Optimal Pooled Testing Design for Prevalence Estimation under Resource Constraints

Ngoc T. Nguyen[1*], Ebru K. Bish[2], Douglas R. Bish[2]

[1] Grado Department of Industrial and Systems Engineering

Virginia Tech, Blacksburg, VA 24061, USA

[2] Department of Information Systems, Statistics, and Management Science

The University of Alabama, Tuscaloosa, AL 35487, USA

December 15, 2020; Revised: March 31, 2021; June 18, 2021

Abstract

Accurate estimation of disease prevalence is essential for mitigation efforts. Due to limited testing resources, prevalence estimation is often conducted via pooled testing, in which multiple specimens are combined and tested via a single test. The *pool design*, i.e., the number and sizes of testing pools, has a substantial impact on estimation accuracy. Determining an optimal pool design is challenging, especially for emerging or seasonal diseases for which information on the status of the disease is unreliable or unavailable prior to testing. We develop novel optimization models for testing pool design under uncertainty and limited resources, and characterize structural properties of optimal pool designs. We apply our models to estimate the prevalence of West Nile virus in mosquitoes (the main vector of transmission to humans). Our findings suggest that estimation accuracy can be substantially improved over the status quo through the proposed optimal pool designs.

Keywords: optimal testing pool design, uncertainty, robust optimization, prevalence estimation, emerging or seasonal diseases.

*Corresponding Author: Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061. E-mail: ntn@vt.edu

1 Introduction

As the current pandemic resulting from the worldwide spread of COVID-19 continues to demonstrate, surveillance is essential for responding to emerging or seasonal diseases. For other recent examples, consider the 2016 outbreak of the Zika virus [5,35], or the 2002-2003 outbreak of the West Nile virus [24]; during both outbreaks, surveillance was crucial for response planning. Surveillance involves the activities of collection, analysis, dissemination, and response [13]. Among these components, collection and analysis activities are conducted frequently at many levels of the healthcare system, including local, state, federal, and international levels, by both public and private agencies, with a main goal of *estimating the prevalence* of the disease in question [13], which is the focus of this paper. An accurate estimate of disease prevalence is important, as it is a key input to various other surveillance activities, including outbreak detection, response and prevention evaluation, and prediction of the impact of various healthcare services (e.g., [13, 16, 58]). However, it is often the case that testing resources (e.g., budget) available for prevalence estimation is very small relative to the needs [13]. As a result, prevalence estimation via individually testing of each subject is either infeasible (e.g., [55, 56]), or highly inefficient in that it leads to small sample sizes, and hence to potentially inaccurate estimates [26, 55].

A solution to both prevalence estimation and subject identification (i.e., identification of all infected subjects) under limited resources came from Dorfman in the 1940's [15]. Dorfman's idea was to use *pooled testing*, by combining specimens (e.g., blood, urine, tissue swabs) from multiple subjects in a single testing pool and testing the pool via a single test [15]. Over the years, pooled testing has been extensively studied, and shown to be a highly efficient approach under limited resources for both prevalence estimation and subject identification problems, and today it is a widely used testing method for both purposes (e.g., [16, 28, 45, 52, 58]). When pooled testing is used for the purpose of prevalence estimation, it often involves testing of the pools only (i.e., without any follow-up testing on individual subjects in positive-testing pools), within a single-stage pooled testing process, as the goal is to derive an accurate estimate of the disease prevalence rate (e.g., [23, 28, 37, 38]). While sequential pooling strategies, i.e., multi-stage pooled testing processes with stage-dependent pool sizes, are possible (e.g., [41, 58, 61]), such strategies are more difficult to implement in practice. As a result, the simple, single-stage pooled testing strategy described

above is commonly studied and used for prevalence estimation. This is especially true when the goal is to estimate the prevalence of "sources" of vector-borne viral or bacterial diseases, e.g., mosquitoes carrying Zika virus or West Nile virus (e.g., [44]) or romaine lettuce carrying *E. coli* bacteria (e.g., [6]). Because one of our objectives is to provide actionable and practical insights for decision-makers, we focus on this single-stage pooling strategy throughout the paper, and propose the study of sequential pooling strategies, and their potential benefit, for prevalence estimation as a future research direction in Section 6.

The test measures the pool's concentration of a certain bio-marker, which serves as an indicator for the presence of the virus or bacteria of interest. The pool's bio-marker concentration is then compared with a preset bio-marker threshold, to produce a binary test outcome: *positive* if the pool's bio-marker concentration equals or exceeds the bio-marker threshold, suggesting the presence of at least one infected specimen in the pool, and *negative* if the pool's bio-marker concentration is lower than the bio-marker threshold, suggesting that all specimens in the pool are infection-free; and inference on the unknown prevalence rate is made based on the testing data. Thus, the test's sensitivity (true positive probability) and specificity (true negative probability) depend on both the number of infected specimens in a pool, and the bio-marker concentration of each infected specimen, where the latter depends on the stage of the infection [40, 41]. In this paper, we make a simplifying assumption, common in the pooling literature, that the test is perfectly reliable, that is, it has perfect sensitivity and specificity. In this case, the test outcome will be positive if there is at least one infected specimen in the pool, and the test outcome will be negative otherwise; we suggest the relaxation of this assumption as a future research direction in Section 6.

The efficiency and effectiveness of the estimation process depend critically on testing *pool design*, that is, the *number of pools* to test and the *pool size* (i.e., the number of specimens to combine in each pool) (e.g., [7, 41, 49, 53]). Determining an optimal testing pool design is difficult, because there is often limited, and highly uncertain, information on the status and dynamics of a disease prior to a surveillance study, especially for emerging or seasonal diseases, but an initial estimate on disease prevalence is a key input to pool design. Further, research that develops optimal pool designs for prevalence estimation is quite limited. The majority of the relevant literature focuses on the "estimation" component, i.e., derivation of an accurate prevalence estimate from testing data, for a *given* pool design. This stream of research includes studies that investigate the characteristics

of the widely used maximum likelihood estimator (MLE) of the prevalence rate in various settings (e.g., [8,9,53]), and develop various approaches for bias reduction in the MLE (e.g., [3,19,26]), as well as studies that investigate alternative methods for deriving an estimator, such as Bayesian analyses (e.g., [17, 43, 51]), or regression analyses (e.g., [10, 18, 27, 57, 60]). On the other hand, only a few studies discuss the pool design component, and mainly in the context of pool size determination for a *fixed (exogenous)* number of testing pools, under both perfect tests (e.g., [29,30,46]), and imperfect tests (e.g., [20, 37, 55, 56, 62]). However, the aforementioned studies on pool size determination are mainly numerical in nature, and the optimal pool size is not fully characterized in an analytical manner, with the notable exception of the work by Liu et al. [37], which derives various properties of the asymptotic variance function (a commonly used metric for pool design). While modeling and optimization efforts for testing pool design are lacking, it has been shown that a pool design that relies highly on an initial point estimate of the prevalence rate, or that corresponds to an exogenously fixed number of testing pools, can result in highly inaccurate estimates of the prevalence rate [29, 30, 41].

Motivated by these gaps in the literature, in this paper we develop and study novel models for testing pool design under uncertainty and limited resources (e.g., testing budget). Specifically, we study the pool design problem in two settings: (i) the number of testing pools is fixed exogenously and cannot be altered (*single-variable pool design problem*), and (ii) the tester has control over both the number of pools and the pool size (*joint pool design problem*). The first setting applies, for example, when there is a limited number of testing kits, which can be an important constraint in disease testing (e.g., [1, 2, 37, 48]), and hence the tester can only control the pool size. Almost all existing studies on pool design for prevalence estimation focus on this first setting (i.e., single-variable pool design problem), as discussed above. Thus, our study of the single-variable pool design problem is important in its own right, and provides a contribution to this literature in that this paper is the first to analytically characterize the optimal solution to the single-variable pool design problem. In addition, this paper contributes to the existing literature on pooled testing for prevalence estimation through introducing novel models, including the joint pool design problem, and novel, robust variations of both the single-variable and the joint pool design problems. These novel formulations are important in practical applications, because they do not require the distributional information of the prevalence rate, but only its support. We establish key structural

properties of optimal pool designs for all models, which allow us to analytically characterize the form of an optimal pool design and obtain an optimal pool design in a highly efficient manner in each setting. Not surprisingly, our study of the joint pool design problem also indicates that a joint optimization of both the pool size and the number of pools can provide substantial benefit, providing a more accurate estimate at the same testing budget. We complement our analytical results with a case study on the prevalence estimation of West Nile virus in mosquitoes. This case study illustrates that a joint optimization of both the pool size and the number of pools can provide substantial benefit, providing a more accurate estimate at the same testing budget; and the use of robust models, and especially the robust version of the joint pool design problem, helps correct inaccuracies in input parameters, further improving estimation accuracy, again without requiring an increase in the testing budget.

The remainder of this paper is organized as follows. Section 2 presents the notation and modeling assumptions, and formulates the pool design optimization models. Then, Section 3 establishes key structural properties of pool design optimization models, Section 4 provides analytical characterization of optimal pool designs, and Section 5 demonstrates the benefits of optimal pool designs through the West Nile virus case study. Finally, Section 6 concludes with a discussion of our findings and suggestions for future research. To facilitate the presentation, some supporting results and all mathematical proofs are relegated to the Appendix.

2 Notation, Assumptions, and Models

Section 2.1 introduces the notation and discusses the modeling assumptions; and Section 2.2 presents pool design optimization models. A summary of notation can be found in Appendix A.

2.1 Notation and Assumptions

Throughout, we denote random variables in upper-case letters and their realization in lower-case letters; and use ";" to denote probabilistic conditioning, or an exogenous parameter.

Consider a disease with prevalence rate, p, which needs to be estimated via pooled testing. The tester models the unknown prevalence rate as continuous random variable P, whose moments

and support are unknown to the tester. We develop distribution-free models for pool design optimization, that is, our models do not require the distributional form of P, and solely rely on either its point estimate, denoted by p_0 (*deterministic pool design*), or estimated support, denoted by $[p_{LB}, p_{UB}]$ (*robust pool design*). We study the *pool design problem* in two settings: the setting in which the tester needs to determine both the pool size, m, and the number of pools to test, n (*joint pool design problem – Problem J)*; and the setting in which the tester needs to determine only the pool size, m, for a *given* number of pools (*single-variable pool design (pool size) problem – Problem S)*. In both settings, the tester has a limited testing budget, and wishes to obtain an accurate estimate of the unknown prevalence rate, i.e., to minimize the asymptotic variance of the estimator, as we discuss below. While the optimal design generated by *Problem J* (i.e., through joint optimization) is clearly more desirable, *Problem S* continues to apply in practice when the number of testing kits is pre-determined and limited (e.g., [1, 2, 37, 48]), and our study of *Problem S* offers practical relevance in those cases.

Testing incurs a fixed testing cost (e.g., cost of the testing kit), of c_f per pool, and a variable cost (e.g., specimen collection and preparation cost), of c_v per specimen tested, with $c_f > c_v$, and the tester has a testing budget of B. While these costs are typically test- and disease-dependent, in general, the fixed testing cost c_f is much larger than the variable cost c_v. For example, considering the West Nile virus, for the reverse-transcription polymerase chain reaction (RT-PCR) assay used in our case study, c_f is around \$72 per test, while c_v is around \$4 per specimen (based on [25]); while for an antibody assay, the fixed and variable costs are around \$1.35 and \$0.04, respectively (e.g., [41, 61]). Then, in *Problem J*, the feasible set for decision variables m and n is given by, $\mathbb{F}(m, n) \equiv \{m, n \in \mathbb{Z}^+ : c_f n + c_v mn \leq B\}$, while in *Problem S*, the feasible set for the single decision variable m, for a *given* value of $n \in \mathbb{Z}^+$, is given by, $\mathbb{F}(m; n) \equiv \left\{m \in \mathbb{Z}^+ : m \leq \overline{M}^S(n) \equiv \left\lfloor \frac{B - c_f n}{c_v n} \right\rfloor \right\}$.

The objective is to design testing pools so as to minimize the asymptotic variance of the estimator, which is a commonly used objective for both pool design (e.g., [30, 37, 54, 56, 57]), and for assessing the accuracy of an estimator (e.g., [36, 50]). The asymptotic variance, $\sigma^2(m, n; p)$, corresponding to a *true* prevalence rate p, represents the limiting behavior of the mean squared error (MSE) (i.e., variance plus bias square) of an estimator $\hat{P}$, $MSE(\hat{P}, m, n; p)$, as the number of pools, n, becomes large, and is commonly used because, in general, the MSE is analytically intractable (e.g., [30, 31, 37]). The asymptotic variance also provides a strong lower bound (i.e.,

through the Cramer-Rao lower bound) on the Fisher's information obtained from the prevalence estimate (e.g., [4]), and hence is also utilized in the pool design literature when the objective is to maximize the Fisher's information, because obtaining an explicit analytical expression for Fisher's information often proves to be intractable (e.g., [29, 58]). Following common terminology, we refer to a pool design as *efficient* if it minimizes the asymptotic variance (e.g., [29, 30, 58]).

