Author/s:
Hepworth, G;Biggerstaff, BJ

Title:
Bias Correction in Estimating Proportions by Pooled Testing

Date:
2017-12-01

Citation:
Hepworth, G. & Biggerstaff, B. J. (2017). Bias Correction in Estimating Proportions by Pooled Testing. Journal of Agricultural Biological and Environmental Statistics, 22 (4), pp.602-614. https://doi.org/10.1007/s13253-017-0297-2.

Persistent Link:
https://hdl.handle.net/11343/283239

Bias Correction in Estimating Proportions by Pooled Testing

Graham Hepworth[1] and Brad J. Biggerstaff[2]

[1] School of Mathematics & Statistics, The University of Melbourne, Victoria 3010, Australia

[2] Centers for Disease Control and Prevention, Fort Collins, Colorado 80526, U.S.A.

Summary. In the estimation of proportions by pooled testing, the MLE is biased, and several methods of correcting the bias have been presented in previous studies. We propose a new estimator based on the bias correction method introduced by Firth (1993), which uses a modification of the score function, and we provide an easily computable, Newton-Raphson iterative formula for its computation. Our proposed estimator is almost unbiased across a range of problems, and superior to existing methods. We show that for equal pool sizes the new estimator is equivalent to the estimator proposed by Burrows (1987). The performance of our estimator is examined using pooled testing problems encountered in plant disease assessment and prevalence estimation of mosquito-borne viruses.

Key Words: Bias correction; Estimation of proportions; Group testing; Pooled testing; Virus prevalence.

Bias Correction in Estimating Proportions by Pooled Testing

Summary. In the estimation of proportions by pooled testing, the MLE is biased, and several methods of correcting the bias have been presented in previous studies. We propose a new estimator based on the bias correction method introduced by Firth (1993), which uses a modification of the score function, and we provide an easily computable, Newton-Raphson iterative formula for its computation. Our proposed estimator is almost unbiased across a range of problems, and superior to existing methods. We show that for equal pool sizes the new estimator is equivalent to the estimator proposed by Burrows (1987). The performance of our estimator is examined using pooled testing problems encountered in plant disease assessment and prevalence estimation of mosquito-borne viruses.

Key Words: Bias correction; Estimation of proportions; Group testing; Pooled testing; Virus prevalence.

1 Introduction

Pooled testing (also known as group testing) occurs when individuals from a population are pooled together and tested as a group for the presence of an attribute, usually a pathogen. Since its introduction to the statistical literature by Dorfman (1943), pooled testing has been applied to a diverse array of fields, including seed health assays (Liu et al., 2011), HIV prevalence estimation (Hund and Pagano, 2013), screening of mycobacteria in rodents (Durnez et al., 2008) and detection of genetically modified organisms (Yamamura and Hino, 2007). Two common areas of application, which correspond to the current authors' involvement, are plant disease assessment (e.g. Freeman et al., 2013) and prevalence of viruses in mosquitoes which transmit them (e.g. Aranda et al., 2009).

Research in pooled testing diverges into two areas — classification, in which the purpose is to identify the positive units, and estimation, where the aim is to estimate the proportion of positives (p) in a population. The statistical issues are quite different, and the literature generally focuses on one or the other. Our concern is with estimation, because of the subject areas which have led to our interest in pooled testing. In plant disease assessment, estimating disease levels in a population such as a field crop, a glasshouse, or a plant production process, has far higher priority than identifying infected plants. With mosquito-borne diseases, the aim is often to estimate the prevalence of virus infection in mosquitoes, which is used in a variety of ways in public health to aid control and prevention of human disease.

The maximum likelihood estimator (MLE) of p is (for a fixed sample size) positively biased, except in the trivial case of all pools consisting of one individual each. This bias, which can be severe, was recognized as an issue in early work on pooled testing (Thompson, 1962), but point estimation has not generated as much interest as interval estimation. Some studies have simply encouraged better design of pooled testing procedures, with s-

tatements such as "the choice of experimental design is crucial to achieve negligible bias" (Schaarschmidt, 2007). It is indeed the case that good planning can alleviate the problem to some extent. For example, Swallow (1985) produced tables showing the bias of the MLE for different numbers of pools and different pool sizes, and these have been used extensively by other studies, even recently (Ding and Xiong, 2016). Hepworth and Watson (2009) showed that bias can be reduced by sequential testing with pools of decreasing size. But in many situations, a design (such as a testing plan with particular numbers of pools of a certain size) does not exist, or if it does, the process may not be able to be fully controlled. For example, the US Centers for Disease Control and Prevention may collect virus vector mosquitoes related to ongoing disease outbreaks or when conducting field-related experiments, and these samples are usually analysed by trap location and collection date, so that pool sizes may not be pre-determined, and pooling of individuals rarely results in equal-sized pools. An example of such a study is described by Godsey et al. (2002).

It is important that less biased alternatives to the MLE be available for the wide range of pooled testing scenarios that occur in practice. In this paper we describe estimators which are almost unbiased, and also have smaller mean squared error than the MLE. We first consider work that has already been done in this area, and then propose a new estimator based on the bias reduction method introduced by Firth (1993). We provide an easily computable, Newton-Raphson iterative formula for its computation. We show that Firth's method is equivalent to that introduced by Burrows (1987) for pools of equal size. We then compare the new estimator with the bias-adjusted MLE resulting from the method described by Gart (1991), which has been previously shown to work well for pools of unequal size. These estimators are evaluated for a variety of pooled testing problems, chosen to reflect real situations encountered by the authors, in either plant disease assessment or mosquito-borne viruses.

