B. Zang, F. Schaarschmidt and J.M. Tebbs, *binGroup: a package for group testing*, The R Journal **2** (2010) 56–60.

[5] C.L. Chen and W.H. Swallow, *Using group testing to estimate a proportion, and to test the binomial model*, Biometrics **46** (1990) 1035–1046.

[6] W.G. Cochran, *The combination of estimates from different experiments*, Biometrics **10** (1954) 101–129.

[7] R. DerSimonian and N. Laird, *Meta-analysis in clinical trials*, Control. Clin. Trials **7** (1986) 177–178.

[8] R. Dorfman, *The detection of defective members in large populations*, Ann. Math. Statistics **14** (1943) 436–440.

[9] H.M. Finuncan, *The blood testing problem*, Appl. Stat. **13** (1964) 43–50.

[10] F.C. Garner, M.A. Stapanian, E.A. Yfantis and L.R. Williams, *Probability Estimation With Sample Compositing Techniques*, Journal of Official Statistics **5** (1989) 365–374.

[11] J. Hartung, G. Knapp and B.K. Sinha, Statistical Meta-Analysis with Applications (John Wiley & Sons, Hoboken, 2008).

[12] P. Holland, *Covariance stabilizing transformations*, Ann. Stat. **14** (1973) 84–92.

[13] J.M. Hughes-Oliver, *Pooling experiments for blood screening and drug discovery*, in: Screening — Methods for Experimentation in Industry, Drug Discovery, and Genetics, Dean and Lewis (Ed(s)), (New York: Springer, 2006) 48–68.

[14] M. Hung and W.H. Swallow, *Robustness of Group Testing in the Estimation of Proportions*, Biometrics **55** (1999) 231–237.

[15] F.K. Hwang, *Group testing with a dilution effect*, Biometrika **63** (1976) 671–673.

[16] N. Johnson, S. Kotz and X. Wu, Inspection Errors for Attributes in Quality Control (Chapman and Hall Ltd., NewYork, 1991).

[17] N. Johnson, S. Kotz and N. Balakrishnan, Continuous Univariate Distributions, Vol. 2 (John Wiley & Sons, NewYork, 1993).

[18] H. Kim, M. Hudgens, J. Dreyfuss, D. Westreich and C. Pilcher, *Comparison of group testing algorithms for case identification in the presence of testing errors*, Biometrics **63** (2007) 1152–1163.

[19] E. Kulinskaya, S. Morgenthaler and R.G. Staudte, Meta Analysis: a guide to calibrating and combining statistical evidence (Wiley, Chichester, 2008).

[20] V.A. Lancaster and S. Keller-McNulty, *A Review of Composite Sampling Methods*, JASA **93** (1998) 1216–1230.

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

[22] M.W. Loyer, *Bad probability, good statistics, and group testing for binomial estimation*, Am. Stat. **37** (1983) 57–59.

[23] A.L. Rukhin, B.J. Biggerstaff and M.G. Vangel, *Restricted maximum likelihood estimation of a common mean and the Mandel-Paul algorithm*, J. Stat. Plan. Infer. **83** (2000) 319–330.

[24] R. Santos, D. Pestana and J.P. Martins, *Extensions of Dorfman's theory*, in: Selected Papers of SPE 2010, Portuguese Statistical Society (Ed(s)), (New York: Springer, 2012) in print.

[25] K.M. Sobel and R.M. Elashoff, *Group testing with a new goal, estimation*, Biometrika **62** (1975) 181–193.

[26] A. Sterret, *On the detection of defective members of large populations*, Ann. Math. Statistics **28** (1957) 1033–1036.

[27] L.M. Wein and S.A. Zenios, *Pooled testing for HIV screening: capturing the dilution effect*, Oper. Res. **44** (1996) 543–569.

[28] M. Xie, K. Tatsuoka, J. Sacks and S.S. Young, *Group testing with blockers and synergism*, JASA **96** (2001) 92–102.

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

Received 16 March 2013
Revised 12 October 2013