Author/s:
Hepworth, G

Title:
Bias correction of estimated proportions using inverse binomial group testing

Date:
2019-03-01

Citation:
Hepworth, G. (2019). Bias correction of estimated proportions using inverse binomial group testing. Australian and New Zealand Journal of Statistics, 61 (1), pp.51-60. https://doi.org/10.1111/anzs.12255.

 doi: 10.1111/j.1467-842X.XXX

Bias correction of estimated proportions using inverse binomial group testing

Graham Hepworth[1]

The University of Melbourne

Summary

Group testing, in which individuals are pooled together and tested as a group, can be combined with inverse sampling to estimate the prevalence of a disease. Alternatives to the MLE are desirable because of its severe bias. We propose an estimator based on the bias correction method of Firth (1993), which is almost unbiased across the range of prevalences consistent with the group testing design. For equal group sizes this estimator is shown to be equivalent to that derived by applying the correction method of Burrows (1987), and better than existing methods. For unequal group sizes the problem has some intractable elements, but under some circumstances our proposed estimator can be found, and we show it to be almost unbiased. Calculation of the bias requires computer-intensive approximation because of the infinite number of possible outcomes.

Key words: estimation of proportions; inverse sampling; negative binomial distribution; pooled testing

1. Introduction

Group testing (or pooled testing) describes the process of pooling individuals and testing them as a group for the presence of an attribute, usually a disease. The field has divided into two fairly distinct areas—*identification* (or classification) of infected individuals, and *estimation* of disease prevalence p. Estimation has received considerable attention in recent years, and is the area in which our interest lies. Some areas of application, such as transmission of viruses by insect vectors (e.g. Walter, Hildreth & Beaty (1980) or HIV prevalence (e.g. Kline et al. (1989)) have employed group testing for considerable time, but others, such as drug discovery (e.g. Hughes-Oliver (2006)), have adopted it more recently. Haber, Malinovsky & Albert (2018) provided an excellent summary of the development of estimation in group testing.

The most obvious benefit of group testing is in reducing the number of tests needed (and hence cost), or for the same number of tests, providing gains in efficiency. Group testing for estimation is most beneficial for small p, though with the appropriate group size

[1]The University of Melbourne

it also provides gains for moderate prevalence, and has the additional benefit of providing confidentiality where it is required Hammick & Gastwirth (1994).

In some situations, the design of a group testing procedure is constrained by the circumstances, such as the limitations of the laboratory equipment or the sensitivity of the test. As a result, and also to make the problem more straightforward, group testing research has generally assumed a fixed number of tests, with a resulting binomial model for the proportion of positive groups. However, sequential designs have also been proposed, and in some cases, implemented. Hughes-Oliver & Swallow (1994) proposed a two-stage adaptive procedure in which the choice of group size at the second stage depends on the maximum likelihood estimate (MLE) from the previous stage, using a criterion of minimising mean squared error (MSE). Hepworth (1996) described a sequential procedure applied to the estimation of virus infection levels in carnations, in which progressively smaller group sizes were used if the previous stage found all groups to be positive. Hardwick, Page & Stout (1998) suggested an adaptive procedure, where at each stage either an individual or a group of pre-chosen size is tested, depending on whether a point estimate exceeds a cut-off. Haber & Malinovsky (2017) proposed a set of random walk designs, in which the group size (potentially) changes at each step.

A particular class of sequential designs is that in which testing proceeds until a set number of positive tests is reached. Katholi & Unnasch (2006) first proposed the application of this inverse sampling arrangement to group testing, with the motivation of reporting estimates early in the testing process. They pointed out that in the case of equal group sizes, the total number of groups required follows a negative binomial distribution. Rodríguez-Pérez et al. (2006) used this process in testing an ocular disease vector (black flies) in groups of 50, until the first positive occurred. Pritchard & Tebbs (2011) used the term "inverse binomial group testing" in their examination of point and interval estimation for this design. Hepworth (2013) further developed their work, proposing an estimator which eliminated most of the bias, and interval estimators with favourable coverage properties. Montesinos-López et al. (2012) compared sample size procedures for estimating prevalence under the negative binomial model. Xiong (2016) considered how to optimise the group size, allowing for misclassification. Haber, Malinovsky & Albert (2018) showed that an unbiased estimator does not exist under this group testing arrangement.