On the testing side, we consider a test that can be applied to pools of specimens collected from subjects (i.e., $m \geq 2$) as well as to individual specimens (i.e., $m = 1$). We assume that the test is perfectly reliable and provides a binary outcome, that is, the test has perfect *sensitivity* and *specificity*, thus providing a *positive* outcome only if there is at least one true-positive specimen in the pool, and a *negative* outcome only if all specimens in the pool are true-negative. This is a reasonable assumption, especially for assays that utilize the nucleic acid amplification technology for detecting the viral RNA in specimens, see, e.g., [42], and we consider such a test, namely the the reverse-transcription polymerase chain reaction (RT-PCR) assay, in our case study (Section **).

We estimate the unknown prevalence rate via the commonly adopted maximum likelihood estimator (MLE) (e.g., [30]), given by:

$$\hat{P} = 1 - \left(1 - \frac{T(m, n; p)}{n}\right)^{\frac{1}{m}}, \tag{1}$$

where $T(m, n; p)$ denotes the random number of positive-testing pools among n pools, each containing m specimens, that is, for $m = 1$ (i.e., individual testing), $T(1, n; p) \sim \text{Binomial}(n, p)$, while for $m \geq 2$ (i.e., pooled testing), $T(m, n; p) \sim \text{Binomial}(n, 1-(1-p)^m)$, given a true prevalence rate of p (i.e., a realization of random variable P), and the term, $1-(1-p)^m$, denotes the probability that a random pool tests positive, i.e., the probability that it contains at least one true-positive specimen. As n becomes large, the bias incurred by $\hat{P}$ becomes negligible, and only its variance remains, i.e., the asymptotic variance of $\hat{P}$ represents the asymptotic behavior of the MSE of $\hat{P}$. The asymptotic variance of $\hat{P}$, given a true prevalence rate p, then follows (e.g., [30]):

$$\sigma^2(m, n; p) = \frac{1-(1-p)^m}{nm^2(1-p)^{m-2}} \ . \tag{2}$$

Note that for the special case of individual testing, we have that $\sigma^2(1,n;p) = \frac{p(1-p)}{n}$. A detailed derivation of $\sigma^2(m,n;p)$ is given in [30].

2.2 Models for Pool Design Optimization

We formulate and study a deterministic optimization model, which relies only on a point estimate of P, given by p_0; and a robust (mini-max) optimization model, which relies only on an estimated support of P, given by $[p_{LB}, p_{UB}]$, for both the single-variable and joint pool design problem settings. Thus, the only input needed for the robust models in both the single-variable and joint pool design problem settings is the support of P, i.e., $[p_{LB}, p_{UB}]$; not its distribution nor its point estimate, both of which can be quite difficult to estimate in practice. This is a highly desirable property of the robust models, facilitating their implementation in practice, especially for emerging or highly dynamic diseases. Further, both robust models perform quite well, as we show subsequently in the case study.

We use model indices D and M to respectively refer to the deterministic and robust models, followed by problem indices S and J to respectively refer to the single-variable and joint pool design problems, e.g., *D-S* denotes the deterministic *Problem S*. We also use the notation *D-X* and *M-X*, $X \in \{S, J\}$, when an expression or a result applies to both *Problems S* and *J*. Model formulations follow:

Joint Pool Design Models:

Deterministic (*D-J*) Model:

$$\underset{(m,n)\in\mathbb{F}(m,n)}{\text{minimize}} \quad \sigma^2(m,n;p_0)$$

Robust (*M-J*) Model:

$$\underset{(m,n)\in\mathbb{F}(m,n)}{\text{minimize}} \quad \left\{\max_{p\in[p_{LB},p_{UB}]} \left\{\sigma^2(m,n;p)\right\}\right\}$$

Single-variable Pool Design Models:

Deterministic (*D-S*) Model:

$$\underset{m\in\mathbb{F}(m;n)}{\text{minimize}} \quad \sigma^2(m;n,p_0)$$

Robust (*M-S*) Model:

$$\underset{m\in\mathbb{F}(m;n)}{\text{minimize}} \quad \left\{\max_{p\in[p_{LB},p_{UB}]} \left\{\sigma^2(m,n;p)\right\}\right\}$$

We use the superscript * to denote an optimal solution, e.g., (m^*_{D-J}, n^*_{D-J}) denotes the optimal solution to the *D-J* Model.

To our knowledge, only the *D-S* Model is utilized in the existing literature, i.e., for a given number of pools, n, under different objective functions, including the minimization of the asymptotic variance [30,37], maximization of the Fisher's information via the Cramer-Rao lower bound, which

reduces to a function of the asymptotic variance [29, 61], or maximization of the probability that a random pool tests positive [55], but research that establishes the model's structural properties and characterizes its optimal solution is limited, with the notable exception of Liu et al. (2011) [37], as we detail below. Thus, our analysis of the *D-S* Model provides a contribution to the literature in its own right; and the *D-J*, *M-S*, and M-J Models that we study are novel. In particular, we build upon the properties developed by Liu et al. (2011) [37], while considering perfect tests ([37] also considers imperfect tests). However, instead of assuming a given number of pools (n), or a given number of specimens tested (mn), i.e., using single-variable optimization as in [37], we also jointly optimize over both the pool size, m, and the number of pools, n, while explicitly accounting for the budget constraint. Further, our formulation of the robust models in both settings is novel, as discussed above. We fully characterize the optimal solutions to all models by establishing new structural properties of optimal pool designs.

3 Properties of the Asymptotic Variance Function

In this section, we establish key properties of the asymptotic variance function, $\sigma^2(m, n; p)$. In particular, we first study the setting with a fixed (exogenous) n value (Section 3.1), followed by the setting where n is optimally set (endogenous) (Section 3.2). All mathematical proofs can be found in Appendix C.

3.1 Asymptotic Variance Function for an Exogenous Number of Pools

The single-variable pool design problem *Problem S*, i.e., with an exogenous number of pools, n, has received some attention in the literature. We first provide the relevant definitions and the properties previously established in the literature.

Definition 1. (From [37]) For any $n, m_1, m_2 \in \mathbb{Z}^+$: $m_1 < m_2$, the *prevalence threshold*, $\pi_0^S(m_1, m_2, n) \in (0, 1)$, is defined as the prevalence rate at which $\sigma^2\big(m_1, n; \pi_0^S(m_1, m_2, n)\big) = \sigma^2\big(m_2, n; \pi_0^S(m_1, m_2, n)\big)$.

By this definition, the prevalence threshold for any pair of pool sizes, m_1 and m_2, is the prevalence rate for which the asymptotic variances at these pool sizes are equal for a given number of pools, n, as we discuss in more detail following Proposition 1. For the exogenous n setting, Liu et al

(2011) [37] provides the following property on the prevalence threshold, $\pi_0^S(m_1, m_2, n)$, and studies the behavior of the asymptotic variance function mainly through numerical studies.

Proposition 1. (From [37]) For any $n, m_1, m_2 \in \mathbb{Z}^+$: $m_1 < m_2$, there exists a unique $\pi_0^S(m_1, m_2, n) \in (0, 1)$. Further:

$$\sigma^2(m_1, n; p) \begin{cases} > \sigma^2(m_2, n; p), & \forall p < \pi_0^S(m_1, m_2, n) \\ < \sigma^2(m_2, n; p), & \forall p > \pi_0^S(m_1, m_2, n) \end{cases}.$$

Proposition 1 illustrates one of the key properties of the prevalence threshold. In particular, as discussed in [37], Proposition 1 implies that, for a given n, a smaller pool (of size m_1) is more efficient than a larger pool (of size m_2), in terms of minimizing the asymptotic variance, only if the prevalence rate p is sufficiently high, in comparison to the prevalence threshold, i.e., $p > \pi_0^S(m_1, m_2, n)$, and vice versa.

In the following, we establish various other properties of the asymptotic variance function, $\sigma^2(m, n; p)$, and the prevalence threshold, $\pi_0^S(m_1, m_2, n)$, for a given n. These properties not only enable us to solve the single-variable pool design problem, i.e., *Problem S*, to optimality but also confirm the numerical observations by Liu et al. (2011) [37], further contributing to the literature on the single-variable pool design problem for prevalence estimation. In particular, Liu et al. (2011) [37] numerically observes that when n is exogenously fixed, the prevalence threshold below which pooled testing (i.e., with $m \geq 2$) is more efficient than individual testing (i.e., with $m = 1$), given by $\pi_0^S(1, m, n)$, decreases in m. In Lemma 1, we formally establish this result for perfect tests, for a general case with any $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$.

Recall that $\sigma^2(m, n; p) = \left(\frac{1}{n}\right)\left(\frac{1-(1-p)^m}{m^2(1-p)^{m-2}}\right)$ (Eqn. (2)). Then, by Definition 1, when n is fixed, for any $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$, the prevalence threshold $\pi_0^S(m_1, m_2, n)$ equals the prevalence rate p for which $\sigma^2(m_1, n; p) = \sigma^2(m_2, n; p)$, equivalently, it is the value of p for which $\left(\frac{1}{n}\right)\left(\frac{1-(1-p)^{m_1}}{m_1^2(1-p)^{m_1-2}}\right) = \left(\frac{1}{n}\right)\left(\frac{1-(1-p)^{m_2}}{m_2^2(1-p)^{m_2-2}}\right)$, or $\frac{1-(1-p)^{m_1}}{m_1^2(1-p)^{m_1-2}} = \frac{1-(1-p)^{m_2}}{m_2^2(1-p)^{m_2-2}}$, that is, the expression of $\pi_0^S(m_1, m_2, n)$ becomes only a function of m_1 and m_2, and is independent of n. Therefore, dropping n, we simply use the notation, $\pi_0^S(m_1, m_2)$. Thus, all subsequent results in this section hold for any $n \in \mathbb{Z}^+$. We first provide an analytical expression for the prevalence threshold function, which allows us to not only determine an optimal solution, but also study the prevalence threshold below which pooled testing is always more efficient than individual testing.

Lemma 1. 1. For any $m_1, m_2 \in \mathbb{Z}^+$: $m_1 < m_2$, $\pi_0^S(m_1, m_2)$ is the unique solution to:

$$\left(\frac{1}{1-\pi_0^S(m_1,m_2)}\right)^{m_1} = 1 + \left(\frac{m_1}{m_2}\right)^2 \left[\left(\frac{1}{1-\pi_0^S(m_1,m_2)}\right)^{m_2} - 1\right]. \tag{3}$$

2. $\pi_0^S(m_1, m_2)$ decreases as m_1 or m_2 increases, for $2 \leq m_1 < m_2$.

3. $\pi_0^S(m-1, m) > \pi_0^S(m, m+1)$, $\forall m \in \mathbb{Z}^+ : m \geq 2$.

Remark 1. Lemma 1 establishes the necessary and sufficient condition for individual testing to be the most efficient method for prevalence estimation, that is, individual testing outperforms pooled testing, with any pool size and for any given number of pools, i.e., $\sigma^2(1, n; p) < \sigma^2(m, n; p), \forall m \geq 2$, $\forall n \in \mathbb{Z}^+$, if and only if, $p > \pi_0^S(1, 2) = \frac{2}{3}$.

Lemma 1 further indicates that as pool size, m, increases, the prevalence threshold below which pooled testing is more efficient than individual testing reduces, that is, larger pools are particularly efficient compared to individual testing for smaller prevalence rates. Lemma 1 also analytically establishes the behavior of $\pi_0^S(1, m)$, numerically observed in [37], as discussed above. Prevalence rates of almost all diseases are well below the threshold of $\frac{2}{3}$, and pooled testing is a commonly used method for disease surveillance. Thus, Lemma 1 provides an analytical justification for the widespread utilization of pooled testing for prevalence estimation of diseases.

Next, we demonstrate the behavior of the prevalence threshold function, π_0^S, in pool sizes, which is analytically established in Lemma 1. To this end, Figure 1 plots the asymptotic variance function, $\sigma^2(m, n; p)$, for various pool sizes m, and number of pools $n = 1$, and shows the prevalence threshold, π_0^S, for various pairs of pool sizes. Because $\sigma^2(m, n; p) = \left(\frac{1}{n}\right)\left(\frac{1-(1-p)^m}{m^2(1-p)^{m-2}}\right)$ (Eqn. (2)), the number of pools, n, alters each asymptotic variance function through a common multiplicative factor, $1/n$, in the single-variable pool design setting. Hence, the relationship between the asymptotic variance functions at different pool sizes remain unchanged for different values of $n \in Z^+$, and $\pi_0^S(m_1, m_2)$ becomes independent of n, as discussed above. By Definition 1, $\pi_0^S(m_1, m_2)$ corresponds to the prevalence rate p for which $\sigma^2(m_1, n; p) = \sigma^2(m_2, n; p)$ (i.e., their intersection point, which is unique [37]), see $\pi_0^S(1, 2), \pi_0^S(1, 3)$, and $\pi_0^S(2, 3)$, marked on the figure. In summary, Figure 1 confirms that $\pi_0^S(m_1, m_2)$ decreases as m_1 or m_2 increases (e.g., $\pi_0^S(1, 2) > \pi_0^S(1, 3)$ and $\pi_0^S(1, 3) > \pi_0^S(2, 3)$), and $\pi_0^S(m-1, m) > \pi_0^S(m, m+1)$, $\forall m \geq 2$ (e.g., $\pi_0^S(1, 2) > \pi_0^S(2, 3)$), in accordance with Lemma 1.