2 Bias Correction of the MLE in Pooled Testing

Suppose that for $i = 1, \ldots, d$, n_i pools of size m_i are tested, of which $X_i = x_i$ pools are positive. Assuming that the individuals in the pools follow i.i.d. Bernoulli distributions with parameter p, and that testing is conducted without error, the binomial parameters for the distribution of X_i are n_i and $1 - (1-p)^{m_i}$, n_i assumed fixed and known. The log-likelihood is therefore

$$l(p\,;\,\boldsymbol{x}) = \sum_{i=1}^{d} \log \binom{n_i}{x_i} + \sum_{i=1}^{d} x_i \log[1-(1-p)^{m_i}] + \log(1-p) \sum_{i=1}^{d} m_i(n_i - x_i) \qquad (1)$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$. The MLE of p is the solution $\hat{p}$ to the score equation

$$S(p) = \frac{1}{1-p} \sum_{i=1}^{d} \left[\frac{m_i x_i}{1-(1-p)^{m_i}} - m_i n_i \right] = 0 \qquad (2)$$

which requires iteration except when all pools are of equal size m, in which case the MLE simplifies to $\hat{p} = 1 - (1 - x/n)^{1/m}$. A convenient, iterative formula for this solution is given in Walter et al. (1980).

Burrows (1987) derived a bias correction for equal pool sizes by writing the MLE as $1 - (y/n)^{1/m}$, where $y = n - x$ is the number of negative pools. Beginning with $[(y+a)/(n+b)]^{1/m}$ instead of the MLE, he found that the bias term of $O(n^{-1})$ is eliminated when $a = b = \frac{1}{2}(m-1)/m$. This results in an estimator which can be expressed as either of the following:

$$\tilde{p} = 1 - \left[\frac{n - x + a}{n + a} \right]^{1/m} = 1 - \left[\frac{2m(n-x) + m - 1}{2mn + m - 1} \right]^{1/m}. \qquad (3)$$

This estimator could be reasonably described as "the MLE with about a half an additional negative pool." In calculations across a range of n, p and m, Burrows found the bias of $\tilde{p}$ to be less than about 5% of that of $\hat{p}$, and the mean squared error (MSE) to be uniformly less than that of $\hat{p}$.

Hepworth and Watson (2009) found the Burrows correction to be extremely effective in reducing bias for small p, and recommended $\tilde{p}$ as an estimator when pool sizes are equal, describing it as "essentially unbiased for values of p consistent with the design of the procedure." Hund and Pagano (2013) concurred, stating that "the Burrows estimator has negligible bias, even for sample sizes as small as 100." Colon et al. (2001) pointed out that the Burrows estimator belongs to a class of shrinkage estimators which shrink the MLE toward zero. They examined in some detail the expansions used to derive it, and compared its bias and MSE properties with the MLE, the jackknife, and other shrinkage estimators. For small sample sizes ($n < 20$) they found the Burrows estimator to have the least bias, and for larger n, it tied for best with the jackknife. Asymptotically, they found that the Burrows estimator tends to overestimate p and the jackknife to underestimate it. However, "for large enough sample sizes the bias of the Burrows estimator is essentially zero" (Mitchell and Pagano, 2013).

For equal pool sizes therefore, the Burrows correction produces an estimator which, for all practical purposes, is effectively unbiased. This perhaps explains the lack of proposed alternatives, though there have been a few. One is the jackknife, mentioned above; another is an empirical Bayes estimator proposed by Bilder and Tebbs (2005), which gave much smaller bias and MSE than $\hat{p}$, but not quite as small as $\tilde{p}$. Gildow et al. (2008) applied this estimator to the transmission of Cucumber mosaic virus by aphids in snap bean.

The only serious competitor to the Burrows estimator is the general bias correction described by Gart (1991) for independent X_i and a single parameter, which matches the usual pooled testing framework we study here. Except for terms of $O(n_i^{-2})$, the bias of the MLE is

$$b(p) = -\frac{2\frac{\partial I}{\partial p} + \mathbb{E}\left[\frac{\partial^3 l}{\partial p^3}\right]}{2[I(p)]^2} \tag{4}$$

where I is the Fisher information. Using the fact that $\mathbb{E}(X_i) = n_i(1-(1-p)^{m_i})$, it can be

shown that

$$I(p) = \sum_{i=1}^{d} \frac{m_i^2 n_i (1-p)^{m_i - 2}}{1 - (1-p)^{m_i}} \tag{5}$$

(Walter et al., 1980). The other terms in (4) were derived for pooled testing by Hepworth (2005), and lead to the following expression for the bias:

$$b(p) = \frac{1}{2[I(p)]^2} \sum_{i=1}^{d} \frac{m_i^2 (m_i - 1) n_i (1-p)^{m_i - 3}}{1 - (1-p)^{m_i}}$$

which provides an estimator

$$\breve{p} = \hat{p} - b(\hat{p})$$

with the bias of $O(n_i^{-1})$ removed.

For equal pool sizes, Hepworth and Watson (2009) found Gart's bias correction to be effective in reducing bias for small p, though not quite as good as Burrows' correction. A situation in which either correction could have been usefully employed was in the testing of blood samples for HIV described by Brookmeyer (1999). Seven hundred individual samples were grouped into 7 pools of 100, and 4 of the pools tested positive, resulting in an MLE of $\hat{p} = 0.00844$. At $p = \hat{p}$, the MLE has a (relative) bias of 243%, the Burrows estimator has a bias of 0.43%, and the Gart correction gives a bias of -1.81%.

The Gart correction has the disadvantage of not providing a result when $\hat{p} = 1$ due to $I(p)$ being zero. For an extensive discussion of the problem of all positive pools in estimation of p, see Hepworth and Watson (2009).

For unequal pool sizes, the Gart correction was also effective. However, their main evaluation involved only one pooled testing procedure, comprising 8 pools of 20 and 8 pools of 5. The context for that particular evaluation was the estimation of virus prevalence in a carnation population, from which 200 plants were sampled, and tested in pools using ELISA. In the present study we evaluate a range of pooled testing scenarios, including larger ones typical of the monitoring of mosquito-borne viruses. Hepworth and Watson (2009) also assessed Gart's correction with sequential pooled testing procedures, and

found that it effectively accounted for the positive bias, but created a negative bias of similar magnitude. Sequential procedures have their own distinctive characteristics, and we do not consider them further here.