Whether a binomial or negative binomial model is used for group testing, the MLE is biased, and several alternative estimators have been proposed for both situations. In this paper we consider bias correction under a negative binomial model. We show that an estimator based on the bias correction method introduced by Firth (1993) is even better than other correction methods, and has other advantages in addition to almost eliminating the bias. Because of the intractability of some aspects of inverse binomial group testing when groups are of different size, most work in this area has been limited to equal group sizes; we show that under certain testing arrangements, the bias correction can also be applied to unequal group sizes, and is very effective.

2. Bias correction for equal group sizes

2.1. Maximum likelihood estimation

We assume that the outcomes for individuals follow independent and identically distributed Bernoulli distributions with parameter p, and that they are randomly pooled into groups of size s. A random number $T = t$ tests are then performed until n positive groups are observed. We also assume that sensitivity and specificity are both 1, i.e., there is no misclassification. Although these assumptions have sometimes been relaxed in group testing studies which assume a binomial model, they have been adhered to in most work on inverse binomial group testing.

The probability of a positive group is $\pi = 1 - (1-p)^s$. The number of tests performed, T, has a negative binomial distribution with parameters n and π. The MLE of p is

$$\hat{p} = 1 - \left(1 - \frac{n}{T}\right)^{1/s}$$

and its bias is even greater than that arising from a binomial model, because the MLE of π, n/T, is already biased. Pritchard & Tebbs (2011) proposed three estimators of p, all of which shrink the proportion of positive groups towards 0, which in turn adjusts the MLE downwards. This reduces the bias and MSE, but the optimisation required to compute them depends on prior knowledge of p, and so their performance can vary widely (Haber, Malinovsky & Albert (2018)).

2.2. Gart's bias correction method

Gart (1991) described a general bias correction to the MLE for a single unknown parameter, which Hepworth (2013) applied to inverse binomial group testing. Gart showed that the bias (except for terms of $O(n^{-2})$) is

$$B(p) = -\frac{2\,dI/dp + \mathrm{E}(d^3l/dp^3)}{2\,I(p)^2} \tag{1}$$

where $l(p)$ is the log-likelihood and $I(p)$ is the Fisher information, which in this situation is

$$I(p) = \frac{ns^2(1-p)^{s-2}}{(1-(1-p)^s)^2} \tag{2}$$

(Pritchard & Tebbs (2011)). Hepworth (2013) showed that

$$\frac{dI}{dp} = -ns^2\left(\frac{(s-2)(1-p)^{s-3}+(s+2)(1-p)^{2s-3}}{(1-(1-p)^s)^3}\right) \tag{3}$$

and

$$\mathrm{E}\left(\frac{d^3l}{dp^3}\right) = \frac{sn}{(1-p)^3} \times \left(\frac{s(s+1)(1-p)^s(1-(1-p)^s)+2(s(1-p)^s+(1-p)^s-1)^2}{(1-(1-p)^s)^3} - \frac{2}{1-(1-p)^s}\right). \tag{4}$$

Substituting $\hat{p}$ into these expressions gives an estimate of the bias, which can be subtracted from $\hat{p}$ to give the bias-corrected estimate, whose estimator we denote $\hat{p}_G$. This estimator is undefined on the boundaries of the parameter space, which is of concern only at $p = 1$, since $\hat{p} = 0$ cannot arise ($T \geq n > 0$). The result $\hat{p} = 1$ (all positive groups, $T = n$) is highly uninformative, and should be unlikely in a well-designed study, but it still needs an ad-hoc solution; if this result occurs it suggests that the group size is too large for the underlying p. To address this issue when evaluating bias, Hepworth & Watson (2009) imposed an upper bound ψ on p, where ψ is the value of p at which the probability of all positive groups is 0.05, which for equal group sizes has the closed form $1-(1-(0.05)^{1/n})^{1/s}$. This restricts p to values consistent with the design of the group testing procedure, since the probability of all positive groups will then be small.

Hepworth (2013) evaluated the performance of $\hat{p}_G$ for $n = 1, 5, 10$ and $s = 5, 10, 20, 30, 50$. The Gart correction was found to be very effective in removing the bias, especially for small p. It resulted in an overcorrection of between about 1% and 2% for p close to ψ. The MSE of $\hat{p}_G$ was also examined, and found to be much less than that of $\hat{p}$.