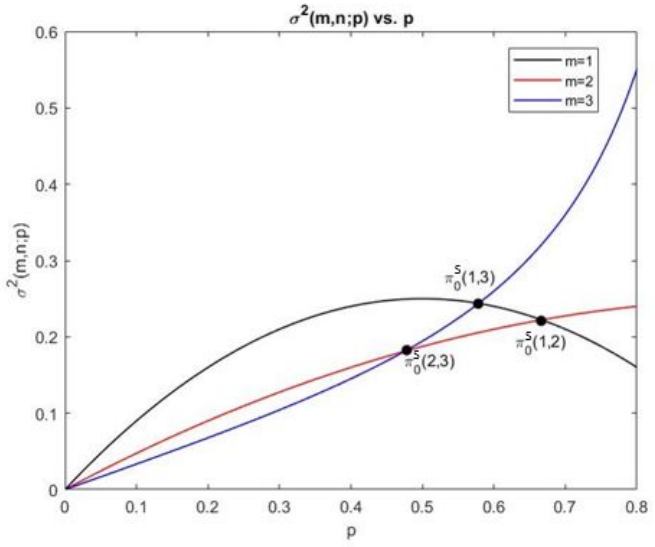

Figure 1: The asymptotic variance function, $\sigma^2(m, n; p)$ for $m = 1, 2, 3$ and $n = 1$, and the prevalence threshold, $\pi_0^S(m_1, m_2)$, for $(m_1, m_2) = (1, 2), (1, 3)$, and $(2, 3)$

We also establish several key properties of the asymptotic variance function to gain intuition. In particular, we show that the asymptotic variance increases as the prevalence rate, p, increases; and decreases as the pool size, n, increases. Further, the asymptotic variance is strictly convex in pool size, m, indicating that an optimal pool size, i.e., one that minimizes the asymptotic variance, will be neither too small nor too large (see Lemma 6 in Appendix B). In addition to providing insight, the results in this section characterize the structure of the prevalence threshold and asymptotic variance functions, allowing us to determine an optimal pool size for *Problem S* in Section 4.

3.2 Asymptotic Variance Function for an Endogenous Number of Pools

Next we study properties of the asymptotic variance function in the joint pool design problem, *Problem J*, i.e., when the tester can set both n and m optimally. To this end, we first derive an expression on the optimal number of testing pools, n^*, in *Problem J* by relaxing the restriction that n is integer, i.e., we consider that n is continuous. (We discuss how to incorporate the integrality constraint on n in Section 5.)

Lemma 2. Consider that n is continuous and non-negative. Then, without loss of optimality, the feasible region for the joint models, *D-J* and *M-J*, given by $\mathbb{F}(m, n)$, can be replaced by $\mathbb{F}(m) =$

$\left\{m \in \mathbb{Z}^+ : m \leq \overline{M}^J \equiv \left\lfloor \frac{B-c_f}{c_v} \right\rfloor\right\}$, with $n^*(m) \equiv \frac{B}{c_f+c_v m}$, and

$$\sigma^2(m, n^*(m); p) = \left(\frac{c_f + c_v m}{B}\right)\left(\frac{1-(1-p)^m}{m^2(1-p)^{m-2}}\right). \tag{4}$$

Thus, each joint model reduces to one with a single decision variable, m, and its asymptotic variance function reduces to a function of m and p only, i.e., $\sigma^2(m, n^*(m); p)$.

We next show that many of the properties established for *Problem S* (Section 3.1) extend to *Problem J*. To this end, we first define the prevalence threshold in the joint setting, which, based on Lemma 2, can be represented as a function of pool sizes only.

Definition 2. For any $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$, the *prevalence threshold*, $\pi_0^J(m_1, m_2) \in (0, 1)$, is defined as the prevalence rate at which $\sigma^2\big(m_1, n^*(m_1); \pi_0^J(m_1, m_2)\big) = \sigma^2\big(m_2, n^*(m_2); \pi_0^J(m_1, m_2)\big)$.

Proposition 2. For any $m_1, m_2 \in \mathbb{Z}^+$: $m_1 < m_2$, there exists a unique $\pi_0^J(m_1, m_2) \in (0, 1)$. Further:

$$\sigma^2(m_1, n^*(m_1); p) \begin{cases} > \sigma^2(m_2, n^*(m_2); p), & \forall p < \pi_0^J(m_1, m_2) \\ < \sigma^2(m_2, n^*(m_2); p), & \forall p > \pi_0^J(m_1, m_2) \end{cases}.$$

Lemma 3. 1. For any $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$, $\pi_0^J(m_1, m_2)$ is the unique solution to:

$$\left(\frac{1}{1-\pi_0^J(m_1,m_2)}\right)^{m_1} = 1 + \left(\frac{m_1}{m_2}\right)^2\left(\frac{c_f + c_v m_2}{c_f + c_v m_1}\right)\left[\left(\frac{1}{1-\pi_0^J(m_1,m_2)}\right)^{m_2} - 1\right]. \tag{5}$$

2. $\pi_0^J(m_1, m_2)$ decreases as m_1 or m_2 increases, for $2 \leq m_1 < m_2$.

3. $\pi_0^J(m-1, m) > \pi_0^J(m, m+1)$, $\forall m \in \mathbb{Z}^+ : m \geq 2$.

Thus, similar to the exogenous number of pools setting (Lemma 1), larger pools remain particularly efficient compared to individual testing for smaller prevalence rates in the more general setting where the decision maker can set both the number of pools and the pool size. Our characterization of the prevalence thresholds for the single-variable and joint pool size problems (Lemmas 1 and 3) allows us to study the impact of joint optimization on pool design.

Lemma 4. For any $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$, we have that $\pi_0^J(m_1, m_2) < \pi_0^S(m_1, m_2)$, with the special case that $\pi_0^J(1, 2) = \frac{2c_f}{3c_f+2c_v} < \pi_0^S(1, 2) = \frac{2}{3}$.

Thus, for any given pair of pool sizes, $m_1 < m_2$, a smaller pool size (m_1) is more efficient than a larger pool size (m_2) for a wider range of prevalence values in *Problem J*, compared to *Problem S*. Under a testing budget, this in turn allows a larger number of pools to be tested in *Problem J*, increasing the efficiency of the estimation.

Similar to the exogenous number of pools setting, we show that the asymptotic variance increases as the prevalence rate, p, increases; decreases as the pool size, n, increases; and is strictly convex in the pool size, m, in the more general setting where the decision maker can set both the number of pools and the pool size. Thus, an optimal pool size m will be neither too small nor too large. Further, given a sufficiently large budget, the optimal number of testing pools will be as large as possible to reduce the asymptotic variance (see Lemma 7 in Appendix B).

4 Pool Design Optimization

With the properties established in Section 3, we are ready to characterize the optimal solutions to the single-variable and joint pool design models. In this section, we limit our analysis to the case where p does not exceed $\frac{1}{2}$; this is the most realistic case for disease surveillance studies, as discussed above.

In the following, we fully characterize the optimal solutions to all the single-variable and joint pool design models, i.e., *D-X* and *M-X*, $X \in \{S, J\}$. Recall that the testing budget constraint imposes an upper bound on pool size m, i.e., $m \leq \overline{M}^S(n) \equiv \left\lfloor \frac{B - c_f n}{c_v n} \right\rfloor$ for *D-S* and *M-S*, and $m \leq \overline{M}^J \equiv \left\lfloor \frac{B - c_f}{c_v} \right\rfloor$ for *D-J* and *M-J*. In *Problem S*, because the number of pools, n, is exogenously fixed, the optimal pool size, m_S^*, depends on n only through the upper bound, $\overline{M}^S(n)$.

The optimal solutions to *D-S* and *D-J* are respectively established by the following theorem.

Theorem 1. For a given p_0 and an exogenously fixed n such that $n \leq \left\lfloor \frac{B}{c_f + c_v} \right\rfloor$, an optimal solution

to the *D-S* Model follows a threshold policy:

$$m^*_{D-S}(p_0) = \begin{cases} \overline{M}^S(n), & \text{if} \quad p_0 \leq \pi_0^S(\overline{M}^S(n)-1, \overline{M}^S(n)) \\ \vdots & \\ m+1, & \text{if} \quad \pi_0^S(m+1,m+2) \leq p_0 \leq \pi_0^S(m,m+1) \\ m, & \text{if} \quad \pi_0^S(m,m+1) \leq p_0 \leq \pi_0^S(m-1,m) \\ m-1, & \text{if} \quad \pi_0^S(m-1,m) \leq p_0 \leq \pi_0^S(m-2,m-1) \\ \vdots & \\ 1, & \text{if} \quad \pi_0^S(1,2) \leq p_0 < 1, \end{cases}$$

where $\pi_0^S(1,2) = \frac{2}{3}$.

Similarly, for a given p_0, an optimal solution to the *D-J* Model follows a threshold policy:

$$m^*_{D-J}(p_0) = \begin{cases} \overline{M}^J, & \text{if} \quad p_0 \leq \pi_0^J(\overline{M}^J-1, \overline{M}^J) \\ \vdots & \\ m+1, & \text{if} \quad \pi_0^J(m+1,m+2) \leq p_0 \leq \pi_0^J(m,m+1) \\ m, & \text{if} \quad \pi_0^J(m,m+1) \leq p_0 \leq \pi_0^J(m-1,m) \\ m-1, & \text{if} \quad \pi_0^J(m-1,m) \leq p_0 \leq \pi_0^J(m-2,m-1) \\ \vdots & \\ 1, & \text{if} \quad \pi_0^J(1,2) \leq p_0 < 1, \end{cases}$$

where $\pi_0^J(1,2) = \frac{2c_f}{3c_f+2c_v}$, and $n^*_J(p_0)$ follows from Lemma 2, that is, $n^*_J(p_0) = \frac{B}{c_f+c_v m^*_J(p_0)}$.

Thus, the optimal *D-X*, $X \in \{S, J\}$, solution is unique if the tester's initial point estimate, p_0, does not correspond to a prevalence threshold point ($p_0 \neq \pi_0^X(m, m+1)$, $\forall m \in \mathbb{Z}^+$); and there are dual optimal solutions if p_0 corresponds to a prevalence threshold point. Further, individual testing, i.e., a pool size of one, becomes optimal if the prevalence rate p_0 is sufficiently high ($p_0 \geq \pi_0^S(1,2)$ and $p_0 \geq \pi_0^J(1,2)$ for the single-variable and joint pool design settings, respectively), and the

optimal pool size increases as the prevalence rate p_0 decreases (see the last part of Lemmas 1 and 3), that is, larger pools are particularly efficient at smaller prevalence rates.

Using the convexity of the asymptotic variance function leads to the following alternative characterization for the optimal solutions to the single-variable and joint pool design problems.

Lemma 5. For the case where $p < 1/3$, consider *D-X*, $X \in \{S, J\}$, and let m' denote the solution to the first-order condition (FOC), that is, $\left\{m' : \frac{\partial}{\partial m}\sigma^2(m, n; p_0)\big|_{m=m'} = 0\right\}$ in *Problem S*, and $\left\{m' : \frac{\partial}{\partial m}\sigma^2(m, n^*(m); p_0)\big|_{m=m'} = 0\right\}$ in *Problem J*, or equivalently,

$$m' : \begin{cases} m' = \frac{2[(1-p_0)^{m'}-1]}{\log(1-p_0)}, & \text{for } \textit{Problem S}; \\ m' = \frac{\left(1+\frac{c_f}{c_f+c_v m'}\right)[(1-p_0)^{m'}-1]}{\log(1-p_0)}, & \text{for } \textit{Problem J}. \end{cases}$$

Then $m^*_{D-X} \in \left\{m', \lceil m' \rceil, \lfloor m' \rfloor, \overline{M}^X\right\}$.

Next, from Lemmas 6 and 7, we have the following characterization of the optimal solutions to the robust models.

Theorem 2. For any given $n \in \mathbb{Z}^+ : n \leq \left\lfloor \frac{B}{c_f+c_v} \right\rfloor$, Model *M-X*, $X \in \{S, J\}$, can be equivalently formulated as its corresponding Model *D-X*, with $p_0 = p_{UB}$; and for $n > \left\lfloor \frac{B}{c_f+c_v} \right\rfloor$, no feasible solution to *M-X* exists.

As input, the robust models, *M-X*, $X \in \{S, J\}$ need only the support, $[p_{LB}, p_{UB}]$, of the prevalence rate random variable, P, and not its distribution. Theorem 2 further states that it is sufficient to provide only the upper bound of the support of P, and not the lower bound. This further simplifies the input estimation process for the robust models, further facilitating their implementation. Theorems 1 and 2, along with Lemmas 6 and 7, lead to the following results.

Corollary 1. For $X \in \{S, J\}$:

1. *D-X* optimal solution, $m^*_{D-X}(p_0)$, is non-increasing as p_0 increases, with $m^*_{D-X}(p_0) = 1$, $\forall p_0 \geq \pi_0^X(1, 2)$.

2. *M-X* optimal solution, $m^*_{D-X}(p_{UB})$, is non-increasing as p_{UB} increases, with $m^*_{M-X}(p_{UB}) = 1$, $\forall p_{UB} \geq \pi_0^X(1, 2)$.