Because of the effectiveness of the Burrows correction for equal pool sizes, an extension of it to unequal pool sizes is desirable. No extension or generalization has yet been derived, however, as even obtaining $a = b = \frac{1}{2}(m-1)/m$ for equal pool sizes, with a closed-form expression for the estimator, is not trivial, and the unequal pool case appears intractable.

Hepworth and Watson (2009) made an attempt at a generalization by defining $y_i = n_i - x_i$ as the number of negative pools for $i = 1, \ldots, d$, replacing y_i in (2) by $y_i + a_i$ and n_i by $n_i + b_i$, where $a_i = b_i = \frac{1}{2}(m_i - 1)/m_i$, and iteratively solving (2). The result was an over-correction, with negative bias for all p, though it was less in absolute value than the positive bias of $\hat{p}$. They also pointed out that there is no inherent reason why a_i must equal b_i, or why either quantity has to equal $\frac{1}{2}(m_i - 1)/m_i$, and so a direct generalization remains elusive.

3 Firth's Bias Correction of the MLE

Firth (1993) proposed a general bias correction to the MLE, which involves a modification to the estimating function. Instead of solving the usual score equation $S(p) = 0$, a solution is found to

$$S(p) - I(p)b(p) = 0\,. \tag{6}$$

This removes bias of $O(n_i^{-1})$, and has the feature of being 'preventative' rather than 'corrective,' as Gart's correction and jackknife methods are. This is an advantage when an estimate can be infinite, as has been seen in logistic regression (Heinze and Schemper, 2002). Firth's method has been applied widely; for example, with a single binomial observation, $\frac{1}{2}$ is added to the number of positives and the number of negatives, which is

not dissimilar from the Burrows correction for equal pool sizes. For pooled testing, we have derived expressions for $S(p)$, $I(p)$ and $b(p)$, and so the numerical solutions to (6) are quite straightforward, though iterative. The resulting estimator, which is our proposed new estimator of p, therefore arises from applying Firth's bias correction method to pooled testing.

To solve equation (6), the Newton-Raphson method provides a convenient iterative scheme for implementation. For each distinct pool size $i = 1, \ldots, d$, write the unique pool size contribution to the information as

$$v_i(p) = \frac{m_i^2 n_i (1-p)^{m_i - 2}}{1 - (1-p)^{m_i}}$$

and set $w_i(p) = v_i(p)/I(p) = v_i(p)/\sum_{j=1}^{d} v_j(p)$. Note that $\sum_{i=1}^{d} w_i = 1$. Next compute

$$\frac{\mathrm{d}v_i(p)}{\mathrm{d}p} = v_i'(p) \;=\; \frac{m_i^2 n_i (1-p)^{m_i - 3}}{\left[1 - (1-p)^{m_i}\right]^2} \left\{2\left[1 - (1-p)^{m_i}\right] - m_i\right\}$$

and

$$\frac{\mathrm{d}w_i(p)}{\mathrm{d}p} = w_i'(p) \;=\; \frac{v_i'(p)\sum_{j=1}^{d} v_j(p) - v_i(p)\sum_{j=1}^{d} v_j'(p)}{\left[\sum_{j=1}^{d} v_j(p)\right]^2}.$$

Using expressions given, the Newton-Raphson recursion may therefore be derived as

$$p_{k+1} \;=\; p_k + \frac{\sum_{i=1}^{d}\left[\frac{m_i x_i}{1-(1-p_k)^{m_i}} - \frac{1}{2} m_i w_i(p_k)\right] - \left(N - \frac{1}{2}\right)}{\sum_{i=1}^{d}\left\{\frac{m_i^2 x_i (1-p_k)^{m_i - 1}}{\left[1-(1-p_k)^{m_i}\right]^2} + \frac{1}{2} m_i w_i'(p_k)\right\}}$$

with $N = \sum_{i=1}^{d} m_i n_i$ equal to the number of individuals. A starting value at $k = 0$ should be pre-determined, and we have found it convenient to use the MLE computed using the proportion of positives pools and average pool size, as follows:

$$p_0 \;=\; 1 - \left(1 - \sum_{i=1}^{d} x_i \Big/ \sum_{i=1}^{d} n_i\right)^{\sum_i n_i / N}.$$

Iteration proceeds until the change in successive iterates p_k and p_{k+1} is less than some desired tolerance. For the case of all positive pools, the starting value for the iteration should be $p_0 = \sum_{i=1}^{d} x_i/N$, the so-called minimum infection rate (MIR), as the preceeding one is 0 in this case, making the Newton-Raphson update undefined. As an alternative to this iterative scheme, commonly used statistical software often provides root-finding functionality.

For equal pool sizes, expressions $S(p)$ and $I(p)$ are (2) and (5), respectively, without the summations or subscripts, and the bias in equation (4) simplifies to

$$b(p) = \frac{(m-1)(1-(1-p)^m)}{2m^2n(1-p)^{m-1}}\,.$$

Piecing these together, equation (6) then becomes

$$S(p) - I(p)b(p) = \frac{1}{1-p}\left(\frac{mx}{1-(1-p)^m} - mn\right) - \frac{m-1}{2(1-p)} = 0$$

which rearranges to

$$(1-p)^m = \frac{2m(n-x)+m-1}{2mn+m-1}$$

whose solution for p is equivalent to the Burrows estimator given in (3). This shows that the Firth bias correction is equivalent to the Burrows bias correction for equal pool sizes, and so may be considered a natural extension of the Burrows estimator for unequal pool sizes.

4 Comparison of Bias-corrected Estimators

4.1 Small-sized Pooled Testing Example

We now compare Gart's and Firth's bias correction methods in some detail using the pooled testing example described above. There were 200 individuals grouped into 8 pools of 20 and 8 pools of 5, for which we adopt the notation $N : m^n = 200 : 5^8\ 20^8$.