2.3. Firth's bias correction method

Hepworth & Biggerstaff (2017) applied the general bias correction method introduced by Firth (1993) to fixed group testing with a binomial model. They found the resulting bias-corrected estimator to be almost unbiased across a range of problems, and less biased overall than the estimator arising from Gart's correction. They also showed that for equal group sizes, Firth's method is equivalent to the bias-correction method introduced by Burrows (1987), which removed nearly all the bias of the MLE and also greatly reduced the MSE. Firth's method does not find the MLE and then correct it, as Gart's method does. Rather, it is based on a modification to the score function, requiring the solution $\hat{p}_F$ to

$$S(p) - I(p)B(p) = 0\,. \tag{5}$$

It is therefore preventative rather than corrective, which has the advantage of avoiding undefined parameter estimates, such as those on the boundary. The score function for inverse binomial group testing is

$$S(p) = -\frac{s}{1-p}\left(t - \frac{n}{1-(1-p)^s}\right) \tag{6}$$

(Hepworth (2013)), and the other two quantities in (5) are already given in 2.2.

It is useful to firstly examine the estimates arising from the Firth correction. Consider $n = 5$ and $s = 20$, the same example used in Hepworth (2013) to illustrate the Gart bias correction. For this example, $\psi = 0.039$. Table 1 shows $\hat{p}$, $\hat{p}_F$ and $\hat{p}_G$ for the first ten outcomes (excluding all positive groups ($t = 5$)). Firth's method, although providing a substantial correction to the MLE, results in a slightly smaller correction than Gart's method, a feature noted by Hepworth & Biggerstaff (2017) for binomial group testing.

Table 1. Firth and Gart bias corrections applied to the MLE, $n = 5$ and $s = 20$.

t	6	7	8	9	10	11	12	13	14	15
$\hat{p}$	0.0857	0.0607	0.0479	0.0397	0.0341	0.0299	0.0266	0.0240	0.0218	0.0201
$\hat{p}_F$	0.0635	0.0470	0.0376	0.0314	0.0270	0.0238	0.0212	0.0191	0.0175	0.0161
$\hat{p}_G$	0.0600	0.0460	0.0372	0.0312	0.0269	0.0237	0.0212	0.0191	0.0174	0.0160

Haber, Malinovsky & Albert (2018) applied the idea of Burrows (1987) to this model, and derived the estimator

$$\hat{p}_B = 1 - \left(\frac{y+\nu}{y+n+\nu-1}\right)^{1/s}, \ \ \nu = \frac{s-1}{2s} \tag{7}$$

where y is the number of negative groups until the nth positive; in our notation, $t = y + n$. Using (1) to (4), (5) can be shown to simplify to

$$\frac{-2ns + (1-p)^s(-2st + s + 1) + 2st + s - 1}{2(1-p)\left((1-p)^s - 1\right)} = 0$$

which requires

$$(1-p)^s = \frac{2st - 2ns + s - 1}{2st - s - 1} = \frac{t - n + \nu}{t - 1 + \nu}$$

whose solution is given by (7). Hence the Firth bias correction is equivalent to the Burrows bias correction for inverse binomial group testing, as found for the binomial model.

3. Bias of corrected estimator

Calculation of bias is more complicated for sequential testing than for fixed testing, especially when there are an infinite number of possible outcomes, as is the case when they follow a negative binomial distribution. The exact bias of either of the corrected estimators is an infinite sum, with no analytical simplification. It is necessary to slightly approximate the bias, which we did by calculating

$$\sum_{t=n}^{t_*}\left(\hat{p}_F\binom{t-1}{n-1}\left(1-(1-p)^s\right)^n(1-p)^{s(t-n)}\right)-p \tag{8}$$

where $\Pr(T > t_*) < 10^{-6}$ for the smallest value of p considered, which we set to $\psi/250$. For p larger than this, $\Pr(T > t_*)$ is even smaller than 10^{-6}, because this function is decreasing in p; the approximation is therefore extremely close to the exact value.

We calculated the expected value, bias and percentage bias of $\hat{p}_F$ for the example above ($n = 5$, $s = 20$). The results are shown in Table 2 for selected values of p, along with those for $\hat{p}_G$ as a comparison.

