Sample Size Calculations for Surveys to Substantiate Freedom of Populations from Infectious Agents

Wesley O. Johnson

Department of Statistics

University of California

Davis, CA 95616

Chun-Lung Su and Ian A. Gardner

Department of Medicine and Epidemiology

School of Veterinary Medicine

University of California

Davis, CA 95616

and

Ronald Christensen

Department of Math and Statistics

University of New Mexico

Albuquerque, NM 87131-1141

Abstract

We develop a Bayesian approach to sample size computations for surveys designed to provide evidence of freedom from a disease or from an infectious agent. A population is considered "disease-free" when the prevalence or probability of disease is less than some threshold value. Prior distributions are specified for diagnostic test sensitivity and specificity and we test the null hypothesis that the prevalence is below the threshold. Sample size computations are developed using hypergeometric sampling for finite populations and binomial sampling for infinite populations. A normal approximation is also developed. Our procedures are compared with the frequentist methods of Cameron and Baldock (1998a) using examples of foot-and-mouth disease and bovine paratuberculosis. User friendly programs for sample size calculation and analysis of survey data are available at

`http://www.epi.ucdavis.edu/diagnostictests/download-bayesfreecalc.html`.

Keywords: Bayesian, Sensitivity, Specificity, Prevalence, Gibbs Sampling, Hypergeometric distribution, Prediction, Risk analysis

1 Introduction

Establishing that populations are free from important pathogens is vital in many applications including animal disease control. Typically, this involves sampling the population using a survey. It is crucial that these samples be large enough to draw appropriate inferences. Sample size computations were developed by Cannon and Roe (1982) for *perfect* diagnostic tests. These were modified by Cameron and Baldock (1998a) to incorporate *known* sensitivities and specificities. Clearly, the sensitivity and specificity of most tests are uncertain. For example, in bovine paratuberculosis an expert's best guess for the sensitivity of serologic testing was 45% with the belief that the sensitivity could easily be as high as 55%. Thus the assumption of known test characteristics is clearly untenable. We extend the work of Cameron and Baldock (1998a) and Cannon (2001) using a Bayesian approach that explicitly models the uncertainty in the sensitivity and specificity, as well as modeling information about the prevalence. We consider sampling only from a *single* homogeneous population.

Because of the complexity of their method, Cameron and Baldock (1998b) provided a widely disseminated computer program, FREECALC. Our software is also freely available on the internet.

Our approach to sample size selection extends the work of Geisser (1992). His predictive approach is detailed in section 3. Alternative Bayesian approaches are based on (1) achieving specified estimation coverage subject to minimizing an average (over future samples) length criterion, see Joseph, Berger and Belisle (1997) or Zou and Norman (2001), (2) minimizing Bayes risk, see Halpern, Brown and Hornberger (2001) or Müller and Parmigiani (1995), and (3) achieving a Bayes factor that exceeds a selected critical value under the hypothesis of interest, see Weiss (1997). Rahme, Joseph and Gyorkos (2000) addressed the problem of assessing sample size in the

infinite population case, as considered here, when the Bernoulli components of the binomial were subject to error. Their sample size criterion was to find the smallest n that gave a specified average (or predictive) coverage probability for intervals of specified length for the disease prevalence. Our approach focuses on the posterior probability that the disease prevalence is below a threshold and then predicts the probability of obtaining data that will make this posterior either large enough to conclude that the infection is under control or small enough to conclude that it is out of control.

Although our examples focus on veterinary medical applications, our methods are also applicable to human populations. For example, one may be interested in determining whether the prevalence of HIV, Dengue fever, or babesiosis is above a threshold beyond which some intervention is to be instituted. Alternatively, some companies may be required to guarantee that it's employees are essentially "drug-free."

In sections 2 and 3 we present background material and the proposed sample size methodology. In section 4 we compare our method with that of Cameron and Baldock (1998a) using foot-and-mouth disease, and present an example based on bovine paratuberculosis. Some theory on the behavior of the sample size calculations is developed in section 5. Final conclusions and a discussion of our software are given in section 6. Technical details are presented in the appendix.

2 Background Material

The presence of an infectious agent is denoted D. Let $\pi = pr(D)$ denote the "prevalence," that is, the proportion of diseased individuals in the population. Suppose that

n individuals are randomly sampled from a population. Each individual is tested for D. A positive result is denoted $+$ and a negative $-$. Let T^+ denote the number of test-positive results out of n. If the sample is with replacement from an infinite population and the test makes no errors, T^+ is binomial, that is, $T^+ \sim \text{Bin}(n, \pi)$. If, for a perfect test, sampling is without replacement from a finite population of N individuals containing d diseased individuals, then $\pi = d/N$ and T^+ is hypergeometric, $T^+ \sim \text{Hyp}(N, d, n)$.

The problem is to use T^+ to determine whether the prevalence of infection, π, is at or below a threshold, say, π_0. If $\pi \leq \pi_0$, no intervention is made in the population, while if $\pi > \pi_0$, the disease is considered "out of control" and measures would be taken to reduce the prevalence. Clearly, a decision must be made between $H_0 : \pi \leq \pi_0$, and $H_1 : \pi > \pi_0$.

In practice, diagnostic tests are imperfect. The sensitivity of a test is the probability that it correctly diagnoses diseased patients, say, $\eta = pr(+|D)$. The specificity is the probability that the test correctly diagnoses healthy individuals, $\theta = pr(-|\bar{D})$. In the infinite population case, $T^+ \sim \text{Bin}(n, \pi\eta + (1-\pi)(1-\theta))$. The distribution is more complicated for finite populations and is given in expression (10) of the appendix. The number T^+ is being used to derive estimates of π, η, and θ. Prior information plays a crucial role in the analysis, see Johnson, Gastwirth and Pearson (2001).

Uncertainty about unknown parameters is modelled with independent probability distributions for the parameters, see Johnson and Gastwirth (1991) and Johnson, Gastwirth and Pearson (2001). For the sensitivity and specificity, we use beta distributions $\eta \sim \text{Beta}(a_\eta, b_\eta)$ and $\theta \sim \text{Beta}(a_\theta, b_\theta)$. For the infinite population case, we take $\pi \sim \text{Beta}(a_\pi, b_\pi)$.

For the prevalence in finite populations, we use a discretized version of a beta. The prevalence is the unknown value $\pi = d/N$, which is restricted to the set $\{0, 1/N, 2/N, \ldots, 1\}$. We begin with a beta prior having $a \geq 1$ and $b \geq 1$. When $a > 1$ and $b > 1$, we use priors that assign to j/N the probability from the beta associated with the interval $j/N \pm 1/2N$. Obvious accommodations are made for $j = 0, 1$. If $a = 1$ or $b = 1$ we use a distribution based on renormalizing the beta densities evaluated at j/N. In the following, we discuss probabilities for infinite populations. Modifications for the discrete case are clear.

Prior information on test accuracies may be derived from published studies or from expert opinion. For example, a validation study of 100 subjects that estimates sensitivity as 0.95 can be modeled as a Beta$(96, 6)$ distribution, see Gastwirth, Johnson and Reneau (1991).