Thus, it becomes more efficient to test smaller pools as p_0 and p_{UB} increase. This property, that pools should get smaller at larger prevalence rates, is so as to gather some "useful" information from testing. In particular, if pools are very large when the disease prevalence is high, then it is likely that many pools will test positive (i.e., contain at least one infected specimen); and similarly, if pools are very small when the disease prevalence is low, then it is likely that many pools will test negative (i.e., will not contain any infected specimen); and in either case, the tester will not obtain much information on the prevalence rate. Thus, an optimal pool design balances these risks, and attempts to gather some useful information from testing. However, when p_0 is a poor estimate of the true p value, $m^*_{D-S}(p_0)$ can still be highly inefficient, and, further, an outcome of all positive-testing pools or all negative-testing pools may still occur, as we discuss in Section 5.

Finally, we also study how the testing cost structure affects the optimal solutions to *D-X* and *M-X*, $X \in \{S, J\}$.

Corollary 2. Let $\gamma = \frac{c_f}{c_v}$. Then, for $X \in \{S, J\}$, both *D-X* and *M-X* optimal solutions, respectively given by m^*_{D-X} and m^*_{M-X}, are non-decreasing as γ increases.

This result directly follows from the testing cost structure, where each test requires a fixed cost of c_f, independently of the pool size, while each specimen included in a pool costs c_v, that is, only the variable cost component, c_v, is a function of pool size. Thus, an increase in the specimen collection cost, c_v, for a given value of c_f (i.e., a reduction in γ), makes it more efficient to use smaller pool sizes, in turn leading to an increase in the number of pools tested in the joint problem. This result is relevant, because new assays, with different testing cost structures, are continuously introduced as technology evolves.

5 Case Study: Estimating West Nile Virus Prevalence in Mosquitoes

Our goal in this section is to gain insights and derive principles on testing pool design for prevalence estimation. For this purpose, we utilize the optimization models, *D-S*, *D-J*, *M-S*, and *M-J*, to design testing pools so as to efficiently estimate the prevalence of mosquitoes carrying West Nile virus (WNV), and compare the outcomes of the single-variable and joint pol design models with each other, and with a benchmark design from the literature.

In the remainder of this section, we first discuss the data sources and model calibration (Section 5.1), and then compare the performance of the optimal pool designs developed by *D-S*, *M-S*, *D-J*, and *M-J* (Section 5.2).

5.1 Data Sources and Model Calibration

WNV infection is a vector-borne disease, and the primary source of disease transmission to humans is a mosquito bite [24]. The mosquito population carrying WNV in any given year depends on various factors, including temperature and humidity [16, 24]. As a result, prevalence rate of WNV in mosquitoes is highly seasonal, and can fluctuate substantially from year to year [16, 24, 34].

WNV infection is a seasonal endemic in the United States, causing many fatalities from neuro-invasive diseases each year [24, 34]. Further, as most cases are asymptomatic, WNV infection in humans is significantly under-reported [59], further contributing to the risk of a transfusion-transmitted WNV infection via blood transfusion or organ transplantation [32,47]. WNV prevalence in humans is highly correlated with that in mosquitoes [11, 14, 33], and an accurate estimate of the prevalence rate of WNV-carrying mosquitoes is essential for outbreak prediction and prevention of WNV infection in humans [24, 25].

We consider the reverse-transcription polymerase chain reaction (RT-PCR) assay for WNV screening in mosquitoes [25]. RT-PCR assays employ the nucleic acid amplification technology for detecting the viral RNA present in the specimens, and are highly sensitive and specific [42]. We also consider a *benchmark pool design*, of $(m = 50, n = 30)$, used in the literature to study WNV prevalence in mosquitoes via RT-PCR assays [43], resulting in a testing budget of $B = \$8,160$, based on the testing cost parameters from [25]; see Table 1. The benchmark pool size, $m = 50$, is also commonly used in pooled testing for WNV prevalence estimation in mosquitoes via RT-PCR assays (e.g., [21, 22, 39, 44]). All data in the numerical study are estimated based on published studies, and are complemented via sensitivity analysis; see Table 1.

As input, deterministic pool design models, *D-S* and *D-J*, require a point estimate of the unknown prevalence rate, p_0, and robust pool design models, *M-S* and *M-J*, require an estimate of the upper bound of the support, p_{UB} (Theorems 1 and 2). We use the four optimization models to develop pool designs under *accurate* (perfect information) and *inaccurate* (imperfect information) model input *scenarios* (Table 1) at a testing budget of $B = \$8,160$, and evaluate their performance

via a Monte-Carlo simulation study, which is described in detail below. To this end, we construct the true support of WNV prevalence based on WNV surveillance data from various parts of the Mid-South region of the United States during the 2002-2005 period, where WNV transmission was the most intense during the outbreak from 2002 to 2003 [24], and WNV-infected mosquito prevalence was reported to be as high as 8.76% in Tennessee Valley [12]In the Monte-Carlo simulation, we use a *true* upper bound, $p_{UB} = 0.09$, *true* lower bound, $p_{LB} = 0.003$ (based on [43]), and consider that $P \sim$ Uniform $(p_{LB}, p_{UB}) = (0.003, 0.09)$, hence has a true mean of 0.0465 (the distribution information is not needed for optimization models). On the other hand, we develop optimal pool designs based on three scenarios, obtained by varying the upper bound, p_{UB}, used by the tester (Table 1): the $p_{UB} = 0.09$ case corresponds to the accurate model input scenario, and the $p_{UB} \in \{0.03, 0.15\}$ cases correspond to the inaccurate model input scenarios; for each scenario, we derive the point estimate required by the deterministic models, *D-S* and *D-J*, as $p_0 = \frac{1}{2}(p_{LB}+p_{UB})$, using $p_{LB} = 0.003$.

Table 1: Data and data sources used in the numerical study.

	Cost Parameters (Source)	
c_f	\$72 [25]	
c_v	\$4 [25]	
	Simulation Parameters	**Model Input Scenarios**
p_{LB}	0.003	not needed
p_{UB}	0.09	$\{0.03, 0.09, 0.15\}$
p_0	N/A	$\frac{1}{2}(p_{LB} + p_{UB})$ (derived based on $p_{LB} = 0.003$)
	Benchmark Design (Source)	**Benchmark Budget (Source)**
	$(m, n) = (50, 30)$ [43]	$B = \$8,160$ [25, 43]

The Monte-Carlo simulation performs 20,000 replications for each of the three scenarios. Specifically, for each scenario, we first determine the optimal pool designs for each of the *D-S*, *D-J*, *M-S*, and *M-J* Models for a testing budget of $B = \$8,160$, based on the inputs provided in Table 1. To obtain an optimal integer number of pools in *D-J* and *M-J*, we utilize Theorem 1 repeatedly within a branch and bound algorithm implemented in MATLAB. Then, in each simulation replication, we use the Monte-Carlo-based random number generator in MATLAB, and randomly generate a realization of P from Uniform$(0.003, 0.09)$ (Table 1), which we denote by p. Based on the generated value of p, we then randomly generate the carrier status of each subject, i.e., specimen from a mosquito (for $m^* \times n$ subjects for *D-S* and *M-S* Models, and $m^* \times n^*(m^*)$ subjects for *D-J* and *M-J* Models), where each specimen is infected with WNV with probability p, and infection-free

with probability 1-p. These specimens are then randomly assigned to the testing pools. If a pool contains at least one infected specimen, then the pool's test outcome will be positive; and otherwise, the pool's test outcome will be negative. For each replication, we compute: the MLE of prevalence ($\hat{p}$, Eqn. (1)), asymptotic variance ($\sigma^2(m^*, n^*; p)$, Eqn. (2)), and the MSE and percent relative bias of the prevalence estimate ($rBias(\%)$), given by:

$$MSE = (\hat{p} - p)^2, \text{ and } rBias(\%) = 100 \times \left| \frac{\hat{p} - p}{p} \right|.$$

These performance metrics are commonly used in the statistics literature to evaluate the efficiency of an estimator (e.g., [29, 30, 61]). Table 2 reports the performance metric in the form, *average ± half width of the 95% confidence interval (CI)*, over 20,000 replications for the accurate and inaccurate input scenarios.

5.2 Pool Design Comparison

Table 2 reports the performance metrics in the form, *average ± half width of the 95% confidence interval (CI)*, over 20,000 replications for the accurate and inaccurate input scenarios. We summarize the findings from the numerical study below.

- **Joint Pool Design:** Not surprisingly, jointly selecting m and n in an optimal manner can substantially improve estimation accuracy, compared to the setting where the number of pools (n) is fixed *a priori*. While this improvement happens for all three scenarios, it is more pronounced when the fixed number of pools is small (e.g., 20), and the tester underestimates the prevalence ($p_{UB} = 0.03$); for example, for this scenario ($n = 20, p_{UB} = 0.03$), joint pool design reduces the MSE more than ten-fold for the deterministic model (from 0.485 to 0.0424), and more than five hundred-fold for the robust model (from 0.485 to 8.64×10^{-4}). Also observe that the benchmark design, ($m = 50, n = 30$), i.e., thirty pools of size fifty, corresponds to the optimal solution for **D-S** and **M-S** only when $p_{UB} = 0.03$, that is, the underestimation scenario, and in general, performs rather poorly compared to the optimal pool designs. To see this, consider the accurate input case (i.e., $p_{UB} = 0.09$): when n is set to 30, it becomes inefficient to use the full budget (which allows a maximum pool size of 50) for both *D-S* and *M-S*; instead it is better to use pool sizes of 33 and 17, respectively for *D-S*

Table 2: Performance (average ± half width of 95% CI) of *D-S*, *M-S*, *D-J* and *M-J* designs, and the benchmark design with varying values of p_{UB} (accurate and inaccurate model inputs; true $p_{UB} = 0.09$)

		Assumed $p_{UB} = 0.03$		Assumed $p_{UB} = 0.09$		Assumed $p_{UB} = 0.15$	
n		***D-S* Model**	***M-S* Model**	***D-S* Model**	***M-S* Model**	***D-S* Model**	***M-S* Model**
20	m^*	84	52	33	17	20	10
	$\hat{p}$	0.5631 ± 0.00674	0.3166 ±0.00610	0.1191 ± 0.00358	0.0494 ±0.00058	0.0528 ± 0.00096	0.0480 ± 0.00044
	$\sigma^2(m^*, n; p)[\times 10^4]$	20.70 ± 0.49	4.18 ± 0.07	2.37 ± 0.03	2.26 ± 0.02	2.17 ± 0.02	2.95 ± 0.03
	$MSE[\times 10^4]$	4,850 ± 60.47	2,530 ± 54.45	671 ± 31.76	10.30 ± 3.48	38.60 ± 7.63	3.32 ± 0.11
	$rBias(\%)$	895 ±12.28	416 ± 8.83	117.73 ±4.52	30.64 ± 0.59	33.05 ± 1.09	34.83 ± 0.46
30	m^*	50	50	33	17	20	10
	$\hat{p}$	0.2411 ±0.00545	0.2411 ±0.00545	0.0838 ±0.00260	0.0479 ± 0.00042	0.0488 ± 0.00053	0.0475 ± 0.00041
	$\sigma^2(m^*, n; p)[\times 10^4]$	2.58 ± 0.04	2.58 ± 0.04	1.58 ± 0.02	1.51 ± 0.01	1.45 ± 0.01	1.97 ± 0.02
	$MSE[\times 10^4]$	1,810 ± 48.52	1,810 ± 48.52	336 ± 22.86	2.21 ± 0.83	7.26 ± 2.97	2.12 ± 0.06
	$rBias(\%)$	286.42 ±7.32	286.42 ±7.32	65.66 ±3.12	23.95 ± 0.32	23.68 ± 0.48	28.23 ± 0.38
50	m^*	22	22	22	17	20	10
	$\hat{p}$	0.0476 ±0.00042	0.0476 ± 0.00042	0.0476 ±0.00042	0.0473 ± 0.00038	0.0474 ± 0.00038	0.0472 ± 0.00039
	$\sigma^2(m^*, n; p)[\times 10^4]$	0.86 ± 0.01	0.86 ± 0.01	0.86 ± 0.01	0.91 ± 0.01	0.87 ± 0.01	1.18 ± 0.01
	$MSE[\times 10^4]$	2.25 ± 1.41	2.25 ± 1.41	2.25 ± 1.41	0.99 ± 0.03	0.99 ± 0.04	1.25 ± 0.03
	$rBias(\%)$	17.22 ±0.27	17.22 ±0.27	17.22 ±0.27	18.23 ± 0.23	17.54 ± 0.22	21.82 ± 0.28
100	m^*	2	2	2	2	2	2
	$\hat{p}$	0.0466 ±0.00041	0.0466 ±0.00041	0.0466 ±0.00041	0.0466 ±0.00041	0.0466 ±0.00041	0.0466 ±0.00041
	$\sigma^2(m^*, n; p)[\times 10^4]$	2.26 ± 0.02	2.26 ± 0.02	2.26 ± 0.02	2.26 ± 0.02	2.26 ± 0.02	2.26 ± 0.02
	$MSE[\times 10^4]$	2.23 ± 0.06	2.23 ± 0.06	2.23 ± 0.06	2.23 ± 0.06	2.23 ± 0.06	2.23 ± 0.06
	$rBias(\%)$	31.29 ±0.42	31.29 ±0.42	31.29 ±0.42	31.29 ±0.42	31.29 ±0.42	31.29 ±0.42

	***D-J* Model**	***M-J* Model**	***D-J* Model**	***M-J* Model**	***D-J* Model**	***M-J* Model**
(m^*, n^*)	(37,37)	(25,47)	(19,55)	(12,68)	(12,68)	(7,81)
$\hat{p}$	0.0930 ±0.00291	0.0487 ± 0.00056	0.0474 ±0.00039	0.0470 ± 0.00037	0.0470 ± 0.00037	0.0468 ± 0.00038
$\sigma^2(m^*, n^*; p)[\times 10^4]$	1.39 ± 0.02	0.92 ± 0.01	0.79 ± 0.01	0.78 ± 0.01	0.78 ± 0.01	0.94 ± 0.01
$MSE[\times 10^4]$	424 ± 25.58	8.64 ± 3.47	1.31 ± 0.82	0.81 ± 0.02	0.81 ± 0.02	0.95 ± 0.02
$rBias(\%)$	75.47 ±3.50	18.37 ± 0.48	16.89 ±0.23	17.55 ± 0.23	17.55 ± 0.23	19.52 ± 0.26

	Benchmark Design
(m, n)	(50,30)
$\hat{p}$	0.2411 ±0.00545
$\sigma^2(m, n; p)[\times 10^4]$	2.58 ± 0.04
$MSE[\times 10^4]$	1,810 ± 48.52
$rBias(\%)$	286.42 ±7.32

and *M-S*, thus leaving some unused budget, which could have been used to test more pools had the number of pools, n, not been fixed. On the other hand, when the fixed number of pools is relatively large with respect to the testing budget (e.g., $n = 100$), the optimization models, *D-S* and *M-S*, do not have much flexibility, and choose the maximum feasible pool size as the optimal pool size, i.e., $m^* = 2$, independent of the model inputs.