Table 2. Expected value, bias and percentage bias for the estimator corrected for bias using the Firth correction or the Gart correction, for $n = 5$ and $s = 20$ ($\psi = 0.039$).

p	0.001	0.003	0.005	0.01	0.02	0.03	0.04	0.05	0.1
$\mathrm{E}(\hat{p}_F)$	0.0010	0.0030	0.0050	0.0100	0.0200	0.0301	0.0400	0.0497	0.0874
$\mathrm{E}(\hat{p}_G)$	0.0010	0.0030	0.0050	0.0100	0.0199	0.0297	0.0394	0.0487	0.0861
$\mathrm{Bias}(\hat{p}_F)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	−0.0003	−0.0126
$\mathrm{Bias}(\hat{p}_G)$	0.0000	0.0000	0.0000	−0.0000	−0.0001	−0.0003	−0.0006	−0.0013	−0.0139
% $\mathrm{Bias}(\hat{p}_F)$	1.12	0.09	0.02	0.06	0.24	0.34	0.11	−0.65	−12.6
% $\mathrm{Bias}(\hat{p}_G)$	0.01	0.01	0.01	−0.06	−0.38	−0.88	−1.59	−2.61	−13.9

Both correction methods are clearly very effective in removing the bias for small p, with Firth's method better overall. For $p < \psi = 0.039$, the mean absolute bias is 0.22% for Firth's method and 0.50% for Gart. For larger p, both methods over-correct, but the Firth correction less so.

Figure 1 (top three rows) displays the bias of $\hat{p}_F$ and $\hat{p}_G$ for $n = 5, 10 \times s = 5, 20, 50$ and $p \leq 1.1\psi$, a range of p broadly consistent with the group testing design. The vertical dashed line in each plot is at $p = \psi$. The plots show that Firth's correction has virtually eliminated the bias for $p < \psi$. The Gart correction has also been very effective, but is not quite as good as the Firth correction, except for the smallest problem ($n = 5, s = 5$).

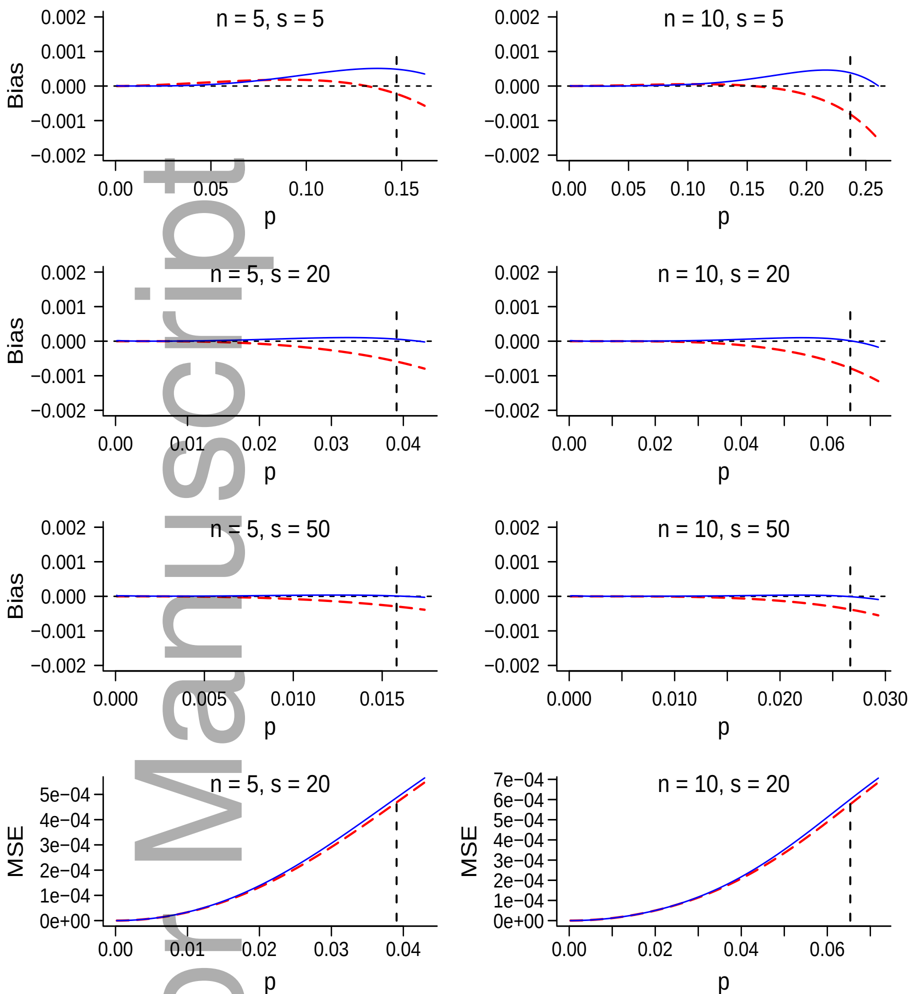


Figure 1. Bias and MSE of estimators corrected by either Firth's or Gart's method, for various combinations of n and s. Firth = continuous line, Gart = broken line. Vertical dashed line is at $p = \psi$.