The prior on the prevalence involves information about the specific population. For example, if the population is a herd of cattle, information on prevalence is available from the owner and the herd veterinarian. Alternatively, it may be desirable to consider a relatively non-informative prior for prevalence.

To elicit expert opinion for prior distributions, the expert might be asked for the most likely value, m, and for one percentile. These determine a beta distribution that satisfies the constraints. The mode of a Beta(a, b) distribution is $m = (a - 1)/(a + b - 2)$, provided $a, b > 1$. For given m, the value of a is a function of b, say $a(b)$, so the distribution becomes Beta$(a(b), b)$. We then either elicit a value l such that $P(\pi > l) = \alpha$ or a value u such that $P(\pi < u) = \alpha$ and numerically determine the value of b so that a Beta$(a(b), b)$ satisfies the probability statement. If the mode is zero or one, we use Beta(1,b) or Beta(a,1) distributions, respectively, along with a probability statement. Our software automatically handles this type of elicitation.

Sometimes the elicitation involves the specification of three conditions, say a mode and two probability statements, in which case we try to find the "best" fitting beta distribution by trial and error. In all cases the prior as determined by the statistician is reconfirmed with the expert as being a reasonable approximation to their opinions. Bedrick, Christensen, and Johnson (1996, 1997) used similar elicitation techniques.

For a given sample size n, the posterior density for π is

$$p_n(\pi|x) = \frac{\int_0^1 \int_0^1 P_n(T^+ = x|\pi, \eta, \theta)p(\pi)p(\eta)p(\theta)d\eta d\theta}{\int_0^1 \int_0^1 \int_0^1 P_n(T^+ = x|\pi, \eta, \theta)dF(\pi)p(\eta)p(\theta)d\eta d\theta}, \tag{1}$$

where $F(\cdot)$ is the CDF for π and $\int dF(\pi)$ denotes Lesbesgue-Stieltjes integration. The posterior is approximated using Gibbs sampling for the binomial problem. The full conditional distributions (9) appear in the first appendix. For finite population problems, the discrete density (mass function) for $d \equiv N\pi$ is given in (11).

Primary interest focuses on

$$P_n(\pi \leq \pi_0|T^+ = x) \equiv P_n(\pi \leq \pi_0|x). \tag{2}$$

The null hypothesis is accepted if this is at least some value p_0 between 0 and 1. The largest value of x^* that results in $P_n(\pi \leq \pi_0|x^*) \geq p_0$ is called a critical value and is denoted $x_{c0} \equiv x_{c0}(n)$. The alternative is accepted if the *complementary* probability is greater than $p_1 \geq 1 - p_0$. The smallest value of x^* that results in $P_n(\pi > \pi_0|x^*) > p_1$ is a critical value denoted $x_{c1} \equiv x_{c1}(n)$. Unless $p_1 = 1 - p_0$, this procedure provides a range in which neither hypothesis is accepted. Making both p_0 and p_1 large ensures a high probability of being correct for both actions. Some combinations of n, p_0, and p_1 may make taking some actions impossible. Note that when $p_1 = 1 - p_0$, $x_{c1} = x_{c0} + 1$, unless $x_{c1} = 0$ or $x_{c0} = n$.

3 Predictive Sample Size Determination

Following Geisser (1992), we develop Bayesian predictive criteria for determining a sample size. The first criterion calculates the predictive probability of being able to make a correct decision in favor of the null. The goal is to find the sample size such that this predictive probability exceeds a pre-specified value, say $\beta_0 = 0.95$. The second criterion involves a similar calculation for "proving" the alternative.

The sample size needed to obtain a correct conclusion depends on the true prevalence. Rather than averaging over the prevalence, we develop sample sizes as a function of the true prevalence. In "proving" that the null hypothesis is true, we presume that there is a "true" value of π, say π_T with $\pi_T < \pi_0$. We calculate the predictive probability

$$\tilde{p}_{n0} \equiv P_n^*\{P_n(\pi \leq \pi_0|x^*) \geq p_0|\pi_T\} = \sum_{x^*=0}^{n} I_{\{P_n(\pi\leq\pi_0|T^+=x^*)\geq p_0\}}(x^*) P_n^*(T^+ = x^*|\pi_T), \tag{3}$$

where the calculation uses the distribution of future data $T^+ = x^*$ of size n, i.e.,

$$P_n^*(T^+ = x^*|\pi_T) = \int P_n(T^+ = x^*|\pi_T, \eta, \theta) p(\eta, \theta) d\eta d\theta,$$

which can either be botained exactly or by Monte Carlo integration. In our examples, we generally select $\pi_T = 0$. More generally, π_T should be taken as a value for which the scientist wants to be confident of a correct decision. The closer π_T is to π_0, the more difficult it will be to satisfy the criterion. Finally, we search for the smallest value of n for which we are $\beta_0 \times 100\%$ certain that we will be able to make a decision, namely, find n such that $\tilde{p}_{n0} \geq \beta_0$.

The second situation corresponds to the goal of "proving" the alternative hypothesis $H_1 : \pi > \pi_0$ when it is true. Thus π_T is selected as a value that is of particular

interest to the scientist with $\pi_T > \pi_0$. The calculation involves finding n such that

$$\tilde{p}_{n1} \equiv P_n^*\{P_n(\pi > \pi_0|x^*) > p_1|\pi_T\} \geq \beta_1. \tag{4}$$

In the two situations, the required sample size is represented as

$$n_{ci} \equiv \arg\min_n \left(\tilde{p}_{ni} \geq \beta_i\right). \tag{5}$$

There are situations where one can determine in advance that it would be impossible to achieve these criteria. These are discussed in the examples and section 5. If both null and alternative criteria are of interest, the selected value of n is simply the larger of the resulting two values, $n_c \equiv \max_i \left(n_{ci}\right)$.

Finding a single value n_{ci} is computationally intensive, but theoretically one could obtain $n_{ci}(\pi)$ for $\pi \in [0, 1]$. Using the prior $p(\pi)$, it is also natural to consider

$$n_{Bi} = \int_{H_i} n_{ci}(\pi) p(\pi|H_i) d\pi,$$

where $p(\pi|H_i)$ is the conditional density for π given H_i. One could then obtain

$$n_B = \max_i \left(n_{Bi}\right) \text{ or } n_B^* = P(H_0) n_{B0} + P(H_1) n_{B1},$$

where $P(H_i)$ is the prior probability of H_i. Note that if $P(H_0)$ is large, $n_{c0}(\pi)$ should be small for all π in H_0, so n_{B0} should be small. In the limit, if $P(H_0) = 1$, $n_B^* = n_{B0} = 0$, whereas no amount of data would make us decide for the alternative.

In our computations, we use a starting value for finding n that is based on a normal approximation to the binomial. This approximation uses the variance stabilizing (arcsin square root) transformation of the binomial proportion T^+/n. (See appendix for details.) The normal approximation is extremely fast and, for some situations, quite accurate. When N is moderate or large, the normal approximation also serves as a first approximation to the hypergeometric model.

4 Illustrations

We illustrate our methods using two examples from Cameron and Baldock (1998a), (CB). While these involve animal infections our results apply more broadly. Thresholds π_0 and population sizes N are selected in the ranges 0.05-0.30 and 50-500, respectively. These inputs are typical for the populations considered. For finite populations, $N\pi_T$ should be an integer; if not, we redefine π_T by rounding $N\pi_T$. The methods also apply when screening test performance is known without error (e.g. take a_η, b_η, a_θ, and b_θ to be large with appropriate modal values).