It is useful to test the methods on a small example such as this, because we cannot rely on asymptotic properties to rescue them from poor performance. It is instructive first to examine the estimates themselves for a range of outcomes; these are presented in Table 1, with the outcomes selected to give a range of values of $\hat{p}$. It is evident that Firth's method makes a slightly smaller correction than Gart's method.

Insert Table 1 here

For evaluation of the bias, there is not much to be gained by examining the entire range of p. A better approach, adopted by Hepworth and Watson (2009), is to concentrate on values of p consistent with the design of the procedure; i.e. values for which the probability of all positive groups is small, because this outcome is highly uninformative, and to be avoided if at all possible. We therefore let ψ be the value of p at which the probability of all positive groups is 0.05, and use ψ as an upper bound on p for purposes of evaluation. For this example, $\psi = 0.211$. Gart's method does not produce an estimate for all positive groups, so it is necessary to allocate a value in an ad-hoc way; we will use the Firth estimate. For small p the actual value chosen makes little difference, because the probability of all positives is negligible.

Table 2 gives the expected value of the estimators corrected by either Gart's or Firth's method, together with the percentage bias and root mean squared error (RMSE), for selected values of p. The corresponding figures for the MLE are also shown. Both corrected

estimators are close to unbiased, with less than 1% absolute bias for $p < \psi$. The bias is slightly negative for the Gart estimator, and mostly slightly positive for the Firth estimator. The RMSE is virtually identical for the two estimators, and here it essentially equals the standard deviation, due to the extremely small bias. Both estimators have smaller RMSE than the MLE, especially for larger p.

Insert Table 2 here

4.2 Medium-sized Pooled Testing Example

We now consider a "medium-sized" example, representative of some procedures used by the CDC in assessing virus infection rates in mosquitoes. This example has 5 pools of 5, 5 pools of 10, 5 pools of 25 and 6 pools of 50, which we write $N : m^n = 500 : 5^5\ 10^5\ 25^5\ 50^6$. Figure 1 shows the bias and RMSE of both estimators, for $p < \psi = 0.183$. For p less than about 0.05, the bias is extremely small for either, with the negative bias for the Gart estimator of about the same magnitude as the positive bias for the Firth estimator. For larger p, both estimators have negative bias, with Gart more negative. For this example, the Firth estimator is better overall, though the Gart estimator still has small bias, with the maximum absolute bias being 1.46% at $p = \psi$. The RMSE is virtually identical for the two methods, and less than the RMSE using the MLE (0.019 at $p = 0.05$, 0.193 at $p = \psi$).

Insert Figure 1 here

Another medium-sized example is the problem described by Brookmeyer (1999), which we introduced in section 2; this can be written $N : m^n = 700 : 100^7$. We do not provide details of the bias here, but overall Firth's estimator performs better than Gart's. At $p = \psi = 0.0105$, the bias is 0.17% for Firth and -2.39% for Gart.

4.3 Large-sized Pooled Testing Examples

We now consider a range of larger examples, with $N = 500$, 1000 or 5000, and between 1 and 4 different pool sizes. Table 3 compares the Gart and Firth estimators for mean absolute percentage bias and RMSE, calculated over 100 equally spaced points in the interval $[0.001, \psi]$. Also shown is the bias at $p = \psi$, where its maximum absolute value usually occurs over the range $(0, \psi)$. Corresponding values for the MLE are not shown, as they always vastly exceed the values for the Gart and Firth estimators. Figure 2 plots the bias of both estimators for six of the procedures listed in Table 3, selected to show a range of $N : m^n$ and the resulting bias patterns. One of the plots arises from equal pool sizes, two of them from 2 pool sizes, two from 3 pool sizes, and one from 4.

Insert Table 3 and Figure 2 here

Some trends emerge in these results. The most obvious and important is that, while the mean percentage (absolute) bias is small for both methods (generally $< 1\%$), it is always smaller for Firth's method than for Gart's. The difference between the methods generally increases with pool size.

The "worst" bias (i.e. at the highest prevalence consistent with the testing procedure) is always much smaller for Firth's method. It is always negative for Gart and usually positive for Firth. In percentage terms, the difference between the methods for worst bias decreases with increasing number of pools, and increases with average pool size.

The average RMSE is always very slightly larger for Firth's method, but the difference is of not practical consequence. For either method, it is generally worse (around 0.03) for procedures involving small pool sizes. However, this is still only about half the corresponding RMSE for the MLE.

5 Discussion

We have considered bias correction in estimation of proportions by pooled testing, in which the MLE is clearly unacceptably biased for routine applications. We have proposed a new estimator based on the general bias reduction method applied to MLEs described by Firth (1993), provided an easily computable formula for its iterative computation, and shown that it is better overall than the method described by Gart (1991). Firstly and most importantly, it results in less absolute bias in a wide range of pooled testing situations. But in addition, being based on a modification of the score function, it is preventative rather than corrective, thus avoiding problems relating to undefined parameter estimates. Finally, we have shown that for pools of equal size, Firth's method is equivalent to the method introduced by Burrows (1987), which is generally viewed as the best estimator available.

Firth's method has been applied to a range of estimation problems. One study of interest in the current context is that of Mehrabi and Matthews (1995), who applied it to estimating the most probable number (the MLE) in dilution assays. Hepworth (1996) pointed out the strong parallel between pooled testing and dilution assays, especially when pools are of unequal size. Mehrabi and Matthews recommended Firth's method, especially with the possibility of an infinite estimate. A close second was a "simple bias corrected estimator," which was an adjustment to the MLE based on Gart's method.

Ding and Xiong (2016) proposed a new estimator for the case of equal pool sizes based on a weighted combination of order statistics. They showed by simulation that it was almost unbiased in most cases they considered, and claimed it to be at least as good as the Burrows estimate. This therefore provides a comparison of their proposed estimator with the Firth-adjusted estimator we propose, since for equal pool sizes Burrows and Firth agree.