We also examined the MSE of $\hat{p}_F$, using the same approximation method as for the bias, with $\hat{p}_F$ replaced by $\hat{p}_F^2$ in the summation shown in (8) to calculate the expected value of the square of the estimator, and hence the variance. For the same reasons described above, this method gives a very close approximation to the true value of the MSE.

Figure 1 (bottom row) shows the MSE of $\hat{p}_F$ and $\hat{p}_G$ for $n = 5$ or 10 and $s = 20$. Plots for the other combinations of n and s have the same shape and pattern, and are not shown. There is clearly very little difference in MSE between the two bias-correction methods. Note that the contribution of the bias (squared) to the MSE is very small, so the plots are essentially showing the variance. It is not surprising that for larger n, the MSE is smaller for the same

p, though it is larger if we consider p relative to ψ. We have not plotted either the bias or the MSE for $\hat{p}$, as it is many times greater than that of either of the bias-corrected estimators. Nor have we plotted the bias for larger n, as it is even smaller than for the values of n shown in Figure 1.

4. Bias correction for unequal group sizes

In some situations, equal group sizes is not a practical option. For example, in the transmission of viruses by mosquitoes, the natural group is a trap, in which the number of individuals (mosquitoes) can vary (Biggerstaff (2008)). In other situations, unequal group sizes are used intentionally because of considerable uncertainty about the prevalence. For example, the study of carnation viruses described by Hepworth (1996) used group sizes of 25, 5 and 1.

When groups are of unequal size, the distribution of T is not negative binomial, and $I(p)$ is intractable, thus inhibiting analytical solutions to finding bias. Pritchard & Tebbs (2011) derived the observed information, and used it to find an approximate confidence interval for p based on the score. They calculated coverage probabilities by simulation, using ranges of group sizes such as 10 to 30. But for point estimation they considered only equal group sizes.

One situation which results in a tractable information function is when each group size has a set number of positives at which testing for that group size stops. Suppose there are d group sizes $\boldsymbol{s} = (s_1, \ldots, s_d)$, and that for each $s_i, i = 1, \ldots, d$, a random number $T_i = t_i$ tests are performed until n_i positive groups are observed. The outcomes $\boldsymbol{T} = (T_1, \ldots, T_d)$ then follow a product of independent but non-identically distributed negative binomial distributions. All the quantities defined in (1) to (6) can be summed across the d group sizes.

This situation is arguably more realistic than one in which there is just an overall number of positives $n = \sum_{i=1}^{d} n_i$. For example, suppose the group sizes alternate between $s_1 = 5$ and $s_2 = 20$ with an overall $n = 10$, and it takes 20 tests to obtain 10 positives. The elements of $\boldsymbol{n} = (n_1, n_2)$ now constitute a random variable; for example, $(6, 4)$ and $(2, 8)$ give different impressions of prevalence, with $\hat{p} = 0.0520$ and 0.0657 respectively. It would likely be more sensible, and no more difficult, to set both n_1 and n_2, for example, to 3 and 7 respectively.

To test the bias correction methods, we combined the three group sizes that we tested for equal size groups, i.e. $\boldsymbol{s} = (5, 20, 50)$. To choose a sensible testing scheme, we assumed $p = 0.02$, a value not incompatible with any of the group sizes. This led to us using $\boldsymbol{n} = (3, 10, 20)$, which is roughly in proportion to the probability of a positive group $(0.096, 0.332, 0.636)$, so that the expected number of tests of each group size is similar. As with equal group sizes, it is useful to first examine the estimates themselves. Table 3 shows $\hat{p}$, $\hat{p}_F$ and $\hat{p}_G$ for a set of outcomes selected to give a range of values of $\hat{p}$ ($\psi = 0.122$).

As observed for equal group sizes, Firth's method results in a smaller correction than Gart's method, though for smaller p the difference is minimal.