4.1 Foot-and-Mouth Disease (FMD)

This example compares our results to CB's. Consider an enzyme-linked immunosorbent (ELISA) test with unknown sensitivity and specificity that is available for testing cattle herds of $N = 50$, 265 and 500 animals that show no clinical evidence of FMD. We take π_T to be zero or 0.3 for the null and alternative calculations, respectively.

Prior information is used to model the uncertainty in test performance. Our expert (Angus Cameron) is 90% sure that the sensitivity is greater than 0.9 and that the specificity is greater than 0.95. His prior modes are 0.95 and 0.98, respectively. These constraints lead to Beta$(68.74, 4.57)$ and Beta$(107.2, 3.17)$ priors for sensitivity and specificity, respectively. A uniform distribution is assumed for the prevalence. This attaches prior probability π_0 to H_0 and probability $1 - \pi_0$ to H_1. For a highly virulent disease such as FMD that displays overt clinical signs, the absence of such signs may provide strong *a priori* evidence that disease is not present. This might justify giving greater weight to the null hypothesis. A Beta(1,b) distribution attaches prior probability $1 - (1 - \pi_0)^b$ to H_0.

In comparing our approach to CB's, their null and alternative hypotheses are reversed from ours since they are most concerned with not making the Type I error of concluding that the prevalence is "small" when in fact it is not. For CB (both veterinary epidemiologists) the status quo was that the prevalence will be large. They would plan to intervene unless there was statistical proof that it was unnecessary. To mimic that behavior, we focus on finding sample sizes n_{c0} to prove low prevalence. To illustrate the comparison, we use $p_0 = 0.95$ which is an intuitively appealing value but bears no direct relationship to CB's procedure.

To use CB's method, we fixed the probabilities of Type I and Type II error at 5%, and assumed the known sensitivity and specificity to be 0.95 and 0.98, respectively. These are the modal values of our prior distributions for η and θ. The "minimum expected prevalence" (MEP) is defined by CB to be the smallest prevalence that is anticipated by veterinarians among those herds that they consider to be having problems with the infection. We treat the MEP like our π_0. CB's procedure produces critical values x_c^* and sample sizes n_c^* so that both $\{1-P(\text{Type II error})\} = P_{n_c^*}(T^+ \leq x_c^* | \pi = 0, \eta = 0.95, \theta = 0.98) \geq 0.95$ and $P(\text{Type I error}) = P_{n_c^*}(T^+ \leq x_c^* | \pi = \pi_0, \eta = 0.95, \theta = 0.98) \leq 0.05$ hold.

For $N = 265$ and $\pi_0 = 0.1$, CB's method generates $n_c^* = 64$, $x_c^* = 3$. To make a direct comparison to Bayesian procedures we found $P_{64}(\pi \leq 0.1|3) = 0.962$. Thus, for $n = 64$, our Bayesian procedure that accepts H_0 when $P_{64}(\pi \leq 0.1|x) \geq 0.962$ is identical to CB's $\alpha = .05$ test. Moreover, for $p_0 = 0.962$, the predictive probability, when $\pi_T = 0$, of accepting H_0 is $\tilde{p}_{64,0} = 0.85$, which is loosely comparable to the 95% power of the CB procedure. Using $p_0 = 0.962$, we require a sample size $n_{c0} = 128$ to achieve our 95% certainty of accepting H_0 when $\pi_T = 0$. For $n = 64$ with $p_0 = 0.95$, because the data are discrete, the predictive probability under $\pi_T = 0$ remains

$\tilde{p}_{64,0} = 0.85$. Nonetheless, using $p_0 = 0.95$, rather than 0.962, we require a smaller sample size, $n_{c0} = 102$, to achieve a predictive probability of $\beta_0 = 0.95$.

Table 1 presents results comparing our null and alternative approaches using $p_0 = p_1 = 0.95$ with the CB method. Our criteria can be considerably more stringent than that used by CB. This, along with our incorporation of uncertainty into the sensitivity and specificity, can cause our sample sizes to be larger than CB's. Also note in Table 1 that there is very little change in the Bayesian numbers for the null among $N = 265$, $N = 500$, and the binomial when $\pi_0 \geq 0.15$. For the alternative, all of the Bayesian procedures are reasonably close when $\pi_0 < 0.15$. (This being less so for $N = 50$.) Similar observations held for other calculations made under other circumstances.

Figures 1 and 2 give plots of sample size, n, versus $\tilde{p}_{n0}$ for the binomial, hypergeometric, and normal null calculations, with the same priors that were given for this example, and for thresholds π_0 between 0.05 and 0.3. For larger thresholds, the normal approximation can be quite reasonable. The saw-tooth shape of the plots, a result of the discreteness of T^+, is discussed in Chernick and Liu (2002). We clarify this behavior in section 5.

Although our priors for η and θ involve reasonably strong prior information, we also tried a considerably more concentrated prior: 0.98 prior probability that sensitivity and specificity are at least 0.94 and 0.97, respectively, with the same modes; 0.95 and 0.98. With $p_0 = \beta_0 = .95$, a sample of size $n = 72$ yielded $\tilde{p}_{n0} = 0.98$ (with $n = 70$ $\tilde{p}_{n0} = 0.93$) based on a null calculation, which is much closer to the CB value of 64. Therefore, assuming known sensitivity and specificity can have a large effect on the sample size calculation.

4.2 Bovine Paratuberculosis

Typically, in cattle herds infected with paratuberculosis, prevalences are lower than 30%. Certification of disease free status is complicated by the low sensitivity of serum ELISA and other tests. Recent evidence (Whitlock et al., 2000) confirms an earlier report by Ridge et al. (1991) that the ELISA sensitivity is lower than 60%. We elicited expert opinion about the sensitivity of the ELISA in stage II paratuberculosis from an expert (Michael Collins, University of Wisconsin, Madison). His mode for the sensitivity was 0.45, and his 90% upper limit was 0.55, resulting in a Beta$(19.34, 23.41)$ distribution. Similarly, his mode and 10% lower limit for specificity were 0.99 and 0.97, respectively, yielding a Beta$(152.08, 2.53)$ distribution. We used a uniform prior for π.

For $N = 50$ with a threshold of $\pi_0 = 0.1$, using $p_0 = 0.95$, and sampling the entire population there is only about a 50% chance of accepting H_0 when $\pi_T = 0$ and an 80% chance of accepting when the threshold is $\pi_0 = 0.2$. With $\pi_T = 0.2$, $\pi_0 = 0.05$, and $p_1 = 0.95$, there is about an 80% chance of accepting H_1. Raising the threshold to $\pi_0 = 0.1$ and keeping $p_1 = 0.95$, it is impossible to accept H_1, but with $p_1 = 0.90$, there is about a 50% chance of accepting H_1.