- **Robust Pool Design:** Considering joint pool design, when the model input, p_{UB}, is inaccurate, the robust model, *M-J*, proves to be especially robust, in terms of minimizing the estimation error (i.e., *MSE* and *rBias*), especially when p_{UB} is an underestimate of the true upper bound of ($p_{UB} = 0.03$) – similarly, p_0 is an underestimate of the true mean of P. In this case, *M-J* performs substantially better than *D-J*. However, when p_{UB} is an overestimate of the true upper bound ($p_{UB} = 0.15$), *M-J* becomes overly conservative and underperforms in comparison to *D-J*, but with only a slight reduction in performance. Consequently, when the inputs for pool design models are uncertain, *M-J* performs well and produces an accurate estimate of the prevalence rate.

When estimating the prevalence of emerging and/or seasonal diseases, the distribution and support of P are highly uncertain at the outset, and thus input parameters to pool design models are likely to be inaccurate. It is in these realistic cases that the joint pool design models, *D-J* and *M-J*, perform significantly better than their single-variable counter-parts, *D-S* and *M-S*. Further, between the joint models, the robust model, *M-J*, often leads to lower *MSE* and relative bias, compared to the deterministic model, *D-J*, even when the input is highly inaccurate. These findings have important implications for designing surveillance studies of emerging and seasonal infectious infections.

6 Conclusions and Future Research Directions

We develop and study deterministic and robust pool design optimization models for prevalence estimation under limited resources in two settings: when the number of pools is fixed exogenously (due, for example, to limited number of testing kits), and when both the pool size and the number of pools are jointly optimized. Our analysis indicates that joint pool design fits especially well with

prevalence estimation of emerging or seasonal diseases, such as Zika, West Nile virus, or COVID-19 infections, for which initial information, prior to testing, is often highly unreliable. We establish key structural properties of optimal pool designs for the different models, and show that the optimal testing pool design for each model follows a threshold-type policy, contributing to the existing literature of testing pool design. Our case study, on estimating the prevalence of West Nile virus in mosquitoes, further underscores the value of robust pool designs, especially for emerging or seasonal diseases.

As immediate, and important, extensions of our study, one can relax some of our modeling assumptions, including that the screening test is perfectly reliable. If the screening test has less than perfect specificity and/or specificity, one needs to consider both the number of infected specimens in the pool and the bio-marker concentration of each infected specimen, which varies with the stage of the infection, to determine the test outcome, similar to [40, 41]. Further, because the joint robust (min-max) model tends to be overly conservative in cases where model inputs are overestimates, other robust pool design models, such as regret-based models, can be investigated. Another interesting research direction is to study optimal sequential (multi-stage) pooling strategies, and their benefit, for prevalence estimation. We study resource (testing budget) allocation for the prevalence estimation of a single disease in a specific region. In practice, health policy makers may need to allocate their testing budget to prevalence estimation activities for a number of diseases in a specific region, or in various regions, each potentially having a different prevalence rate of the disease in question. Therefore, extending our models to study pool design optimization and budget allocation for prevalence estimation for multiple diseases or multiple regions is an important research direction. Additionally, practitioners may need to select a screening test, among a set of commercially available tests, including combo tests, i.e., tests that can simultaneously detect a number of diseases. Thus, another interesting future research direction is to study the problem of test selection and pool design for prevalence estimation of multiple diseases. It is our hope that the proposed pool design models spur further research in academia, and our insights and principles on testing pool design have an impact on surveillance studies in practice.

Acknowledgments: We would like thank Dr. Joe Zhu, the Area Editor, and the two anonymous reviewers for great suggestions and comments that improved the presentation of this paper.

This material is based upon work supported in part by the National Science Foundation under Grant No.# 1761842. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] K. Altucker. Labs Are Testing 100,000 People Each Day for the Coronavirus. That's Still Not Enough, April 2020. Available at `https://www.usatoday.com/story/news/health/2020/04/02/coronavirus-testing-number-labs-covid/5099458002/` (accessed: April 24, 2020).

[2] K.V. Brown and E. Court. U.S. Labs Face Crisis After Crisis Despite Improvements in Testing, April 2020. Available at `https://www.bloomberg.com/news/articles/2020-04-07/coronavirus-testing-accuracy-and-availability-shortages-remain` (accessed: April 24, 2020).

[3] P. M. Burrows. Improved Estimation of Pathogen Transmission Rates by Group Testing. *Phytopathology*, 77(2):363–365, 1987.

[4] G. Casella and R. L. Berger. Statistical Inference. Volume 2, chapter 7, page 335. Duxbury Pacific Grove, CA, 2002.

[5] Centers for Disease Control and Prevention. CDC Awards Nearly $184 Million to Continue the Fight against Zika, December 2016. Available at `https://www.cdc.gov/media/releases/2016/p1222-zika-funding.html` (accessed: June 11, 2018).

[6] Centers for Disease Control and Prevention. Multistate Outbreak of E. coli O157:H7 Infections Linked to Romaine Lettuce, June 2018. Available at `https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html` (accessed: June 14, 2018).

[7] Y.P. Chaubey and W. Li. Estimation of Fraction Defectives Through Batch Sampling. *Proc. Qual. Productvty Sect. Am. Statist Ass*, pages 198–206, 1993.

[8] C. L. Chen and W. H. Swallow. Using Group Testing to Estimate a Proportion, and to Test the Binomial Model. *Biometrics*, 46(4):1035–1046, 1990.

[9] C. L. Chen and W. H. Swallow. Sensitivity Analysis of Variable-Size Group Testing and its Related Continuous Models. *Biometrical Journal*, 37(2):173–181, 1995.

[10] P. Chen, J. M. Tebbs, and C. R. Bilder. Group Testing Regression Models with Fixed and Random Effects. *Biometrics*, 65(4):1270–1278, 2009.

[11] W. M. Chung, C. M. Buseman, S. N. Joyner, S. M. Hughes, T. B. Fomby, J. P. Luby, and R. W. Haley. The 2012 West Nile Encephalitis Epidemic in Dallas, Texas. *Journal of the American Medical Association*, 310(3):297–307, 2013.

[12] E. W. Cupp, H. K. Hassan, X. Yue, W. K. Oldland, B. M. Lilley, and T. R. Unnasch. West Nile Virus Infection in Mosquitoes in the Mid-South USA, 2002–2005. *Journal of Medical Entomology*, 44(1):117–125, 2007.

[13] J. R. Davis, J. Lederberg, et al. Public Health Systems and Emerging Infections: Assessing the Capabilities of the Public and Private sectors. Workshop Summary. In *Public Health Systems and Emerging Infections: Assessing the Capabilities of the Public and Private sectors. Workshop Summary.* National Academy Press, 2000.

[14] N. B. DeFelice, E. Little, S. R. Campbell, and J. Shaman. Ensemble Forecast of Human West Nile Virus Cases and Mosquito Infection Rates. *Nature Communications*, 8:14592, 2017.

[15] R. Dorfman. The Detection of Defective Members of Large Populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

[16] H. El-Amine, E. K. Bish, and D. R. Bish. Robust Postdonation Blood Screening under Prevalence Rate Uncertainty. *Operations Research*, 66(1):1–17, 2018.

[17] X. Fang, W. W. Stroup, and S. Zhang. Improved Empirical Bayes Estimation in Group Testing Procedure for Small Proportions. *Communications in Statistic–Theory and Methods*, 36(16):2937–2944, 2007.

[18] C. P. Farrington. Estimating Prevalence by Group Testing Using Generalized Linear Models. *Statistics in Medicine*, 11(12):1591–1597, 1992.

[19] J. J. Gart. An Application of Score Methodology: Confidence Intervals and Tests of Fit for One-hit Curves. *Handbook of Statistics*, 8:395–406, 1991.

[20] J. L. Gastwirth and P. A. Hammick. Estimation of the Prevalence of a Rare Disease, Preserving the Anonymity of the Subjects by Group Testing: Application to Estimating the Prevalence of AIDS Antibodies in Blood Donors. *Journal of Statistical Planning and Inference*, 22(1):15–27, 1989.

[21] W. Gu, T. R. Unnasch, C. R. Katholi, R. Lampman, and R. J. Novak. Fundamental Issues in Mosquito Surveillance for Arboviral Transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(8):817–822, 2008.

[22] T.L. Hadfield, M. Turell, M.P. Dempsey, J. David, and E.J Park. Detection of West Nile Virus in mosquitoes by RT-PCR. *Molecular and Cellular Probes*, 15(3):147–150, 2001.

[23] P. A. Hammick and J. L. Gastwirth. Group Testing for Sensitive Characteristics: Extension to Higher Prevalence Levels. *International Statistical Review/Revue Internationale de Statistique*, pages 319–331, 1994.

[24] E. B. Hayes, N. Komar, R. S. Nasci, S. P. Montgomery, D. R. O'Leary, and G. L. Campbell. Epidemiology and Transmission Dynamics of West Nile Virus Disease. *Emerging Infectious Diseases*, 11(8):1167, 2005.

[25] J. M. Healy, W. K. Reisen, V. L. Kramer, M. Fischer, N. P. Lindsey, R. S. Nasci, P. A. Macedo, G. White, R. Takahashi, L. Khang, et al. Comparison of the Efficiency and Cost of West Nile Virus Surveillance Methods in California. *Vector-Borne and Zoonotic Diseases*, 15(2):147–155, 2015.

[26] G. Hepworth and R. Watson. Debiased Estimation of Proportions in Group Testing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):105–121, 2009.

[27] X. Huang. An Improved Test of Latent-Variable Model Misspecification in Structural Measurement Error Models for Group Testing Data. *Statistics in Medicine*, 28(26):3316–3327, 2009.

[28] J. M. Hughes-Oliver. Pooling Experiments for Blood Screening and Drug Discovery. In A. Dean and S. Lewis, editors, *Screening*, chapter 3, pages 48–68. Springer, 2006.

[29] J. M. Hughes-Oliver and W. F. Rosenberger. Efficient Estimation of the Prevalence of Multiple Rare Traits. *Biometrika*, 87(2):315–327, 2000.

[30] J. M. Hughes-Oliver and W. H. Swallow. A Two-Stage Adaptive Group-Testing Procedure for Estimating Small Proportions. *Journal of the American Statistical Association*, 89(427):982–993, 1994.

[31] M. Hung and W. H. Swallow. Robustness of Group Testing in the Estimation of Proportions. *Biometrics*, 55(1):231–237, 1999.

[32] M. Iwamoto, D. B. Jernigan, A. Guasch, M. J. Trepka, C. G. Blackmore, W. C. Hellinger, S. M. Pham, S. Zaki, R. S. Lanciotti, S. E. Lance-Parker, et al. Transmission of West Nile Virus from an Organ Donor to Four Transplant Recipients. *New England Journal of Medicine*, 348(22):2196–2203, 2003.

[33] A. M. Kilpatrick and W. J. Pape. Predicting Human West Nile Virus Infections with Mosquito Surveillance Data. *American Journal of Epidemiology*, 178(5):829–835, 2013.

[34] C. T. Korves, S. J. Goldie, and M. B. Murray. Cost-effectiveness of Alternative Blood-screening Strategies for West Nile Virus in the United States. *PLoS Medicine*, 3(2):e21, 2006.

[35] W.M. Kyaw, H.Y. Loke, A. Chow, M. Chan, and Y.S. Leo. Surveillance of Zika Virus Infection: The Experience of an Adult Tertiary Care Hospital in Singapore. *International Journal of Infectious Diseases*, 53:81–82, 2016.

[36] C. T. Le. A New Estimator for Infection Rates Using Pools of Variable Size. *American Journal of Epidemiology*, 114(1):132–136, 1981.