Table 3. Firth and Gart bias corrections applied to the MLE for a range of outcomes, $n = (3, 10, 20)$ and $s = (5, 20, 50)$.

t	$(4, 11, 20)$	$(8, 12, 21)$	$(15, 15, 22)$	$(24, 20, 24)$	$(30, 30, 30)$	$(40, 36, 35)$	$(50, 60, 55)$
$\hat{p}$	0.1381	0.0721	0.0486	0.0336	0.0211	0.0165	0.0093
$\hat{p}_F$	0.1173	0.0670	0.0463	0.0323	0.0204	0.0159	0.0090
$\hat{p}_G$	0.1152	0.0667	0.0462	0.0323	0.0204	0.0159	0.0090

Calculation of bias is more difficult again for unequal group sizes, because the number of possible outcomes involved in the summation of probabilities can be extremely large, even with an approximation corresponding to (8). For the current example, the number of outcomes for which $\Pr(T > t_*) < 10^{-6}$ for $p = \psi/250$ is 3822, 1630 and 966 respectively, for the three group sizes. The total number of outcomes to consider is the product of these numbers, which is around 6×10^9. With more group sizes than three, the number would be even more computationally prohibitive.

Instead of completely enumerating the (approximate) bias across all possible outcomes, we randomly sampled 100,000 outcomes, weighted by their probabilities. These were averaged to obtain the expected value of the estimator. Table 4 shows the expected value, bias and percentage bias of $\hat{p}_F$ for the current example.

Table 4. Expected value, bias and percentage bias for the estimator corrected for bias using Firth's correction, for $n = (3, 10, 20)$ and $s = (5, 20, 50)$ ($\psi = 0.122$).

p	0.001	0.005	0.01	0.03	0.05	0.1	0.125	0.15
$\mathrm{E}(\hat{p}_F)$	0.0010	0.0050	0.0100	0.0300	0.0500	0.1004	0.1242	0.1449
$\mathrm{Bias}(\hat{p}_F)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004	−0.0008	−0.0051
% $\mathrm{Bias}(\hat{p}_F)$	0.83	0.01	0.06	0.06	0.10	0.43	−0.68	−3.43

The Firth correction has virtually eliminated the bias for small p, with less than 0.1% bias even at $p = 0.05$. For larger p it is still very small, with well under 1% absolute bias at $p = \psi$. The virtues of this correction that were exhibited for equal group sizes clearly extend to unequal group sizes.

In situations where there is an overall number of positives n but no separate values of n_i, there is no analytical way to derive Firth's correction, but it would be a reasonable approximation to apply it using the observed values of n_i. Consider the example described above, in which the group sizes alternated between $s_1 = 5$ and $s_2 = 20$ with an overall

$n = 10$, and it took 20 tests; suppose that $n_1 = 2$ and $n_2 = 8$. Assuming the n_i's are fixed, and applying the steps above, results in $\hat{p} = 0.0657$ and $\hat{p}_F = 0.0582$. Note that if the sequential testing were ignored and binomial sampling assumed, the bias-corrected estimate would be 0.0624.

5. Discussion

We derived an estimator based on the bias correction method of Firth (1993), for group testing problems in which a negative binomial model is appropriate. The correction has been shown to be extremely effective, virtually eliminating the bias when the prevalence is small, and greatly reducing it for any p consistent with the group testing design. It is an improvement on the bias correction method of Gart (1991), and uses the same algebraic components related to the log-likelihood which were derived by Hepworth (2013). Unlike Gart's method, Firth's method involves the solution of an equation rather than requiring a plug-in estimate. It is therefore preventative rather than corrective, which has the advantage of avoiding undefined parameter estimates. Both methods result in similar MSE, so Firth's method can be generally recommended for inverse binomial group testing.

We showed that for groups of equal size, the method of Firth is equivalent to the method introduced by Burrows (1987), which has previously been shown to be a very good estimator in a range of group testing problems. Haber, Malinovsky & Albert (2018), who derived the Burrows-type estimator for the negative binomial model, also did so for group testing which proceeds until a set number of negative tests is reached. They found an unbiased estimator for that problem, and also showed that no unbiased estimator exists for the negative binomial model with a set number of positive tests. Having an unbiased estimator is generally desirable, though practitioners may be less interested in a method whose end-point depends on the number of negative tests rather than the number of positive tests.

Unequal group sizes creates a formidable problem for inverse binomial group testing, with the intractability of the distribution of the number of tests adding to the complexity of the infinite number of possible outcomes. However, we made inroads into this problem by proposing a design which considers each group size separately, and then combines the results. Not only is this a reasonable approach, but the outcomes can also be used to approximate the results from a more general design which considers only the total number of tests.