With $N = 265$, $\pi_0 = 0.1$ and $p_i = 0.95$, it is not possible to achieve the $\beta_i = 0.95$ null or alternative criteria. By sampling the entire population, we can only establish about a 90% chance of accepting H_0 when $\pi_T = 0$ and, even by lowering p_1 to 0.9, an 83% chance of accepting H_1 when $\pi_T = 0.2$. For this latter situation, sampling only 100 animals results in a 70% probability of accepting H_1. These calculations illustrate the fact that our criteria may not be achievable.

With $N = 500$, $\pi_0 = 0.1$ and $p_0 = 0.95$, we can achieve a 90% chance of accepting H_0 when $\pi_T = 0$ with $n = 350$. Sampling all $n = 500$ animals only results in

$\tilde{p}_{n0} = 0.93$. On the other hand, if $\pi_0 = 0.2$, a sample of only $n = 113$ gives a 95% chance of accepting the null. The normal approximation is much better for the null calculation when the threshold is 0.2 than when it is 0.1.

Sample size results are given in Table 2. For the null and alternative, minimum samples sizes n_{ci} are given for $p_0 = p_1 = 0.95$ and $\beta_0 = \beta_1 = 0.95$ when possible. When the criteria are not attainable, a variety of results are illustrated.

Finally, consider changing the prior on the prevalence. Suppose we are 95% sure *a priori* that the prevalence is less than 0.10 and assume the mode of the prior is zero. This results in a Beta(1, 28.46) prior. With all other information the same, let $\pi_0 = 0.1$ and $p_0 = 0.95$. For $N = 50$ and a sample size of $n = 50$, $\tilde{p}_{n0} = 0.93$ as compared to $\tilde{p}_{n0} = 0.49$ with the uniform prior. With $N = 265$, a sample of $n = 110$ results in $\tilde{p}_{n0} = 0.96$, whereas with the uniform $n = 265$ only gave $\tilde{p}_{n0} = 0.89$. For $N = 500$ or an infinite population, $n = 150$ results in $\tilde{p}_{n0} = 0.95$ whereas with the uniform it took a sample of $n = 350$ out of $N = 500$ to achieve $\tilde{p}_{n0} = 0.90$ and in the infinite N case, $n = 339$ to achieve $\tilde{p}_{n0} = 0.88$. The normal approximation results in $n = 165$.

5 Behavior of the Sample Size Calculations

We now give a result that can be used to develop efficient search procedures for finding sample sizes and another that explicates the saw-toothed shape of $\tilde{p}_{n0}$ in Figures 1 and 2. We provide a formal justification in the binomial case for the null hypothesis procedure. The results are proved in the appendix.

Proposition 1. In the binomial case, provided $\eta + \theta > 1$ with prior probability 1,

$$\begin{aligned} P_n(\pi \leq \pi_0|x) &> P_n(\pi \leq \pi_0|x+1), \qquad x < n \\ P_n(\pi \leq \pi_0|x) &< P_{n+1}(\pi \leq \pi_0|x), \qquad x \leq n. \end{aligned} \tag{6}$$

In our experience, the results in Proposition 1 appear to also hold Hypergeometric case, though we have not proven this analytically. Efficient search procedures based on Proposition 1, and Proposition 2 below, can be devised. For example, for a given n, check whether $P_n(\pi \leq \pi_0|0) \geq p_0$. If not, then n must be increased since there is no possibility that the selected n will be appropriate. If this criterion is satisfied, then the procedure will continue with increasing x until a critical value x_{c0} is obtained such that $P_n(\pi \leq \pi_0|x_{c0}) \geq p_0$ and $P_n(\pi \leq \pi_0|x_{c0}+1) < p_0$. For the given n, values larger than x_{c0} result in posterior probabilities of H_0 that are smaller than p_0, thus $\tilde{p}_{n0} = P_n^*(T^+ \leq x_{c0}(n)|\pi_T)$. Moreover, by (6), we also have $x_{c0}(n) \leq x_{c0}(n+1)$, because $P_{n+1}(\pi \leq \pi_0|x_{c0}(n)) > P_n(\pi \leq \pi_0|x_{c0}(n)) \geq p_0$. Finally, using Proposition 2, if $x_{c0}(n+1) = x_{c0}(n)$, then it is impossible to improve the sample size criterion by increasing from n to $n+1$.

Our programs for obtaining sample sizes are described in the discussion section below, and our (crude) algorithms are presented in the appendix. The current version of our program does not use these efficient methods for the binomial, but does use them for the hypergeometric, if requested. Results are virtually identical using either method.

The calculations to obtain (3) and (4) are different for the finite and infinite population cases but both can be slow if n is large. Results can be obtained quickly using our computer code for moderate $N, n < 500$ using an efficient search for the n that achieves either criterion. However, with large n or N this search can take

considerable time. We have obtained results for $500 < n, N < 1000$ in reasonable amounts of time using a simple trial-and-error approach. To execute this, a value of n is posited and the corresponding value $\tilde{p}_{ni}$ is obtained. If $\tilde{p}_{ni} < \beta_i$, simply try a larger n and continue until finding an n with $\tilde{p}_{ni} \geq \beta_i$ or you realize that β_i can never be achieved.

The saw-tooth behavior follows from

Proposition 2. Let $m < n$ be given and define $\delta_T = \pi_T\eta + (1 - \pi_T)(1 - \theta)$. Then in the binomial case, and provided $0 < \delta_T < 1$ with probability one,

$$P_m(T^+ \leq x|\pi_T, \eta, \theta) = P_n(T^+ \leq x|\pi_T, \eta, \theta) + \delta_T \sum_{k=m}^{n-1} P_k(T^+ = x|\pi_T, \eta, \theta). \tag{7}$$

Moreover, with $h_n^*(x) = \int P_n(T^+ \leq x|\pi_T, \eta, \theta)p(\eta, \theta)d\eta d\theta$,

$$h_m^*(x) > h_n^*(x), \quad m < n, \quad x \leq m; \qquad h_{n+1}^*(x_1) > h_n^*(x_0), \quad x_0 < x_1 \leq n. \tag{8}$$

For null calculations, suppose that $x_0 \equiv x_{c0}(m)$ is the critical value for both m and $n > m$. Then (8) implies that $\tilde{p}_{n0} < \tilde{p}_{m0}$. It is only when $x_0 \equiv x_{c0}(n) < x_{c0}(n+1) \equiv x_1$ that $\tilde{p}_{n+1,0} > \tilde{p}_{n0}$, by the second part of (8). Thus $\tilde{p}_{n0}$ is decreasing in n until it gets so low that the critical value is increased to $x_0 + 1$, at which point $\tilde{p}_{n0}$ jumps. This is precisely the saw-tooth behavior that is exhibited in Figures 1 and 2.

6 Discussion

We developed a Bayesian sample size calculator for determining "disease-free-status" in a single homogeneous population based on infinite or finite population sampling when test parameters are unknown and where prior information on prevalence and testing parameters are incorporated. Our methods work well over the dissimilar examples considered.

Because uncertainty about the sensitivity and specificity are incorporated into the model, larger sample sizes are required than when these parameters are treated as known. Table 1 indicates the need for larger sample sizes but we also showed that precise knowledge of test accuracy resulted in sample sizes that were comparable to those in CB. Since *sensitivity and specificity are rarely, if ever, known precisely,* incorporating uncertainty about the test accuracies should give more realistic results. Another factor in the larger sample sizes is that we incorporated more stringent criteria for decision making than are imposed by CB. In practice, the values of p_i and β_i should be chosen so that the required sample size is within the resources of the experimenter.