[37] A. Liu, C. Liu, Z. Zhang, and P. S. Albert. Optimality of Group Testing in the Presence of Misclassification. *Biometrika*, 99(1):245–251, 2012.

[38] C. S. McMahan, J. M. Tebbs, and C. R. Bilder. Regression Models for Group Testing Data with Pool Dilution Effects. *Biostatistics*, 14(2):284–298, 2012.

[39] Navy Entomology Center for Excellence. West Nile Virus Surveillance and Control Guide For U.S. Navy and Marine Corps Installation 2014, October 2014. Available at `https://www.med.navy.mil/sites/nmcphc/Documents/nece/WNV-Surveillance-and-Control-Guide-2014.pdf` (accessed: January 17, 2019).

[40] N. T. Nguyen, H. Aprahamian, E. K. Bish, and D. R. Bish. A Methodology for Deriving the Sensitivity of Pooled Testing, Based on Viral Load Progression and Pooling Dilution. *Journal of Translational Medicine*, 17(1):1–10, 2019.

[41] N. T. Nguyen, E. K. Bish, and H. Aprahamian. Sequential Prevalence Estimation with Pooling and Continuous Test Outcomes. *Statistics in Medicine*, 37(15):2391–2426, 2018.

[42] L. Overbergh, A. Giulietti, D. Valckx, B. Decallonne, R. Bouillon, and C. Mathieu. The Use of Real-Time Reverse Transcriptase PCR for the Quantification of Cytokine Gene Expression. *Journal of Biomolecular Techniques: JBT*, 14(1):33, 2003.

[43] N. A. Pritchard and J. M. Tebbs. Bayesian Inference for Disease Prevalence using Negative Binomial Group Testing. *Biometrical Journal*, 53(1):40–56, 2011.

[44] C. R. Rutledge, J. F. Day, C. C. Lord, L. M. Stark, and W. J. Tabachnick. West Nile Virus Infection Rates in Culex Nigripalpus (Diptera: Culicidae) Do not Reflect Transmission Rates in Florida. *Journal of Medical Entomology*, 40(3):253–258, 2003.

[45] P. Sham, J. S. Bader, I. Craig, M. O'Donovan, and M. Owen. DNA Pooling: a Tool for Large-Scale Association Studies. *Nature Reviews Genetics*, 3(11):862–871, 2002.

[46] M. Sobel and R.M. Elashoff. Group Testing with a New Goal, Estimation. *Biometrika*, 62(1):181–193, 1975.

[47] S. L. Stramer, C. T. Fang, G. A. Foster, A. G. Wagner, J. P. Brodsky, and R. Y. Dodd. West Nile Virus among Blood Donors in the United States, 2003 and 2004. *New England Journal of Medicine*, 353(5):451–459, 2005.

[48] L. Strickler and A. Kaplan. Private Labs Do 85 Percent of U.S. COVID-19 Tests but Still Struggle with Backlogs, Shortages, April 2020. Available at `https://www.nbcnews.com/health/health-news/private-labs-do-85-percent-u-s-covid-19-tests-n1177866` (accessed: April 24, 2020).

[49] W. H. Swallow. Group Testing for Estimating Infection Rates and Probabilities of Disease Transmission. *Phytopathology*, 75(8):882–889, 1985.

[50] J. M. Tebbs and C. R. Bilder. Confidence Interval Procedures for the Probability of Disease Transmission in Multiple-Vector-Transfer Designs. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(1):75, 2004.

[51] J. M. Tebbs, C. R. Bilder, and B. K. Moser. An Empirical Bayes Group-Testing Approach to Estimating Small Proportions. *Communications in Statistics-Theory and Methods*, 32(5):983–995, 2003.

[52] J. M. Tebbs, C. S. McMahan, and C. R. Bilder. Two-Stage Hierarchical Group Testing for Multiple Infections with Application to the Infertility Prevention Project. *Biometrics*, 69(4):1064–1073, 2013.

[53] J. M. Tebbs and W. H. Swallow. Estimating Ordered Binomial Proportions with the Use of Group Testing. *Biometrika*, pages 471–477, 2003.

[54] K. H. Thompson. Estimation of the Proportion of Vectors in a Natural Population of Insects. *Biometrics*, 18(4):568–578, 1962.

[55] X. M. Tu, E. Litvak, and M. Pagano. Screening Tests: Can We Get More by Doing Less? *Statistics in Medicine*, 13(19-20):1905–1919, 1994.

[56] X. M. Tu, E. Litvak, and M. Pagano. On the Informativeness and Accuracy of Pooled Testing in Estimating Prevalence of a Rare Disease: Application to HIV Screening. *Biometrika*, 82(2):287–297, 1995.

[57] S. Vansteelandt, E. Goetghebeur, and T. Verstraeten. Regression Models for Disease Prevalence with Diagnostic Tests on Pools of Serum Samples. *Biometrics*, 56(4):1126–1133, 2000.

[58] L. M. Wein and S. A. Zenios. Pooled Testing for HIV Screening: Capturing the Dilution Effect. *Operations Research*, 44(4):543–569, 1996.

[59] World Health Organization. West Nile virus, October 2017. Available at `http://www.who.int/news-room/fact-sheets/detail/west-nile-virus` (accessed: September 5, 2018).

[60] M. Xie. Regression Analysis of Group Testing Samples. *Statistics in Medicine*, 20(13):1957–1969, 2001.

[61] S. A. Zenios and L. M. Wein. Pooled Testing for HIV Prevalence Estimation: Exploiting the Dilution Effect. *Statistics in Medicine*, 17(13):1447–1467, 1998.

[62] Z. Zhang, C. Liu, S. Kim, and A. Liu. Prevalence Estimation Subject to Misclassification: the Mis-substitution Bias and Some Remedies. *Statistics in Medicine*, 33(25):4482–4500, 2014.

Appendix

Appendix A: Summary of Notation

Decision Variables	
m	Pool size, i.e., number of specimens in each pool
n	Number of testing pools (parameter in the single-variable models; decision variable in the joint models)
Random Variables	
P	True (unknown) prevalence rate of the disease (with realization p)
$\hat{P}$	Maximum likelihood estimator (MLE) of P (with realization $\hat{p}$)
$T(m,n;p)$	Number of positive-testing pools among n pools, each containing m specimens, given a true prevalence rate of p
Objective Function and Performance Metrics	
$\sigma^2(m,n;p)$	Asymptotic variance of $\hat{P}$ for pool design (m,n), given a true prevalence rate of p
$MSE(\hat{P},m,n;p)$	Mean square error of $\hat{P}$ for pool design (m,n), given a true prevalence rate of p
$rBias(\%) = 100\left(\left\lvert\frac{\hat{p}-p}{p}\right\rvert\right)$	Relative bias of $\hat{p}$ (in percentage), given a true prevalence rate of p
Model Parameters	
p_0	An initial point estimate of P
p_{LB}	Lower bound of P
p_{UB}	Upper bound of P
c_f	Fixed testing cost per pool tested
c_v	Variable testing cost per specimen tested
B	Testing budget

B: Supporting Results

The following results establish key properties of the asymptotic variance function in the exogenous number of pools and endogenous number of pools settings, respectively.

Lemma 6. For any $n \in \mathbb{Z}^+$, $\sigma^2(m,n;p)$ has the following properties:

1. For $m \geq 2$, $\sigma^2(m,n;p)$ increases as p increases, $\forall p < 1$.
2. $\sigma^2(1,n;p)$increases as p increases, $\forall p < \frac{1}{2}$.
3. $\sigma^2(m,n;p)$ decreases as n increases.
4. $\sigma^2(m,n;p)$ is strictly convex in m.

Lemma 7. $\sigma^2(m,n^*(m);p)$ has the following properties:

1. For $m \geq 2$, $\sigma^2(m,n^*(m);p)$ increases as p increases, $\forall p < 1$.
2. $\sigma^2(1,n;p)$ increases as p increases, $\forall p < \frac{1}{2}$.
3. $\sigma^2(m,n^*(m);p)$ is strictly convex in m, $\forall p < \frac{1}{3}$.

Appendix C: Proofs

Proof of Lemma 1: To simplify the notation, we represent $\pi_0^S(m_1, m_2)$ as π_0.

Part 1. By Definition 1 and Proposition 1, π_0 is the unique solution to:

$$\begin{aligned}
&\frac{1-(1-\pi_0)^{m_1}}{nm_1^2(1-\pi_0)^{m_1-2}} = \frac{1-(1-\pi_0)^{m_2}}{nm_2^2(1-\pi_0)^{m_2-2}} \\
\Leftrightarrow\ &\left(\frac{1}{1-\pi_0}\right)^{m_1-2} - (1-\pi_0)^2 = \left(\frac{m_1}{m_2}\right)^2\left[\left(\frac{1}{1-\pi_0}\right)^{m_2-2} - (1-\pi_0)^2\right] \\
\Leftrightarrow\ &\left(\frac{1}{1-\pi_0}\right)^{m_1} = 1 + \left(\frac{m_1}{m_2}\right)^2\left[\left(\frac{1}{1-\pi_0}\right)^{m_2} - 1\right].
\end{aligned}$$

Part 2. Define $\theta(m_1, m_2) \equiv 1-\pi_0$, for $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$. Then, $\theta(m_1, m_2) \in (0,1)$. First, note that the equation provided in Part 1 is equivalent to:

$$(1-\pi_0)^{m_2} = (1-\pi_0)^{m_2-m_1}\Big(\frac{m_2^2}{m_2^2-m_1^2}\Big) - \Big(\frac{m_1^2}{m_2^2-m_1^2}\Big). \tag{6}$$

a. Proof that $\pi_0^S(m_1, m_2)$ is decreasing in m_1:

Taking the derivative of the RHS of Eqn. (6) with respect to m_1 yields:

$$\begin{aligned}
\frac{\partial}{\partial m_1}\left\{(1-\pi_0)^{m_2-m_1}\Big(\frac{{m_2}^2}{{m_2}^2-{m_1}^2}\Big) - \Big(\frac{{m_1}^2}{{m_2}^2-{m_1}^2}\Big)\right\} &= -\left[\theta(m_1,m_2)\right]^{m_2-m_1}\log[\theta(m_1,m_2)]\Big(\frac{m_2^2}{m_2^2-m_1^2}\Big) \\
&\quad + \frac{2m_1m_2^2}{(m_2^2-m_1^2)^2}[(\theta(m_1,m_2))^{m_1-m_2}-1] > 0,
\end{aligned}$$

which follows because $\theta(m_1, m_2) < 1$ and $m_1 < m_2$, and hence, we have that $\log(\theta(m_1, m_2)) < 0$ and $(\theta(m_1, m_2))^{m_1-m_2} > 1$. Therefore, the RHS of Eqn (6) is increasing in m_1. Then, the LHS of Eqn. (6) must also increase in m_1 to preserve the equality, which in turn implies that $\theta(m_1, m_2)$ is increasing in m_1, and, equivalently, $\pi_0^S(m_1, m_2)$ is decreasing in m_1.

b. Proof that $\pi_0^S(m_1, m_2)$ is decreasing in m_2:

We prove this result by contradiction. Assume, to the contrary, that $\pi_0^S(m-2, m) > \pi_0^S(m-2, m-1)$ for some $m \in \mathbb{Z}^+ : m > 2$. From Part 1, we have that $\pi_0^S(m-1, m) < \pi_0^S(m-2, m)$, $\forall m \in \mathbb{Z}^+ : m > 2$. Therefore, we have two cases:

Case 1. $\pi_0^S(m-2, m-1) < \pi_0^S(m-1, m) < \pi_0^S(m-2, m)$:

Consider any $p \in \Big(\pi_0^S(m-1, m), \pi_0^S(m-2, m)\Big)$. Since $\pi_0^S(m-2, m-1) < \pi_0^S(m-1, m)$ (by

assumption of this case), $p > \pi_0^S(m-2, m-1)$. Therefore, by Proposition 1, $\sigma^2(m-2, n; p) < \sigma^2(m-1, n; p)$. Further, since $\pi_0^S(m-1, m) < p < \pi_0^S(m-2, m)$, by Proposition 1, $\sigma^2(m-1, n; p) < \sigma^2(m, n; p)$, and $\sigma^2(m, n; p) < \sigma^2(m-2, n; p)$, leading to a contradiction, i.e., $\sigma^2(m-2, n; p) < \sigma^2(m, n; p)$ *and* $\sigma^2(m-2, n; p) > \sigma^2(m, n; p)$. Hence, Case 1 is not possible.

Case 2. $\pi_0^S(m-1, m) < \pi_0^S(m-2, m-1) < \pi_0^S(m-2, m)$:

Consider any $p \in \Big(\pi_0^S(m-2, m-1), \pi_0^S(m-2, m)\Big)$. Since $\pi_0^S(m-2, m-1) > \pi_0^S(m-1, m)$ (by assumption of this case), $p > \pi_0^S(m-1, m)$. Therefore, by Proposition 1, $\sigma^2(m-1, n; p) < \sigma^2(m, n; p)$. Further, since $\pi_0^S(m-2, m-1) < p < \pi_0^S(m-2, m)$, by Proposition 1, $\sigma^2(m-2, n; p) < \sigma^2(m-1, n; p)$, and $\sigma^2(m, n; p) < \sigma^2(m-2, n; p)$, leading to a contradiction, i.e., $\sigma^2(m-2, n; p) > \sigma^2(m-1, n; p)$, *and* $\sigma^2(m-2, n; p) < \sigma^2(m-1, n; p)$. Hence, Case 2 is not possible.