Another way to address the intractability of the information function would be to use the observed information, which Pritchard & Tebbs (2011) derived for this problem, and used to construct confidence intervals for p. Firth (1993) used both expected and observed information in illustrating his method, and found both to be acceptable. However, the expression for the observed information (and its derivative) is much more complicated, adding to an already complex and computer-intensive calculation. This would be an interesting extension to our work in this paper.

Another extension of our work would be to allow for misclassification. Xiong (2016) allowed for imperfect testing in optimizing group size for inverse binomial group testing, but that study assumed equal group sizes. All the derivations from the log-likelihood could include sensitivity and specificity parameters not equal to 1, though this would add to the complexity. Misclassification also creates issues near the boundaries of the parameter space.

References

BIGGERSTAFF, B.J. (2008). Confidence intervals for the difference of two proportions estimated from pooled samples. *Journal of agricultural, biological, and environmental statistics* **13**, 478–496.

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

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

GART, 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., 395–406, Amsterdam: Elsevier .

HABER, G. & MALINOVSKY, Y. (2017). Random walk designs for selecting pool sizes in group testing estimation with small samples. *Biometrical Journal* **59**, 1382–1398.

HABER, G., MALINOVSKY, Y. & ALBERT, P. (2018). Sequential estimation in the group testing problem. *Sequential Analysis* **37**, 1–17.

HAMMICK, P.A. & GASTWIRTH, J.L. (1994). Group testing for sensitive characteristics: Extension to higher prevalence levels. *International Statistical Review* **62**, 319–331.

HARDWICK, J., PAGE, C. & STOUT, Q.F. (1998). Sequentially deciding between two experiments for estimating a common success probability. *Journal of the American Statistical Association* **93**, 1502–1511.

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

HEPWORTH, G. (2013). Improved estimation of proportions using inverse binomial group testing. *Journal of agricultural, biological, and environmental statistics* **18**, 102–119.

HEPWORTH, G. & BIGGERSTAFF, B.J. (2017). Bias correction in estimating proportions by pooled testing. *Journal of Agricultural, Biological and Environmental Statistics* **22**, 602–614.

HEPWORTH, G. & WATSON, R. (2009). Debiased estimation of proportions in group testing. *Journal of the Royal Statistical Society C* **58**, 105–121.

HUGHES-OLIVER, J.M. (2006). Pooling experiments for blood screening and drug discovery. In *Screening*. Springer, pp. 48–68.

HUGHES-OLIVER, J.M. & SWALLOW, W.H. (1994). A two-stage adaptive group-testing procedure for estimating small proportions. *Journal of the American Statistical Association* **89**, 982–993.

KATHOLI, C.R. & UNNASCH, T.R. (2006). Important experimental parameters for determining infection rates in arthropod vectors using pool screening approaches. *The American journal of tropical medicine and hygiene* **74**, 779–785.

KLINE, R.L., BROTHERS, T.A., BROOKMEYER, R., ZEGER, S. & QUINN, T. (1989). Evaluation of human immunodeficiency virus seroprevalence in population surveys using pooled sera. *Journal of clinical microbiology* **27**, 1449–1452.