The program developed to implement the calculations, BDFree 1.0, is coded in Fortran with a Visual Basic graphic user interface (GUI). The parameters for the beta priors are automatically calculated when one inputs the mode m along with α and either l or u as described in section 2. For a given n, users can obtain $\tilde{p}_{ni}$ based on all of our proposed methods. They can subsequently iterate by trial and error until a suitable n is found that achieves the criterion or they can choose an option whereby the program automatically determines n_{ci}. Our GUI indicates when the criterion is not achievable. If a suitable n were selected and data subsequently collected, the GUI can also be used to calculate the posterior probability of the null for the given data. The beta version of the program is available on the web at http://www.epi.ucdavis.edu/diagnostictests/download-bayesfreecalc.html.

Many pathogen surveys involve two-stage cluster sampling with selection of multiple subpopulations and then selection of a subset of individuals from each subpopulation for testing, see Cameron and Baldock (1998b) or Suess, Gardner and Johnson (2002). We are currently developing sample size methods for these designs.

Acknowledgements:

This research was partially supported by the USDA NRI Competitive Grants Program award No. 01-02494. The authors thank Doctors Angus Cameron and Michael Collins for providing expert opinion used to derive the prior distributions and Dr. Patrick McInturff and Adam Branscum for their help in preparing the beta version of the GUI and for their comments on the manuscript. We also thank an associate editor and two referees for insightful comments that lead to a much improved manuscript.

References

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**, 1450-1460.

Bedrick, E. J., Christensen, R., and Johnson, W. (1997). Bayesian binomial regression. *The American Statistician*, **51**, 211-218.

Cameron, A.R., Baldock, F.C. (1998a). A new probability formula for surveys to substantiate freedom from disease. *Preventive Veterinary Medicine* **34**, 1-17.

Cameron, A.R., Baldock, F.C, (1998b). Two-stage sampling in surveys to substantiate freedom from disease. *Preventive Veterinary Medicine* **34**, 19-30.

Cannon, R.M., Roe, R.T. (1982). Livestock disease surveys. A field manual for veterinarians. Bureau of Range Science, Department of Primary Industry. Australian Government Publishing Service, Canberra.

Cannon, R.M. (2001). Sense and sensitivity - designing surveys based on an imperfect test. *Preventive Veterinary Medicine* **49**, 141-63.

Chernick, M.R. and Liu, C.Y. (2002). The saw-toothed behavior of power versus sample size and software solutions: Single binomial proportion using exact methods. *The American Statistician* **56**, 149-155.

Gastwirth, J.L., Johnson, W.O. and Reneau, D.M. (1991). Bayesian analysis of screening data: Application to AIDS in blood donors. *Canadian Journal of Statistics*, **19**, 135-150.

Geisser, S. (1992). On the curtailment of sampling. *Canadian Journal of Statistics* **20**, 297-309.

Halpern, J., Brown, B.W. and Hornberger, J. (2001). The sample size for a clinical trial: A Bayesian-decision theoretic approach. *Statistics in Medicine* **20**, 841-858.

Johnson, W.O. and Gastwirth, J.L. (1991). Bayesian inference for medical screening tests: Approximations useful for the analysis of AIDS data. *Journal of the Royal Statistical Society, B* **53**, 427-439.

Johnson, W.O., Gastwirth, J.L., and Pearson, L.M. (2001). Screening without a gold standard: The Hui-Walter paradigm revisited. *American Journal of Epidemiology* **153**, 921-924.

Joseph L, du Berger R, Belisle P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in Medicine* **16**, 769-781.

Müller, P. and Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association* **90**, 1322-1330.

Rahme, E., Joseph, L. and Gyorkos, T. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Applied Statistics* **49**, 119-228.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications, 2nd Edn.* Wiley, New York.

Ridge, S.E., Morgan, I.R., Sockett, D.C., Collins, M.T., Condon, R.J., Skilbeck, N.W., Webber, J.J. (1991). Comparison of the Johne's absorbed EIA and the complement-fixation test for the diagnosis of Johne's disease in cattle. *Australian Veterinary Journal* **68**, 253-257.

Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods* Springer-Verlag, New York.

Suess, E., Gardner, I.A. and Johnson, W.O. (2002). Hierarchical Bayesian model for prevalence inferences and determination of a country's status for an animal pathogen. *Preventive Veterinary Medicine* **55**, 155-171.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *The Statistician* **46**, 185-191.

Whitlock, R.,H., Wells, S.J., Sweeney, R.W. and Van Tiem, J. (2000). ELISA and fecal culture for paratuberculosis (Johne's disease): Sensitivity and specificity of each method. *Veterinary Microbiology* **77**, 387-398.

Zou, K.H. and Normand, S.-L. T. (2001). On determination of sample size in hierarchical binomial models. *Statistics in Medicine* **20**, 2163-2182.

Appendix 1: Infinite Population Case

Our method is based on data augmentation and Gibbs sampling. The true disease state (D) and test results (T) are dichotomous. Consider latent variables $z = (z_1, z_2)$ giving the numbers of true positives and negatives, respectively. (See Table 3.) We calculate each full conditional posterior distribution and iteratively sample from these

to numerically approximate the posterior distribution of $\pi|x$.

The joint posterior based on augmented data satisfies $p(\pi,\eta,\theta|x,z) \propto \pi^{z_1+n-x-z_2+a_\pi-1}(1-\pi)^{x-z_1+z_2+b_\pi-1}\eta^{z_1+a_\eta-1}(1-\eta)^{n-x-z_2+b_\eta-1}\theta^{z_2+a_\theta-1}(1-\theta)^{x-z_1+b_\theta-1}$. After some calculation, the full conditional posterior distributions of π,η,θ,z_1,z_2 are

$$z_1|\text{else} \sim \text{Bin}\left(x, \frac{\pi\eta}{\pi\eta+(1-\pi)(1-\theta)}\right),\ z_2|\text{else} \sim \text{Bin}\left(n, \frac{(1-\pi)\theta}{\pi(1-\eta)+(1-\pi)\theta}\right),$$

$$\eta|\text{else} \sim \text{Beta}(z_1+a_\eta, n-x-z_2+b_\eta),\ \theta|\text{else} \sim \text{Beta}(z_2+a_\theta, x-z_1+b_\theta), \qquad (9)$$

$$\pi|\text{else} \sim \text{Beta}(z_1+n-x-z_2+a_\pi, x-z_1+z_2+b_\pi),$$

where "else" indicates conditioning on x and the other four variables.

The sample size calculation algorithm proceeds as follows.