If MSE or RMSE (which is composed largely of variance here) were used as the main criterion for choosing an estimator, we might place the Gart and Firth methods on an equal footing. As Firth (1993) states, "the merits of bias reduction in any particular problem will depend on a number of factors." But given that the RMSE associated with the two methods is very similar, and much less than that of the MLE, the smaller bias of the Firth estimator gives it an advantage. This is likely to be particularly true for practitioners using pooled testing, whose disposition is often to prefer an unbiased estimator, even at the expense of a mild increase in variance.

We have assumed in this study that positive and negative bias are of equal detriment to an estimator. The fact that the corrected estimator has negative bias for Gart's method (i.e. it is a slight over-correction) and generally positive bias for Firth's method has therefore not been a consideration in recommending the Firth method.

Computation is not a major issue in deciding on an appropriate estimator. Estimation of p generally requires iteration when pools are of unequal size, even if there is no bias correction. In our simulations, some of the bias calculations took considerable computing time when there were a large number of outcomes, e.g. $N : m^n = 1000 : 5^{10}\, 10^{10}\, 25^{10}\, 50^{12}$, for which there are 17303 outcomes. If the number of different group sizes was large, this was accentuated. However, the computation of the estimates themselves took very little time, and so provided a practitioner has access to statistical software, bias-corrected estimates can be found readily. R code to implement the methods in this paper is available from the authors, and for Firth's estimator, R code implementing the Newton-Raphson iteration is provided in an online supplement accompanying the article at the journal's website.

References

Aranda, C., Sanchez-Seco, M. P., Caceres, F., Escosa, R., Galvez, J. C., Masia, M., Marques, E., Ruz, S., Alba, A., Busquets, N., Vazquez, A., Castella, J., and Tenorio, A. (2009). Detection and monitoring of mosquito flaviviruses in Spain between 2001 and 2005. (2009) *Vector-borne and Zoonotic Diseases* **9**, DOI: 10.1089/vbz.2008.0073.

Bilder, C. R. and Tebbs, J. M. (2005). Empirical Bayes estimation of the disease transmission probability in multiple-vector-transfer designs. *Biometrical Journal* **47**, 502–516.

Brookmeyer, R. (1999). Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence. *Biometrics* **55**, 608–612.

Burrows, P. M. (1987). Improved estimation of pathogen transmission rates by group testing. *Phytopathology*, **77**, 363–365.

Colon, S., Patil, G. P., and Taillie, C. (2001) Estimating prevalence using composites. *Environmental and Ecological Statistics*, **8**, 213–236.

Ding, J. and Xiong, W. (2016) A new estimator for a population proportion using group testing. *Communications in Statistics–Simulation and Computation* **45**, 10.1080/03610918.2013.854909.

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

Durnez, L., Eddyani, M., Mgode, G. F., Katakweba, A., Katholi, C. R., Machang'u, R. R., Kazwala, R. R., and Portaeis, F. and Leirs, H. (2008) First detection of mycobacteria in African rodents and insectivores, using stratified pool screening. *Applied and Environmental Microbiology* **74**, 768–773.

Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.

Freeman, A. J., Spackman, M. E., Aftab, M., McQueen, V., King, S., van Leur, J. A., Loh, M. H., and Rodoni, B. (2013) Comparison of tissue blot immunoassay and reverse transcription polymerase chain reaction assay for virus-testing pulse crops from a South-Eastern Australia survey. *Australasian Plant Pathology* **42**, 675–683.

Gart, J. J. (1991) An application of score methodology: Confidence intervals and tests of fit for one-hit curves. In *Handbook of Statistics*, C. R. Rao, R. Chakraborty (eds), **8**, 395–406. Amsterdam: Elsevier.

Gildow, F. E., Shah, D. A., Sackett, W. M., Butzler, T., Nault, B. A., and Fleischer, S. J. (2008) Transmission efficiency of *Cucumber mosaic virus* by aphids associated with virus epidemics in snap bean. *Phytopathology* **98**, 1233–1241.

Godsey, M. S., Nasci, R., Savage, H. M., Aspen, S., King, R., Powers, A. M., Burkhalter, K., Colton, L., Charnetzky, D., Lasater, S., Taylor, V., and Palmisano, C. T. (2005) West Nile Virus-infected mosquitoes, Louisiana, 2002. *Emerging Infectious Diseases* **11**, 1399–1404.

Heinze, G. and Schemper, M. (2002) A solution to the problem of separation in logistic regression. *Statistics in Medicine* **21**, 2409–2419.

Hepworth, G. (1996) Exact confidence intervals for proportions estimated by group testing. *Biometrics* **52**, 1134–1146.

Hepworth, G. (2005) Confidence intervals for proportions estimated by group testing with groups of unequal size. *JABES* **10**, 478–497.

Hepworth, G. and Watson, R. (2009) Debiased estimation of proportions in group testing. *JRSS-C* **58**, 105–121.

Hund, L. and Pagano, M. (2013) Estimating HIV prevalence from surveys with low individual consent rates: annealing individual and pooled samples. *Emerging Themes in Epidemiology* **10:2**.

Liu, S-C., Chiang, K-S., Lin, C-H., and Deng, T-C. (2011) Confidence interval procedures for proportions estimated by group testing with groups of unequal size adjusted for overdispersion. *Journal of Applied Statistics* **38**, 1467–1482.

Mehrabi, Y. and Matthews, J. N. (1995) Likelihood-based methods for bias reduction in limiting dilution assays. *Biometrics* **51**, 1543–1549.

Mitchell, S. and Pagano, M. (2012) Pooled testing for effective estimation of the prevalence of *Schistosoma mansoni*. *American Journal of Tropical Medicine and Hygiene* **87**, 850–861.

Schaarschmidt, F. (2007) Experimental design for one-sided confidence intervals or hypothesis tests in binomial group testing. *Communications in Biometry and Crop Science* **2**, 32–40.