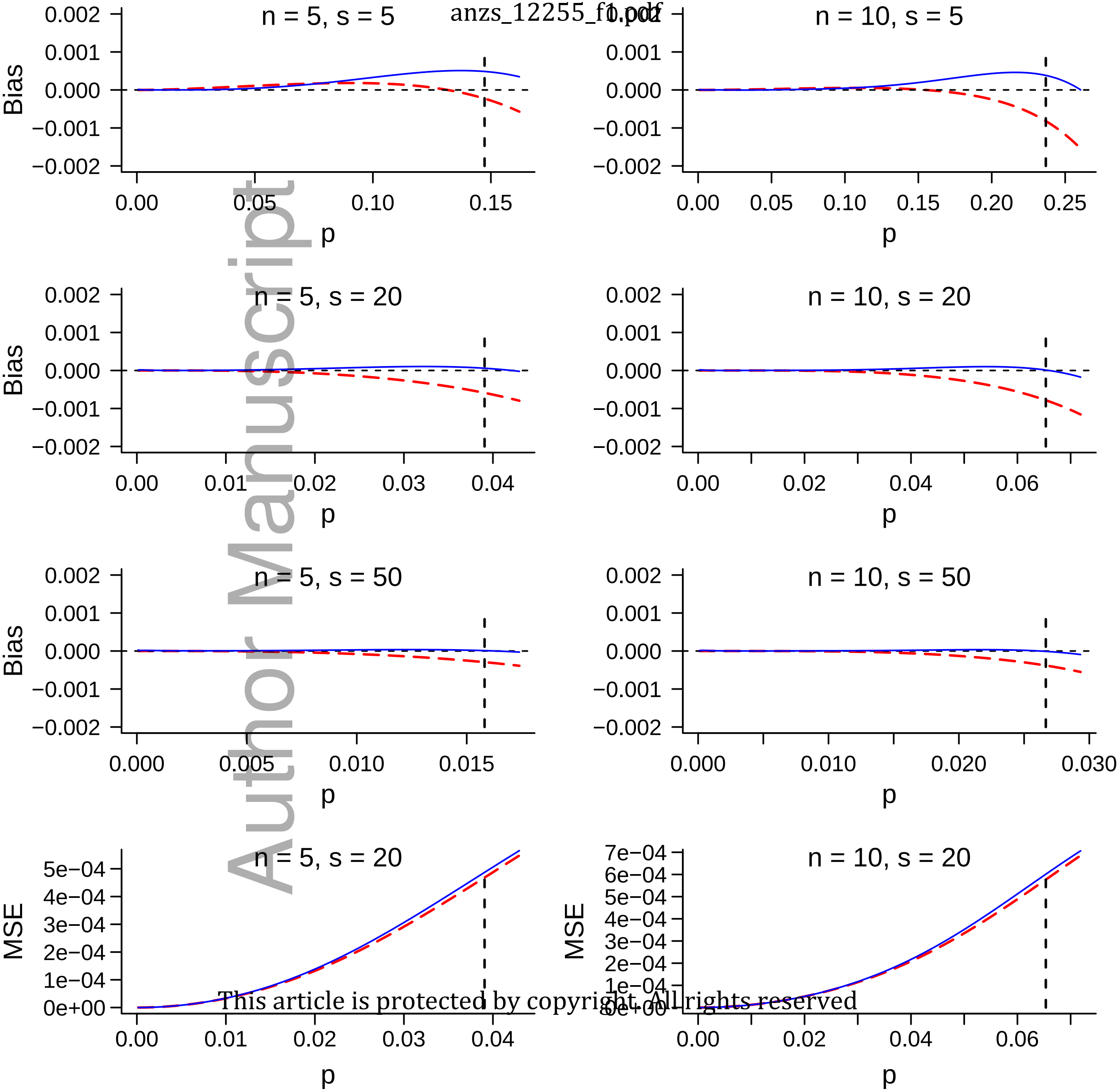
MONTESINOS-LÓPEZ, O.A., MONTESINOS-LÓPEZ, A., CROSSA, J. & ESKRIDGE, K. (2012). Sample size under inverse negative binomial group testing for accuracy in parameter estimation. *PloS one* **7**, e32250.

PRITCHARD, N.A. & TEBBS, J.M. (2011). Estimating disease prevalence using inverse binomial pooled testing. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 70–87.

RODRÍGUEZ-PÉREZ, M.A., KATHOLI, C.R., HASSAN, H.K. & UNNASCH, T.R. (2006). Large-scale entomologic assessment of onchocerca volvulus transmission by poolscreen pcr in mexico. *The American journal of tropical medicine and hygiene* **74**, 1026–1033.

WALTER, S.D., HILDRETH, S.W. & BEATY, B.J. (1980). Estimation of infection rates in populations of organisms using pools of variable size. *American Journal of Epidemiology* **112**, 124–128.

XIONG, W. (2016). The optimal group size using inverse binomial group testing considering misclassification. *Communications in Statistics-Theory and Methods* **45**, 4600–4610.


n = 5, s = 5
n = 10, s = 5
n = 5, s = 20
n = 10, s = 20
n = 5, s = 50
n = 10, s = 50
n = 5, s = 20
n = 10, s = 20
Bias
MSE
p
0.002
0.001
0.000
−0.001
−0.002
0.00
0.05
0.10
0.15
0.20
0.25
0.01
0.02
0.03
0.04
0.06
0.000
0.005
0.010
0.015
0.020
0.030
0e+00
1e−04
2e−04
3e−04
4e−04
5e−04
6e−04
7e−04