1. Sample $\eta^i \sim \text{Beta}(a_\eta, b_\eta)$, $\theta^i \sim \text{Beta}(a_\theta, b_\theta)$, $i = 1,\ldots,m$.

2. Given n, sample $x^i \sim \text{Bin}(n, \pi_T\eta^i + (1-\pi_T)(1-\theta^i))$, $i = 1,\ldots,m$.

   (a) For each x^i, $i = 1,\ldots,m$, iteratively sample $(z^i_{1j}, z^i_{2j}, \eta^i_j, \theta^i_j, \pi^i_j)$, $j = 1,\ldots,m_1$ from the full conditional distributions in (9). Approximate $P_n(\pi \le \pi_0|x^i) \approx \sum_{j=1}^{m_1} \frac{I(\pi^i_j \le \pi_0)}{m_1} \equiv g_{n0}(x^i)$ for the null case or $P_n(\pi > \pi_0|x^i) \approx \sum_{j=1}^{m_1} \frac{I(\pi^i_j > \pi_0)}{m_1} \equiv g_{n1}(x^i)$ for the alternative.

   (b) Calculate $\tilde{p}_{n0} \approx \sum_{i=1}^{m} \frac{I(g_{n0}(x^i) \ge p_0)}{m}$ or $\tilde{p}_{n1} \approx \sum_{i=1}^{m} \frac{I(g_{n1}(x^i) > p_1)}{m}$.

3. Increase n and repeat step (2) until finding the smallest n that satisfies equation (3) or equation (4), that is, n_{c0} or n_{c1}, respectively. Also report the corresponding number of test positive outcomes, $x_{c0}(n_{c0}) = \arg\max_x\{P_{n_{c0}}(\pi \le \pi_0|x) \ge p_0\}$ for the null or $x_{c1}(n_{c1}) = \arg\min_x\{P_{n_{c1}}(\pi > \pi_0|x) > p_1\}$ for the alternative case.

Similar ideas are applied to the finite and large sample cases.

Appendix 2: Finite Population Case

Following CB, with number of infected animals d, sensitivity η, and specificity θ,

$$P_n(T^+ = x|d,\eta,\theta) = \sum_{y=max(n-N+d,0)}^{min(n,d)} \frac{\binom{d}{y}\binom{N-d}{n-y}}{\binom{N}{n}} \sum_{j=max(x-n+y,0)}^{min(x,y)} \binom{y}{j}\eta^j(1-\eta)^{y-j}\binom{n-y}{x-j}\theta^{n-x-y+j}(1-\theta)^{x-j}. \tag{10}$$

Assuming the discrete prior density $p(d)$ and Beta priors for η and θ, the posterior mass function for $d|x$ is,

$$\begin{aligned} p(d|x) \propto & \sum_{y=max(n-N+d,0)}^{min(n,d)} \frac{\binom{d}{y}\binom{N-d}{n-y}}{\binom{N}{n}} \sum_{j=max(x-n+y,0)}^{min(x,y)} \binom{y}{j} B(j+a_\eta, y-j+b_\eta)\times \\ & \binom{n-y}{x-j} B(n-x-y+j+a_\theta, x-j+b_\theta)p(d), \end{aligned} \tag{11}$$

where $B(\cdot,\cdot)$ is the beta function.

1. Sample η^i and θ^i from their corresponding priors, $i = 1,\ldots,m$.

2. Given n, sample $y^i \sim$ Hyp$(N, N\pi_T, n)$. Sample t_{1i} and t_{2i} from Bin(y^i,η^i) and Bin$(n-y^i, 1-\theta^i)$, respectively. Let $x^i = t_{1i} + t_{2i}, i = 1,\ldots,m$.

3. For each x^i, $i = 1,\ldots,m$, calculate the discrete posterior distribution for d, $p_n(d|x^i)$, by normalizing (11). Obtain $g_{n0}(x^i) = P_n(d \le N\pi_0|x^i)$ or $g_{n1}(x^i) = P_n(d > N\pi_0|x^i)$ for the alternative.

4. Calculate $\sum_{i=1}^m \frac{I(g_{n0}(x^i)\ge p_0)}{m}$ or $\sum_{i=1}^m \frac{I(g_{n1}(x^i)>p_1)}{m}$.

5. Find n_{ck} and x_{ck} using the same procedure as in step 3 for the infinite case.

Remarks: If the sample size is near N, the range of possible values for y in the first summation of (11) is reduced dramatically, resulting in faster calculations. If $n = N$, only one term $y = d$ needs to be considered. Thus, one can easily check whether

sampling the entire herd would satisfy the criterion before searching for the optimal sample size. For instance, it took less than 30 seconds to run the null calculation with $N = n = 2000$ under the same conditions as in Table 1 with a 466 CPU Celeron PC. For each d, the binomial coefficients and beta functions can be reused because y's and j's are both consecutive.

We have done analytical integration to obtain the marginal posterior for d in the hypergeometric case whereas we performed numerical integration in the binomial case. A result that is analogous to (10) is obtainable for the binomial case.

Appendix 3: Normal Approximation

Let $\delta = \pi\eta + (1-\pi)(1-\theta)$ and $x|\delta \sim \text{Bin}(n, \delta)$ and let $y_n \equiv \sin^{-1}\sqrt{x/n}$. For large n, we have $y_n|\delta \dot{\sim} N(\rho, \frac{1}{4n})$ where $\rho = \sin^{-1}\sqrt{\delta}$, cf. Rao (1973, p. 427). If we assume the (induced) prior on ρ is approximately normal, that is $\rho \dot{\sim} N(\mu_\rho, 1/\tau_\rho)$, then the posterior is $\rho|x \dot{\sim} N(wy_n + (1-w)\mu_\rho, \frac{1}{4n+\tau_\rho})$, where $w = \frac{4n}{4n+\tau_\rho}$ and μ_ρ and τ_ρ are approximated by sampling. The predictive (marginal) distribution is $y_n \dot{\sim} N(\mu_\rho, 1/\tau_\rho + 1/4n)$.

Let $\delta_0 = \pi_0\eta + (1-\pi_0)(1-\theta)$. Provided $\eta + \theta > 1$,

$$g_{n0}(x) = P_n(\pi \le \pi_0|x) = P_n(\delta \le \delta_0|x) = P_n(\rho \le \sin^{-1}\sqrt{\delta_0}|x),$$

and

$$\begin{aligned} & g_{n0}(x) \ge p_0 \iff P_n(\rho \le \sin^{-1}\sqrt{\delta_0}|x) \ge p_0 \\ & \iff \sqrt{\tau_\rho + 4n}\left[\sin^{-1}\sqrt{\delta_0} - \{wy_n + (1-w)\mu_\rho\}\right] \ge \Phi^{-1}(p_0) \qquad (12) \\ & \iff y_n \le y_{n0} \end{aligned}$$

where $y_{n0} = \left[\sin^{-1}\sqrt{\delta_0} - (1-w)\mu_\rho - \Phi^{-1}(p_0)/\sqrt{\tau_\rho + 4n}\right]/w$.

For the purpose of determining sample size when $\pi = \pi_T$, we consider the distribution of future data $x^* = T^+$ of size n, or more particularly, of $y_n^* \equiv \sin^{-1}(\sqrt{T^+/n})$. As seen earlier, this is $y_n^*|\pi_T \dot{\sim} N(\mu_{\rho_T}, 1/4n + 1/\tau_{\rho_T})$. Here π_T, δ_T, and ρ_T are particular values of π, δ, and ρ and μ_{ρ_T} and τ_{ρ_T} are approximated by sampling. Asymptotically,

$$\begin{aligned} & P_n^*(g_{n0}(x^*) \geq p_0|\pi_T) \geq \beta_0 \Longleftrightarrow P_n^*(y_n^* \leq y_{n0}|\pi_T) \geq \beta_0 \\ & \Longleftrightarrow \frac{y_{n0} - \mu_{\rho_T}}{\sqrt{1/4n + 1/\tau_{\rho_T}}} \geq \Phi^{-1}(\beta_0). \end{aligned} \tag{13}$$

We find the smallest n that satisfies this inequality.