Swallow, W. H. (1985) Group testing for estimating infection rates and probabilities of disease transmission. *Phytopathology* **75**, 882–889.

Thompson, K. H. (1962) Estimation of the proportion of vectors in a natural population of insects. *Biometrics* **18**, 568–578.

Walter, S. D., Hildreth, S. W. and Beaty, B.J. (1980) Estimation of infection rates in populations of organisms using pools of variable size. *American Journal of Epidemiology* **112**, 124–128.

Yamamura, K. and Hino, A. (2007) Estimation of the proportion of defective units by using group testing under the existence of a threshold of detection. *Communications in Statistics–Simulation and Computation* **36**, 949–957.

Table 1: Estimates of p from the Gart and Firth bias correction methods, for selected outcomes of a procedure testing 8 pools of 20 and 8 pools of 5

	Number of positive groups (x_1, x_2)									
Method	$(1,2)$	$(4,0)$	$(2,5)$	$(3,7)$	$(6,4)$	$(5,7)$	$(7,5)$	$(7,8)$	$(8,7)$	$(8,8)$
MLE	0.016	0.025	0.042	0.067	0.085	0.099	0.128	0.205	0.341	1
Gart	0.015	0.024	0.039	0.062	0.079	0.091	0.116	0.180	0.291	—
Firth	0.015	0.024	0.040	0.064	0.080	0.093	0.118	0.187	0.296	0.455

Table 2: Bias of estimators corrected by either Gart's or Firth's method, for testing 8 pools of 20 and 8 pools of 5

	MLE			Gart			Firth		
p	$\mathbb{E}(\hat{p})$	% bias	RMSE	$\mathbb{E}(\tilde{p})$	% bias	RMSE	$\mathbb{E}(\breve{p})$	% bias	RMSE
0.01	0.0105	4.6	0.0077	0.0100	−0.06	0.0074	0.0100	0.13	0.0074
0.02	0.0210	5.0	0.0115	0.0200	−0.07	0.0108	0.0200	0.15	0.0109
0.03	0.0317	5.5	0.0150	0.0300	−0.08	0.0139	0.0301	0.17	0.0139
0.04	0.0424	6.0	0.0184	0.0400	−0.09	0.0168	0.0401	0.19	0.0168
0.05	0.0533	6.6	0.0219	0.0499	−0.10	0.0196	0.0501	0.22	0.0197
0.07	0.0755	7.8	0.0297	0.0699	−0.14	0.0256	0.0702	0.25	0.0258
0.10	0.1099	9.9	0.0450	0.0998	−0.22	0.0357	0.1003	0.25	0.0359
0.15	0.1715	14.3	0.0927	0.1494	−0.41	0.0554	0.1501	0.09	0.0558
0.20	0.2448	22.4	0.1703	0.1983	−0.85	0.0752	0.1994	−0.30	0.0758
0.25	0.3363	34.5	0.2586	0.2459	−1.64	0.0910	0.2474	−1.05	0.0913
0.30	0.4447	48.2	0.3391	0.2910	−3.00	0.1003	0.2927	−2.42	0.1000

Table 3: Mean percentage bias, RMSE, and bias ($\times 10^4$) at $p = \psi$, for estimators corrected by either Gart's or Firth's method, for a range of pooled testing procedures.

			Gart			Firth		
N	m^n	ψ	mean |% bias|	mean RMSE	bias at $p = \psi$	mean |% bias|	mean RMSE	bias at $p = \psi$
500	5^{100}	0.506	0.020	0.0288	-9.81	0.015	0.0289	2.12
500	10^{50}	0.248	0.085	0.0220	-13.04	0.032	0.0222	0.57
500	20^{25}	0.103	0.224	0.0137	-9.80	0.066	0.0139	0.13
500	50^{10}	0.027	0.735	0.0063	-4.81	0.200	0.0065	0.22
500	100^{5}	0.008	2.069	0.0033	-2.46	0.553	0.0034	0.27
1000	20^{50}	0.133	0.122	0.0123	-9.46	0.033	0.0125	-0.10
1000	100^{10}	0.013	0.791	0.0033	-2.52	0.204	0.0034	0.10
1000	5^{200}	0.569	0.012	0.0231	-8.53	0.009	0.0232	1.75
5000	5^{1000}	0.687	0.004	0.0135	-6.17	0.003	0.0135	1.21
1000	$5^{100}\,50^{10}$	0.506	0.023	0.0286	-9.81	0.019	0.0287	2.12
1000	$25^{20}\,50^{10}$	0.078	0.251	0.0099	-8.61	0.067	0.0101	-0.20
5000	$5^{500}\,50^{50}$	0.641	0.006	0.0170	-7.09	0.005	0.0171	1.41
5000	$25^{100}\,50^{50}$	0.132	0.076	0.0083	-7.90	0.018	0.0084	-0.33
1000	$10^{20}\,25^{8}\,50^{12}$	0.180	0.203	0.0219	-15.18	0.024	0.0222	-1.12
1000	$10^{50}\,25^{12}\,50^{4}$	0.248	0.085	0.0207	-13.12	0.026	0.0209	0.23
1000	$5^{100}\,10^{40}\,25^{4}$	0.507	0.019	0.0262	-9.74	0.014	0.0263	1.98
5000	$10^{200}\,25^{60}\,50^{30}$	0.344	0.032	0.0155	-11.33	0.010	0.0156	0.18
1000	$5^{10}\,10^{10}\,25^{10}\,50^{12}$	0.261	0.199	0.0335	-19.35	0.037	0.0339	-0.46
1000	$5^{20}\,10^{40}\,25^{12}\,50^{4}$	0.350	0.065	0.0283	-14.37	0.043	0.0286	3.49
5000	$10^{50}\,25^{40}\,50^{30}\,100^{20}$	0.248	0.080	0.0187	-13.17	0.019	0.0188	-0.28