From Cases 1 and 2, it follows that $\pi_0^S(m-2, m) < \pi_0^S(m-2, m-1)$, $\forall m \in \mathbb{Z}^+ : m > 2$.

Part 3. As a result of Part 2, $\pi_0^S(m-1, m) > \pi_0^S(m-1, m+1) > \pi_0^S(m, m+1)$, completing the proof. □

Proof of Lemma 6:

Part 1. From Eqn. (2), we can derive:

$$\frac{\partial}{\partial p}\sigma^2(m, n; p) = \frac{1}{nm^2}\Big[\frac{m-2}{(1-p)^{m-1}} + 2(1-p)\Big] > 0 \ \forall m \geq 2 \text{ and } p < 1,$$

and the result follows.

Part 2. From the derivation in Part 1, it follows that $\frac{\partial}{\partial p}\sigma^2(1, n; p) = 1 - 2p > 0$, $\forall p < \frac{1}{2}$, and the result follows.

Part 3. Since $\sigma^2(m, n; p) = \frac{1-(1-p)^m}{nm^2(1-p)^{m-2}}$, the result trivially follows.

Part 4. We have the following derivatives:

$$\frac{\partial}{\partial m}\sigma^2(m, n; p) = \frac{-2}{m^3(1-p)^{m-2}} - \frac{\log(1-p)}{m^2(1-p)^{m-2}} + \frac{2(1-p)^2}{m^3}, \text{ and}$$

$$\begin{aligned}\frac{\partial^2}{\partial m^2}\sigma^2(m, n; p) = \frac{6}{m^4}&\left\{\frac{m-2}{(1-p)^{m-1}} + 2(1-p)\right\} \\ &+ \frac{m(m-2)\log^2(1-p) - 2m\log(1-p) + 4(m-2)\log(1-p) - 4}{m^3(1-p)^{m-1}}.\end{aligned} \tag{7}$$

Consider the following:

$$\frac{\partial}{\partial p}\Big(\frac{\partial^2}{\partial m^2}\sigma^2(m,n;p)\Big) = \frac{6(m-2)+12m^4(1-p)^m+m^2(m-2)[\log(1-p)]^2-\log(1-p)(8m-2m^2)-4m}{m^4(1-p)^{m-1}}. \quad (8)$$

To find the root of $\frac{\partial}{\partial p}\Big(\frac{\partial^2}{\partial m^2}\sigma^2(m,n;p)\Big)$, we define $x \equiv \log(1-p)$, and solve the following quadratic equation:

$$m^2(m-2)x^2 - 2m(4-m)x - 4m + 6(m-2) + 12m^4(1-p)^m = 0,$$

where $\Delta = b^2-4ac = m^2[48m^4(1-p)^m(2-m) - 4m^2 + 32m - 32] < 0$, $\forall m \in \mathbb{R}^+$. Thus, the quadratic equation has no real root. Further, $\frac{\partial}{\partial p}\Big(\frac{\partial^2}{\partial m^2}\sigma^2(m,n;p)\Big)\Big|_{p=0} = \frac{12m^4+2m-12}{m^4} > 0$, $\forall m \geq 1$. Therefore, $\frac{\partial^2}{\partial m^2}\sigma^2(m,n;p)$ starts out as a positive function in p, and is increasing in p, $\forall m \geq 1$, $p \in (0,1)$. Hence, $\frac{\partial^2}{\partial m^2}\sigma^2(m,n;p) > 0$, $\forall m \geq 1$ and $p \in (0,1)$, which implies that $\sigma^2(m,n;p)$ is strictly convex in m, $\forall m \geq 1$ and $p \in (0,1)$. □

Proof of Lemma 2: The results follow because, by Lemma 6, $\sigma^2(m,n;p)$ is decreasing in n, for any given $m \in \mathbb{Z}^+$. Thus, in order to minimize $\sigma^2(m,n;p)$, for any given $m \in Z^+$, $n^*(m) = \frac{B}{c_f+c_v m}$, completing the proof. □

Proof of Proposition 2: For any $m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$, we study the ratio $\frac{\sigma^2(m_1,n^*(m_1);p)}{\sigma^2(m_2,n^*(m_2);p)}$, which, from Lemma 2, equals:

$$\frac{\sigma^2(m_1,n^*(m_1);p)}{\sigma^2(m_2,n^*(m_2);p)} = \left(\frac{c_f+c_v m_1}{c_f+c_v m_2}\right)\Big(\frac{m_2}{m_1}\Big)^2\left\{\frac{(1-p)^{m_2-2}\big[1-(1-p)^{m_1}\big]}{(1-p)^{m_1-2}\big[1-(1-p)^{m_2}\big]}\right\}, \text{ where}$$

$\lim\limits_{p\to 0}\dfrac{\sigma^2(m_1,n^*(m_1);p)}{\sigma^2(m_2,n^*(m_2);p)} = \dfrac{c_f m_2 + c_v m_1 m_2}{c_f m_1 + c_v m_1 m_2} > 1$ (by L'Hospital's Rule and since $m_2 > m_1 \geq 1$), and $\lim\limits_{p\to 1}\dfrac{\sigma^2(m_1,n^*(m_1);p)}{\sigma^2(m_2,n^*(m_2);p)} = 0$.

Further, from [37], the ratio $\Big(\frac{m_2}{m_1}\Big)^2\left\{\frac{(1-p)^{m_2-2}\big[1-(1-p)^{m_1}\big]}{(1-p)^{m_1-2}\big[1-(1-p)^{m_2}\big]}\right\}$ is decreasing in p, $\forall m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$. Thus, the result follows. □

Proof of Lemma 3:

Part 1. To simplify the notation, we represent $\pi_0^J(m_1,m_2)$ as π_0. From Definition 2 and Propo-

sition 2, π_0 is the unique solution to:

$$\left(\frac{c_f + c_v m_1}{B}\right)\left(\frac{1-(1-\pi_0)^{m_1}}{m_1^2(1-\pi_0)^{m_1-2}}\right) = \left(\frac{c_f + c_v m_2}{B}\right)\left(\frac{1-(1-\pi_0)^{m_2}}{m_2^2(1-\pi_0)^{m_2-2}}\right)$$

$$\Leftrightarrow \left(\frac{1}{1-\pi_0}\right)^{m_1-2} - (1-\pi_0)^2 = \left(\frac{m_1}{m_2}\right)^2\left(\frac{c_f + c_v m_2}{c_f + c_v m_1}\right)\left[\left(\frac{1}{1-\pi_0}\right)^{m_2-2} - (1-\pi_0)^2\right]$$

$$\Leftrightarrow \left(\frac{1}{1-\pi_0}\right)^{m_1} = 1 + \left(\frac{m_1}{m_2}\right)^2\left(\frac{c_f + c_v m_2}{c_f + c_v m_1}\right)\left[\left(\frac{1}{1-\pi_0}\right)^{m_2} - 1\right].$$

Part 2. We first show that $\pi_0^J(1,m)$ is decreasing in m, $\forall m \in \mathbb{Z}^+ : m \geq 2$. By Part 1 of this lemma (with proof above), $\forall m \in \mathbb{Z}^+ : m \geq 2$, $\pi_0^J(1,m)$ is the unique solution to:

$$\pi_0^J(1,m) = \left(\frac{c_f + c_v m}{c_f + c_v}\right)\left\{\frac{1-\left(1-\pi_0^J(1,m)\right)^m}{m^2\left(1-\pi_0^J(1,m)\right)^{m-1}}\right\}. \tag{9}$$

Taking the derivative of the RHS of Eqn. (9) with respect to m yields:

$$\frac{\partial}{\partial m}RHS = \left(\frac{1}{c_f + c_v}\right)\left[\left(\frac{-2c_f - c_v m}{m^3}\right)\left(\frac{1}{\left(1-\pi_0^J(1,m)\right)^{m-1}} - \left(1-\pi_0^J(1,m)\right)\right) + \left(\frac{c_f + c_v m}{m^2}\right)\left(\frac{-\log\left(1-\pi_0^J(1,m)\right)}{\left(1-\pi_0^J(1,m)\right)^{m-1}}\right)\right] > 0,$$

where the last inequality follows because $m \geq 2$, leading to $\frac{c_f + c_v m}{m^2} \geq \frac{2c_f + c_v m}{m^3}$; and because $\log\left(1-\pi_0^J(1,m)\right) < -1$, $\forall \pi_0^J(1,m) < \frac{2}{3}$ (by Lemma 1), leading to $\frac{-\log\left(1-\pi_0^J(1,m)\right)}{\left(1-\pi_0^J(1,m)\right)^{m-1}} > \frac{1}{\left(1-\pi_0^J(1,m)\right)^{m-1}} - \left(1-\pi_0^J(1,m)\right)$. Thus the RHS of Eqn. (9) is increasing in m. Further, the RHS of Eqn. (9), which equals $\frac{B}{c_f + c_v}\sigma^2(m, n^*(m), \pi_0^J(1,m))$, is also increasing in $\pi_0^J(1,m)$. Thus, $\pi_0^J(1,m)$ must decrease as m increases so as to preserve the equality in Eqn. (9).

We are now ready to show that $\pi_0^J(m_1,m_2)$ is decreasing in each of m_1 and m_2, $\forall m_1, m_2 \in \mathbb{Z}^+ : m_1 < m_2$.

a. Proof that $\pi_0^J(m_1,m_2)$ is decreasing in m_1:

The proof follows by induction and contradiction. To this end, we first show that $\pi_0^J(2,m_2) < \pi_0^J(1,m_2)$, $\forall m_2 > 2$. Suppose, to the contrary, that $\pi_0^J(2,m_2) > \pi_0^J(1,m_2)$. Since $\pi_0^J(1,m_2) < \pi_0^J(1,2)$, $\forall m_2 > 2$, there are two cases:

Case 1. $\pi_0^J(2,m_2) > \pi_0^J(1,2)$: By Proposition 2, we have that $\forall p \in \left(\pi_0^J(1,m_2), \pi_0^J(1,2)\right)$, $\sigma^2(1, n^*(1); p) <$

$\sigma^2(m_2, n^*(m_2); p) < \sigma^2(2, n^*(2); p)$, and $\sigma^2(1, n^*(1); p) > \sigma^2(2, n^*(2); p)$, leading to a contradiction.

Case 2. $\pi_0^J(1, m_2) < \pi_0^J(2, m_2) < \pi_0^J(1, 2)$: By Proposition 2, we have that $\forall p \in \left(\pi_0^J(1, m_2), \pi_0^J(2, m_2)\right)$, $\sigma^2(1, n^*(1); p) < \sigma^2(m_2; n^*(m_2); p) < \sigma^2(2, n^*(2); p)$ (since $m_2 > 2$), but since $\pi_0^J(2, m_2) < \pi_0^J(1, 2)$ (by assumption of this case), we also have that $\sigma^2(2, n^*(2); p) < \sigma^2(1, n^*(1); p)$ (by Proposition 2), leading to a contradiction.

Thus, it follows that $\pi_0^J(2, m_2) < \pi_0^J(1, m_2)$. It then follows, by induction, that $\pi_0^J(m_2 - 1, m_2) < \cdots < \pi_0^J(2, m_2) < \pi_0^J(1, m_2)$.

b. Proof that $\pi_0^J(m_1, m_2)$ is decreasing in m_2:

From Part a, we have that $\forall m_2 \in \mathbb{Z}^+$, $\pi_0^J(m_2 - 1, m_2) < \pi_0^J(m_2 - 2, m_2)$. We want to show that $\pi_0^J(m_2 - 2, m_2 - 1) > \pi_0^J(m_2 - 2, m_2)$. Suppose, to the contrary, that $\pi_0^J(m_2 - 2, m_2 - 1) < \pi_0^J(m_2 - 2, m_2)$. We have the following cases:

Case 1. $\pi_0^J(m_2 - 2, m_2 - 1) < \pi_0^J(m_2 - 1, m_2)$:

By Proposition 2, $\forall p \in \left(\pi_0^J(m_2-1, m_2), \pi_0^J(m_2-2, m_2)\right)$, $\sigma^2(m_2-1, n^*(m_2-1); p) < \sigma^2(m_2, n^*(m_2); p) < \sigma^2(m_2 - 2, n^*(m_2 - 2); p)$, and $\sigma^2(m_2 - 2, n^*(m_2 - 2); p) < \sigma^2(m_2 - 1, n^*(m_2 - 1); p)$, leading to a contradiction.

Case 2. $\pi_0^J(m_2 - 1, m_2) < \pi_0^J(m_2 - 2, m_2 - 1) < \pi_0^J(m_2 - 2, m_2)$:

By Proposition 2, $\forall p \in \left(\pi_0^J(m_2 - 2, m_2 - 1), \pi_0^J(m_2 - 2, m_2)\right)$, $\sigma^2(m_2 - 2, n^*(m_2 - 2); p) < \sigma^2(m_2 - 1, n^*(m_2 - 1); p) < \sigma^2(m_2, n^*(m_2); p)$, but $\sigma^2(m_2, n^*(m_2); p) < \sigma^2(m_2 - 2, n^*(m_2 - 2); p)$, leading to a contradiction.