In particular,

1. Sample π^i, η^i, and θ^i from the corresponding priors, $i = 1, \ldots, m$. Obtain $\rho^i = \sin^{-1}\sqrt{\pi^i\eta^i + (1-\pi^i)(1-\theta^i)}$, $\rho_T^i = \sin^{-1}\sqrt{\pi_T\eta^i + (1-\pi_T)(1-\theta^i)}$, $i = 1, \ldots, m$, and obtain approximations to μ_ρ, μ_{ρ_T}, τ_ρ, and τ_{ρ_T} of $\bar{\mu}_\rho \equiv \sum_1^m \rho^i/m$, $\bar{\mu}_{\rho_T} \equiv \sum_1^m \rho_T^i/m$, $\bar{\tau}_\rho \equiv m/\sum_1^m(\rho^i - \bar{\mu}_\rho)^2$, and similarly for $\bar{\tau}_{\rho_T}$. Denote the prior means for η and θ as μ_η and μ_θ, respectively, and obtain $\bar{\delta}_0 = \pi_0\mu_\eta + (1-\pi_0)(1-\mu_\theta)$.

2. Fix n. Let $\bar{w} = \frac{4n}{4n+\bar{\tau}_\rho}$. Substitute appropriate terms with bars over them into the formula for y_{n0}, and define the result as $\bar{y}_{no}$. Substitute $\bar{\mu}_{\rho_T}$, $\bar{\tau}_{\rho_T}$ and $\bar{y}_{n0}$ into (13) and note whether the inequality is satisfied.

3. Increment n and repeat step 2 until we find the smallest n, say n_{c0} that satisfies the inequality in (13).

The alternative case is the same as the null case described above except the inequalities are reversed.

Appendix 4: Proofs of Propositions 1 and 2

Proof of Proposition 1: Let $0 < \pi_0 < 1$, fix η, θ and define $g(\pi) = \frac{\pi\eta+(1-\pi)(1-\theta)}{1-\{\pi\eta+(1-\pi)(1-\theta)\}} = \delta/(1-\delta)$ and $F_x(\pi_0) = P_n(\pi \leq \pi_0|x, \eta, \theta)$; f is the corresponding density of F_x. Note that $g'(\pi) > 0$ provided $\eta + \theta > 1$. Then

$$P_n(\pi \leq \pi_0|x, \eta, \theta) > P_n(\pi \leq \pi_0|x+1, \eta, \theta) \Leftrightarrow F_x(\pi_0) > \frac{\int_0^{\pi_0} f(\pi)g(\pi)d\pi}{\int_0^1 f(\pi)g(\pi)d\pi}$$

$$\Leftrightarrow F_x(\pi_0)\{g(1) - \int_0^1 F_x(\pi)dg(\pi)\} - F_x(\pi_0)g(\pi_0) + \int_0^{\pi_0} F_x(\pi)dg(\pi) > 0$$

$$\Leftrightarrow F_x(\pi_0)\{g(\pi_0)F_x(\pi_0) + \int_{\pi_0}^1 g(\pi)dF_x(\pi) - g(\pi_0)\} + \{1 - F_x(\pi_0)\}\int_0^{\pi_0} F_x(\pi)dg(\pi) > 0,$$

after successive integrations by parts. But since $g'(\pi) > 0$, $\int_0^{\pi_0} F_x(\pi)dg(\pi) > 0$ and $g(\pi_0)F_x(\pi_0) + \int_{\pi_0}^1 g(\pi)dF_x(\pi) - g(\pi_0) > g(\pi_0)F_x(\pi_0) + g(\pi_0)\{1 - F_x(\pi_0)\} - g(\pi_0) = 0$. A modification of this argument is used to establish the unconditional result. In the modified argument, $f(\cdot)$ becomes the joint density of (π, η, θ) given x and $g(\cdot)$ is also regarded as a function of all parameters. In the final expression above, substitute $F_x(\pi_0)$ with the corresponding marginal probability. The next term in brackets, and the final term, are replaced by the integrals of the given terms against the joint density for (η, θ) given x. The final result follows by the same reasoning as above. The second result is obtained by letting $g(\pi) = 1 - \delta$ and noticing that $g'(\pi) < 0$ provided $\eta + \theta > 1$.

Proof of Proposition 2: Observe that

$$\begin{aligned} P_{n+1}(T^+ \leq x|\pi_T, \eta, \theta) &= \sum_{y=0}^{x} \binom{n+1}{y} \delta_T^y (1-\delta_T)^{n+1-y} \\ &= \sum_{y=1}^{x} \left\{ \binom{n}{y} + \binom{n}{y-1} \right\} \delta_T^y (1-\delta_T)^{n+1-y} + (1-\delta_T)^{n+1} \\ &= P_n(T^+ \leq x|\pi_T, \eta, \theta) - \delta_T P_n(T^+ = x|\pi_T, \eta, \theta). \end{aligned}$$

The result (7) is then derived by induction and the first part of (8) is obtained integrating against the distribution for (η, θ). The second part of (8) is established using (7), that is $h^*_{n+1}(x_1) - h^*_n(x_0) =$

$$\int \{P_n(x_0 < T^+ < x_1 | \pi_T, \eta, \theta) + P_n(T^+ = x_1 | \pi_T, \eta, \theta)(1 - \delta_T)\} p(\eta, \theta) d\eta d\theta > 0.$$

	$N = 50$		$N = 265$		$N = 500$		Bin	Normal
	Bayes	CB	Bayes	CB	Bayes	CB	Bayes	Bayes
π_0	n_{ci}/x_{ci}	n_c/x_c	n_{ci}/x_{ci}	n_c/x_c	n_{ci}/x_{ci}	n_c/x_c	n_{ci}/x_{ci}	n_{ci}
	H_0 with $\pi_T = 0$							
0.10	48/4	40/2	102/7	64/3	116/8	65/3	132/9	127
0.15	35/3	29/2	51/4	36/2	52/4	37/2	53/4	61
0.20	21/2	25/2	33/3	28/2	34/3	28/2	35/3	39
0.25	19/2	16/1	22/2	17/1	22/2	17/1	22/2	28
0.30	16/2	13/1	18/2	14/1	18/2	14/1	18/2	21
	H_1 with $\pi_T = 0.3$							
0.05	17/3		18/3		23/4		23/4	22
0.10	32/7		44/9		45/9		45/9	44
0.15	38/9		78/18		91/21		109/25	96

Table 1: Sample sizes for the FMD investigation. The Bayesian method uses $p_0 = p_1 = 0.95$ and $\beta_0 = \beta_1 = 0.95$. The Cameron-Baldock (CB) criterion uses minimum expected prevalence equal to π_0 and Type I and Type II error probabilities of 0.05.