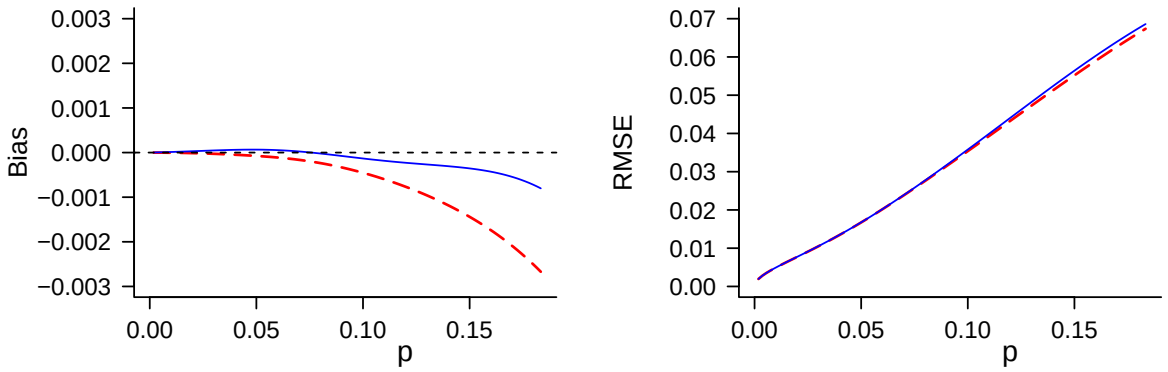

Figure 1: Bias and root mean squared error of estimators corrected by either Gart's or Firth's method, for pooled testing with $N : m^n = 500 : 5^5\ 10^5\ 25^5\ 50^6$. Gart = broken line, Firth = unbroken line.

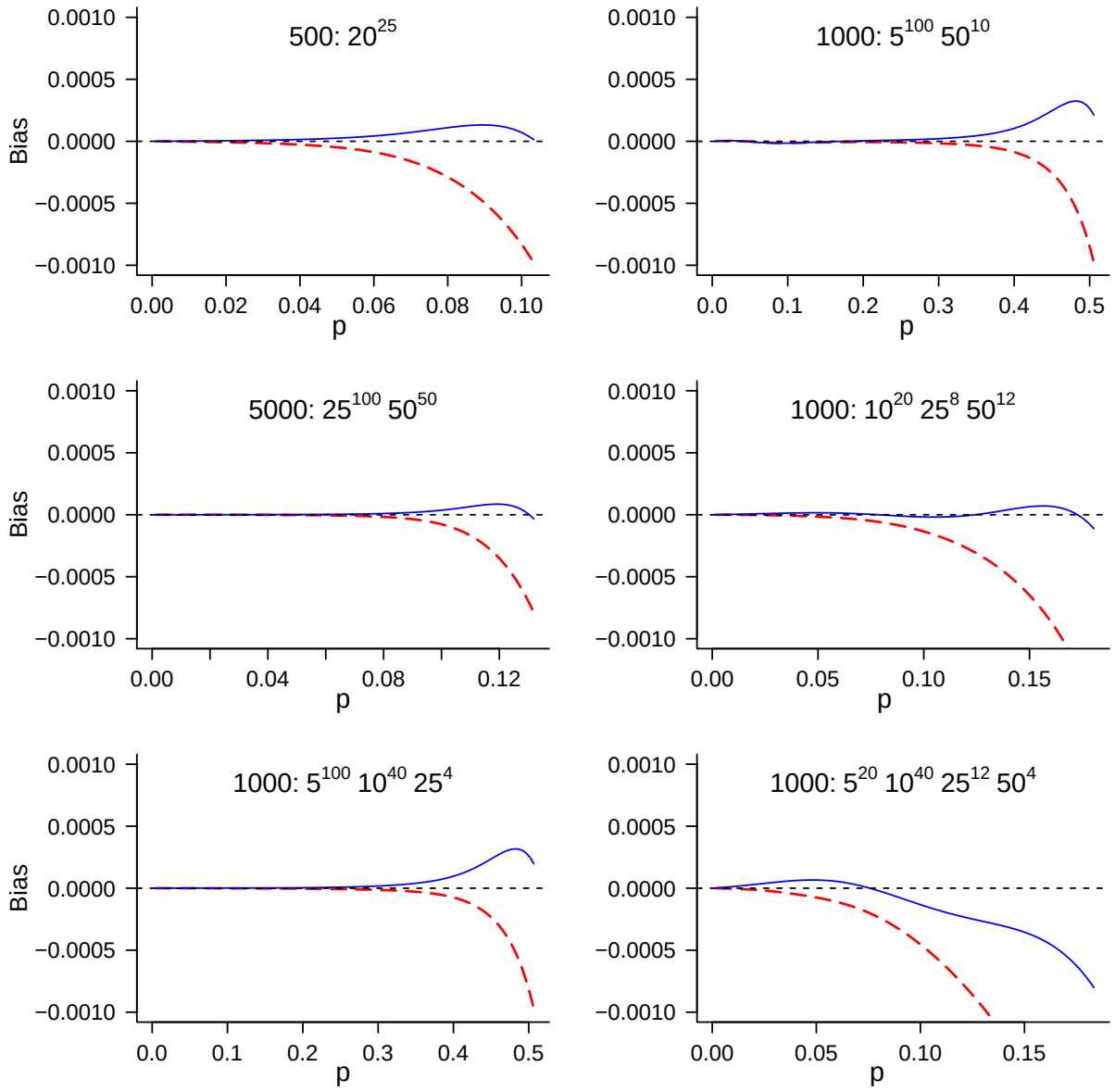

Figure 2: Bias of estimators over $p < \psi$ corrected by either Gart's or Firth's method, for a range of pooled testing procedures. Gart = broken line, Firth = unbroken line.