Thus, it follows that $\pi_0^J(m_2 - 2, m_2 - 1) > \pi_0^J(m_2 - 2, m_2)$.

Part 3. From Part 2, $\pi_0^J(m - 1, m) > \pi_0^J(m - 1, m + 1) > \pi_0^J(m, m + 1)$, and the result follows. This completes the proof. □

Proof of Lemma 4: Recall that $\gamma = \frac{c_f}{c_v}$. From Part 1 of Lemma 3, we have the following:

$$\frac{m_2^2(1 - \pi_0^J(m_1, m_2))^{m_2}\left[1 - (1 - \pi_0^J(m_1, m_2))^{m_1}\right]}{m_1^2(1 - \pi_0^J(m_1, m_2))^{m_1}\left[1 - (1 - \pi_0^J(m_1, m_2))^{m_2}\right]} = \frac{c_f + c_v m_2}{c_f + c_v m_1} = \frac{\gamma + m_2}{\gamma + m_1} = 1 + \frac{(m_2 - m_1)}{\gamma + m_1}.$$

Note that $\frac{m_2^2(1-\pi_0^J(m_1,m_2))^{m_2}\left[1-(1-\pi_0^J(m_1,m_2))^{m_1}\right]}{m_1^2(1-\pi_0^J(m_1,m_2))^{m_1}\left[1-(1-\pi_0^J(m_1,m_2))^{m_2}\right]}$ is decreasing in $\pi_0^J(m_1, m_2)$ (from Liu *et al* [37]) and is constant in γ, while $1 + \frac{(m_2-m_1)}{\gamma+m_1}$ is decreasing in γ (since $m_2 > m_1$) and is constant in $\pi_0^J(m_1, m_2)$. Thus, as γ increases, $\pi_0^J(m_1, m_2)$ also increases. Further, we observe that when

$c_v = 0$, i.e., when $\gamma \to \infty$, $\pi_0^J(m_1, m_2) \to \pi_0^S(m_1, m_2)$. Since $\pi_0^J(m_1, m_2)$ is increasing in γ, $\pi_0^J(m_1, m_2) < \pi_0^S(m_1, m_2)$. □

Proof of Lemma 7: Parts 1 and 2. Observe that $\sigma^2(m, n^*(m); p) = \left(\frac{c_f + c_v m}{B}\right)\left(\frac{\sigma^2(m,n;p)}{n}\right)$. Then, the proof follows directly from Lemma 6.

Part 3. $\sigma^2(m, n^*(m); p)$ can be expressed as follows:

$$\begin{aligned}\sigma^2(m, n^*(m); p) &= \frac{1}{B}\left\{c_f\left[\frac{1-(1-p)^m}{m^2(1-p)^{m-2}}\right] + c_v\left[\frac{1-(1-p)^m}{m(1-p)^{m-2}}\right]\right\} \\ &= \frac{1}{B}\left[\frac{1}{(1-p)^{m-2}} - (1-p)^2\right]\left(\frac{c_f}{m^2} + \frac{c_v}{m}\right).\end{aligned}$$

Let $g(m; p) \equiv B\sigma^2(m, n^*(m); p)$. We have the following:

$$\frac{\partial}{\partial m} g(m; p) = \left[-\frac{\log(1-p)}{(1-p)^{m-2}}\right]\left(\frac{c_f}{m^2} + \frac{c_v}{m}\right) + \left[\frac{1}{(1-p)^{m-2}} - (1-p)^2\right]\left(\frac{-2c_f}{m^3} - \frac{c_v}{m^2}\right), \text{ and}$$

$$\begin{aligned}\frac{\partial^2}{\partial m^2} g(m; p) = &\left[\frac{\left[\log(1-p)\right]^2}{(1-p)^{m-2}}\right]\left(\frac{c_f}{m^2} + \frac{c_v}{m}\right) + 2\left[\frac{\log(1-p)}{(1-p)^{m-2}}\right]\left(\frac{2c_f}{m^3} + \frac{c_v}{m^2}\right) \\ &+ \left[\frac{1}{(1-p)^{m-2}} - (1-p)^2\right]\left(\frac{6c_f}{m^4} + \frac{2c_v}{m^3}\right).\end{aligned}$$

We now show that $\frac{\partial^2}{\partial m^2} g(m; p) > 0$, $\forall m > 0$ and $\forall p \in (0, \frac{1}{3})$. Noting that $(1-p)^{m-2} > 0$, $\forall m > 0$ and $p \in (0, \frac{1}{3})$, and multiplying $\frac{\partial^2}{\partial m^2} g(m; p)$ by $\left\{(1-p)^{m-2}\right\}$ yields the following :

$$h(m; p) = \Big(\log(1-p)\Big)^2\left(\frac{c_f}{m^2} + \frac{c_v}{m}\right) + 2\log(1-p)\left(\frac{2c_f}{m^3} + \frac{c_v}{m^2}\right) + \left[1-(1-p)^m\right]\left(\frac{6c_f}{m^4} + \frac{2c_v}{m^3}\right).$$

$$\frac{\partial}{\partial p} h(m; p) = \frac{-2\log(1-p)}{1-p}\left(\frac{c_f}{m^2} + \frac{c_v}{m}\right) - \frac{2}{1-p}\left(\frac{2c_f}{m^3} + \frac{c_v}{m^2}\right) + (1-p)^{m-1}\left(\frac{6c_f}{m^3} + \frac{2c_v}{m^2}\right).$$

Note that the sign of $\frac{\partial}{\partial p} h(m; p)$ equals the sign of the following function:

$$-2\log(1-p)\left(\frac{c_f}{m^2}+\frac{c_v}{m}\right)-2\left(\frac{2c_f}{m^3}+\frac{c_v}{m^2}\right)+(1-p)^m\left(\frac{6c_f}{m^3}+\frac{2c_v}{m^2}\right)$$
$$=-\frac{2\log(1-p)c_v}{m}+\frac{1}{m^2}\Big(-2c_f\log(1-p)-2c_v+2c_v(1-p)^m\Big)+\frac{c_f}{m^3}\Big(-4+6(1-p)^m\Big)>0,$$

where the last inequality follows because $c_f > c_v$ and $\log(1-p) < 0$, $\forall p \in \left(0, \frac{1}{3}\right)$. Thus, $\frac{\partial^2}{\partial m^2} g(m;p)$ is increasing in p, $\forall m > 0$ and $\forall p \in \left(0, \frac{1}{3}\right)$. At $p = 0$, $\frac{\partial^2}{\partial m^2} g(m;p) = \frac{6c_f}{m^4} + \frac{2c_v}{m^3} > 0$. Thus, $\frac{\partial^2}{\partial m^2} g(m;p) > 0$, $\forall m > 0$ and $\forall p \in \left(0, \frac{1}{3}\right)$. This completes the proof. □

Proof of Theorem 1: By Properties 1 and 2, and Lemmas 1 and 3, we have that $\pi_0^X(m, m+1) \leq \pi_0^X(m-1, m)$, $\forall m \in \mathbb{Z}^+ : m > 2$, $X \in \{S, J\}$. Then, we have that $\sigma^2(m, n; p) < \sigma^2(k, n; p)$, $\forall k \in \mathbb{Z}^+ : k \neq m$, if and only if $\pi_0^X(m, m+1) < p < \pi_0^X(m-1, m)$. If $p = \pi_0^X(m, m+1)$, then $\sigma^2(m+1, n; p) = \sigma^2(m, n; p)$, and, if $p = \pi_0^X(m-1, m)$, then $\sigma^2(m, n; p) = \sigma^2(m-1, n; p)$.

As a result, we have the following:

$\cdots\ \pi_0^S(m+1, m+2) < p < \pi_0^S(m, m+1)$	$\pi_0^S(m, m+1) < p < \pi_0^S(m-1, m)$	$\pi_0^S(m-1, m) < p < \pi_0^S(m-2, m-1)\ \cdots$
$\sigma^2(m+1, n; p) < \sigma^2(k, n; p)$, $\forall k \in \mathbb{Z}^+ : k \neq m+1$	$\sigma^2(m, n; p) < \sigma^2(k, n; p)$, $\forall k \in \mathbb{Z}^+ : k \neq m$	$\sigma^2(m-1, n; p) < \sigma^2(k, n; p)$, $\forall k \in \mathbb{Z}^+ : k \neq m-1$.

Then, if $\pi_0^S(m, m+1) < p < \pi_0^S(m-1, m)$, then m is the unique optimal solution to *D-S*. If $p = \pi_0^S(m, m+1)$, then m and $m+1$ are both optimal, and, if $p = \pi_0^S(m-1, m)$, then $m-1$ and m are both optimal for *D-S*.

For *D-J*, we have the following:

$\cdots\ \pi_0^J(m+1, m+2) < p < \pi_0^J(m, m+1)$	$\pi_0^J(m, m+1) < p < \pi_0^J(m-1, m)$	$\pi_0^J(m-1, m) < p < \pi_0^J(m-2, m-1)\ \cdots$
$\sigma^2(m+1, n*(m+1); p) < \sigma^2(k, n^*(k); p)$, $\forall k \in \mathbb{Z}^+ : k \neq m+1$	$\sigma^2(m, n^*(m); p) < \sigma^2(k, n^*(k); p)$, $\forall k \in \mathbb{Z}^+ : k \neq m$	$\sigma^2(m-1, n^*(m-1); p) < \sigma^2(k, n^*(k); p)$, $\forall k \in \mathbb{Z}^+ : k \neq m-1$.

Then, if $\pi_0^J(m, m+1) < p < \pi_0^J(m-1, m)$, then m is the unique optimal solution to *D-J*. If $p = \pi_0^J(m, m+1)$, then m and $m+1$ are both optimal, and, if $p = \pi_0^J(m-1, m)$, then $m-1$ and m are both optimal for *D-J*. This completes the proof. □

Proof of Lemma 5: By Lemmas 6 and 7, m' is the unique solution to the first-order condition:

$$\frac{-2}{(m')^3(1-p_0)^{m'-2}} - \frac{\log(1-p_0)}{(m')^2(1-p_0)^{m'-2}} + \frac{2(1-p_0)^2}{(m')^3} = 0, \text{ for } D\text{-}S, \text{ and}$$
$$\left(\frac{c_f}{(m')^2} + \frac{c_v}{m'}\right)\left[-\frac{\log(1-p_0)}{(1-p_0)^{m'-2}}\right] - \left(\frac{2c_f}{(m')^3} + \frac{c_v}{(m')^2}\right)\left[\frac{1}{(1-p_0)^{m'-2}} - (1-p_0)^2\right] = 0, \text{ for } D\text{-}J,$$

which are respectively equivalent to:

$$m' = \frac{2[(1-p_0)^{m'}-1]}{\log(1-p_0)}, \text{ for } D\text{-}S, \text{ and}$$
$$m' = \frac{\left(1+\frac{c_f}{c_f+c_v m'}\right)[(1-p_0)^{m'}-1]}{\log(1-p_0)}, \text{ for } D\text{-}J.$$

This completes the proof. □

Proof of Theorem 2: From Lemmas 6 and 7, since $\sigma^2(m,n;p)$ and $\sigma^2(m,n^*(m);p)$ are increasing in p, $\max_{p\in[p_{LB},p_{UB}]}\left\{\sigma^2(m,n;p)\right\} \equiv \sigma^2(m,n;p_{UB})$, and $\max_{p\in[p_{LB},p_{UB}]}\left\{\sigma^2(m,n^*(m);p)\right\} \equiv \sigma^2(m,n^*(m);p_{UB})$. Thus, the results follow. □

Proof of Corollary 1: The results for *D-X*, $X \in \{S,J\}$, follow from Theorem 1; and the results for *M-X*, $X \in \{S,J\}$, follow from Theorem 2. □

Proof of Corollary 2: Recall that $\gamma = \frac{c_f}{c_v}$. From Lemma 3, we have the following:

$$\frac{m_2^2(1-\pi_0^J(m_1,m_2))^{m_2}\left[1-(1-\pi_0^J(m_1,m_2))^{m_1}\right]}{m_1^2(1-\pi_0^J(m_1,m_2))^{m_1}\left[1-(1-\pi_0^J(m_1,m_2))^{m_2}\right]} = \frac{c_f+c_v m_2}{c_f+c_v m_1} = \frac{\gamma+m_2}{\gamma+m_1} = 1+\frac{(m_2-m_1)}{\gamma+m_1}.$$

Note that $\frac{m_2^2(1-\pi_0^J(m_1,m_2))^{m_2}\left[1-(1-\pi_0^J(m_1,m_2))^{m_1}\right]}{m_1^2(1-\pi_0^J(m_1,m_2))^{m_1}\left[1-(1-\pi_0^J(m_1,m_2))^{m_2}\right]}$ is decreasing in $\pi_0^J(m_1,m_2)$ (from Liu *et al* [37]) and is constant in γ, while $1+\frac{(m_2-m_1)}{\gamma+m_1}$ is decreasing in γ (since $m_2 > m_1$) and is constant in $\pi_0^J(m_1,m_2)$. Thus, as γ increases, $\pi_0^J(m_1,m_2)$ also increases, and $\pi_0^S(m_1,m_2)$ is independent of γ. The results then follow from Theorems 1 and 2, and Corollary 1. □