π_0	$N = 50$ $n/x_{ci}/\tilde{p}_{ni}$	$N = 265$ $n/x_{ci}/\tilde{p}_{ni}$	$N = 500$ $n/x_{ci}/\tilde{p}_{ni}$	Binomial $n/x_{ci}/\tilde{p}_{ni}$	Normal $n/\tilde{p}_{ni}$
	H_0 with $\pi_T = 0$				
0.10	50/0/0.49	265/8/0.89	350/11/0.9	339/11/0.88	253/0.9
0.20	50/0/0.8	106/5/0.95	117/5/0.95	113/5/0.95	100/0.95
	H_1 with $\pi_T = 0.2$				
0.05	50/5/0.79	225/15/0.95	227/15/0.95	250/16/0.95	163/0.95
0.10*	50/0/0.49	265/22/0.83	500/39/0.90	500/39/0.89	503/0.95
		180/15/0.80	250/21/0.80	250/21/0.80	340/0.90
		150/13/0.75	200/17/0.79	150/13/0.74	169/0.80
		100/9/0.70	100/9/0.70	123/11/0.70	90/0.70

Table 2: Sample sizes for bovine paratuberculosis. $p_i = 0.95$ for all cases except $p_1 = 0.90$ for $\pi_0 = 0.1$ and $\pi_T = 0.2$. Uniform prior for prevalence.

	D^+	D^-	Total
T^+	z_1	$x - z_1$	x
T^-	$n - x - z_2$	z_2	$n - x$

Table 3: Augmented Data Representation

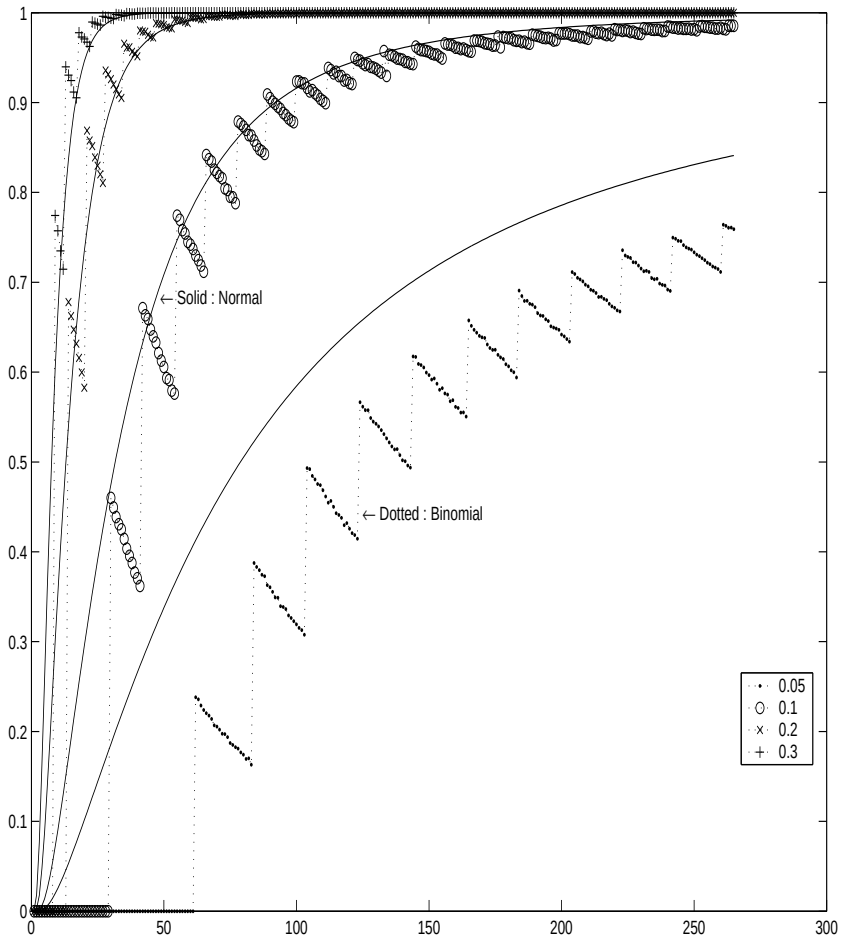

Figure 1. Sample size n (for null calculation with $\pi_T = 0$) versus corresponding $\tilde{p}_{n0}$ for $\pi_0 \in (0.05, 0.10, 0.20, 0.30)$; normal (solid line) and binomial (dotted). $p_1 = p_2 = 0.95$, $\pi \sim$ Beta(1,1), $\eta \sim$ Beta(68.74, 4.57), $\theta \sim$Beta(107.20, 3.17).

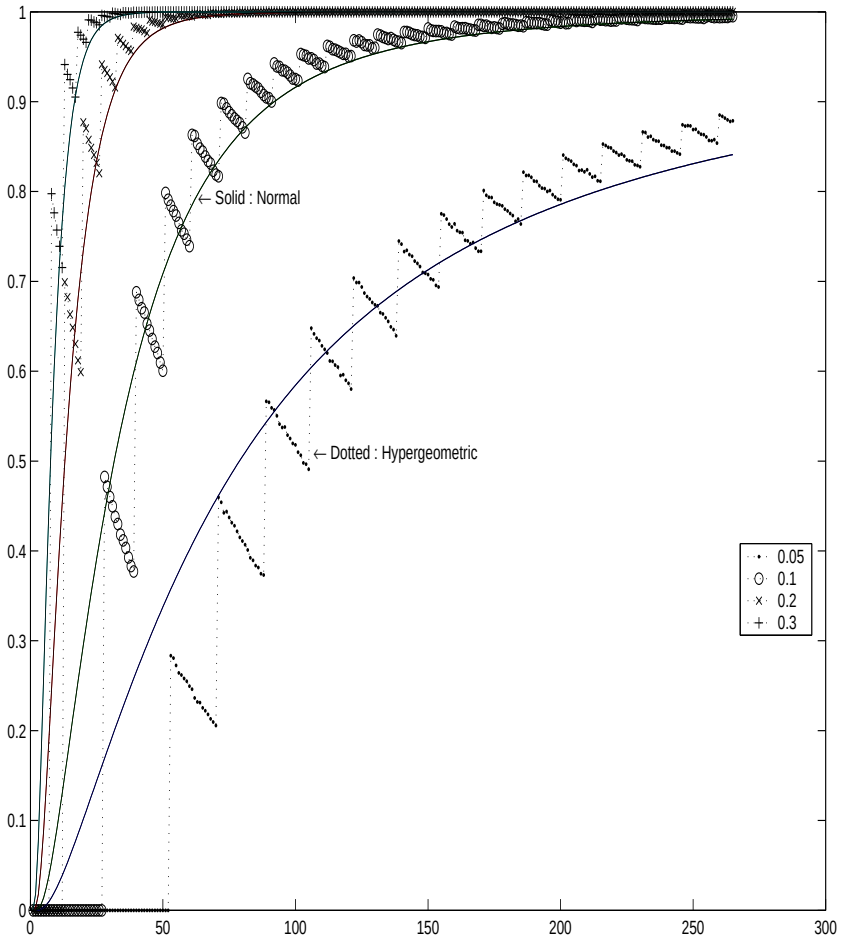

Figure 2. Normal and Hypergeometric sample size curves under the same conditions as given in Figure 1