Bias Correction in Estimating Proportions by Pooled Testing: Supplement
by Hepworth and Biggerstaff, 2017

Derivation of the Newton-Raphson formula sequence for estimation of the Firth-corrected estimate of p in pooled testing, with accompanying R code.

Write the individual-pool contribution to the information as

$$v_i(p) = \frac{m_i^2 n_i (1-p)^{m_i-2}}{1-(1-p)^{m_i}}$$

and, taking all summations from 1 to d, set $w_i(p) = v_i(p)/I(p) = v_i(p)/\sum_j v_j(p)$, emphasizing the dependence on p. Note that $\sum_i w_i = 1$. Then with the score function $S(p)$, the Firth adjusted score, $S^*(p)$, is

$$\begin{aligned}
S^*(p) &= S(p) - I(p)b(p) \\
&= S(p) - \frac{1}{2I(p)} \sum_i \frac{m_i^2 (m_i-1) n_i (1-p)^{m_i-3}}{1-(1-p)^{m_i}} \\
&= S(p) - \frac{1}{2} \sum_i w_i(p)(m_i-1)(1-p)^{-1} \\
&= S(p) - \frac{1}{2(1-p)} \left[\sum_i m_i w_i(p) - 1 \right] \\
&= \frac{1}{1-p} \left\{ \sum_i \left[\frac{m_i x_i}{1-(1-p)^{m_i}} - \frac{1}{2} m_i w_i(p) \right] - \left(N - \frac{1}{2} \right) \right\}
\end{aligned}$$

In finding the root to $S^*(p)$, the leading term $1/(1-p)$ has no role, so we need to solve

$$\tilde{S}^*(p) = \sum_i \left[\frac{m_i x_i}{1-(1-p)^{m_i}} - \frac{1}{2} m_i w_i(p) \right] - \left(N - \frac{1}{2} \right) = 0$$

A couple of expressions for derivatives will be needed. Note that

$$\frac{\mathrm{d}v_i(p)}{\mathrm{d}p} = v_i'(p) = \frac{m_i^2 n_i (1-p)^{m_i-3}}{\left[1-(1-p)^{m_i}\right]^2} \left\{ 2\left[1-(1-p)^{m_i}\right] - m_i \right\}$$

and

$$\frac{\mathrm{d}w_i(p)}{\mathrm{d}p} = w_i'(p) = \frac{v_i'(p) \sum_j v_j(p) - v_i(p) \sum_j v_j'(p)}{\left[\sum_j v_j(p)\right]^2}$$

and so

$$\frac{\mathrm{d}\tilde{S}^*(p)}{\mathrm{d}p} = \tilde{S}^{*\prime}(p) = -\sum_i \frac{m_i^2 x_i (1-p)^{m_i-1}}{\left[1-(1-p)^{m_i}\right]^2} - \frac{1}{2} \sum_i m_i w_i'(p)$$

The Newton-Raphson recursion is therefore

$$p_{k+1} = p_k - \frac{\tilde{S}^*(p_k)}{\tilde{S}^{*\prime}(p_k)} = p_k + \frac{\sum_i \left[\frac{m_i x_i}{1-(1-p_k)^{m_i}} - \frac{1}{2} m_i w_i(p_k)\right] - \left(N - \frac{1}{2}\right)}{\sum_i \left\{\frac{m_i^2 x_i (1-p_k)^{m_i-1}}{[1-(1-p_k)^{m_i}]^2} + \frac{1}{2} m_i w_i'(p_k)\right\}}$$

with starting value at $k = 0$ pre-determined, using, for example, the MLE computed with the proportion of positives pools and average pool size

$$p_0 \;=\; 1 - \left(1 - \sum_i x_i \Big/ \sum_i n_i\right)^{\sum_i n_i/N}$$

or, especially when all pools are positive, the minimum infection rate (MIR), given by

$$p_0 \;=\; \frac{1}{N} \sum_i x_i$$

R code for this is

```
"pooledbinom.firth" <-
function(x, m, n = rep(1., length(x)), tol = 1e-008){
   # x is a vector of the numbers of positive pools of corresponding
   #   sizes given in m
   # m is a vector of pool sizes
   # n is a vector of numbers of pools of sizes m
   # tol is absolute tolerance desired
   if(length(m) == 1.) m <- rep(m, length(x))
   if(length(m) != length(x))
      stop("\n x and m must have same length if length(m) > 1")
   if(any(x > n)) stop("x elements must be <= n elements")
   if(all(x == 0.)) return(0.)
   N <- sum(n * m)
   # use proportion of positive pools, sum(x)/sum(n)
   # and average pool size, sum(n)/N
   p.new <- 1 - (1 - sum(x)/sum(n))^(sum(n)/N)
   if(sum(x) == sum(n)) p.new <- sum(x)/N

   done <- FALSE
   while(!done) {
      p.old <- p.new
      vi.p <- m^2 * n * (1-p.old)^(m-2) / (1 - (1-p.old)^m)
      wi.p <- vi.p / sum(vi.p)
      vi.p.prime <- m^2 * n * (1-p.old)^(m-3) *
               (2 * (1-(1-p.old)^m) - m) /
            (1-(1-p.old)^m)^2
```

```
      wi.p.prime <- (vi.p.prime * sum(vi.p) - vi.p *
              sum(vi.p.prime)) / sum(vi.p)^2
      p.new <- p.old +
        (sum(m*x/(1-(1-p.old)^m) - m*wi.p/2) - (N-1/2)) /
          sum(m^2*x*(1-p.old)^(m-1) / (1-(1-p.old)^m)^2 +
          m * wi.p.prime / 2)
      if(abs(p.new - p.old) < tol)
         done <- TRUE
   }
   p.new
}

#
# Examples
#
x <- c(1,0,0)
m <- c(5,25,50)
pooledbinom.firth(x, m)
# [1] 0.01035495

n <- c(100, 50, 200)
pooledbinom.firth(x,m,n)
# [1] 8.496025e-05
```