Published in final edited form as:

Biometrics. 2009 September ; 65(3): 903–910. doi:10.1111/j.1541-0420.2008.01158.x.

Three-Dimensional Array-Based Group Testing Algorithms

Hae-Young Kim^{1,*} and Michael G. Hudgens²

¹ Center for Statistical Analysis and Research, New England Research Institutes, 9 Galen Street, Watertown, Massachusetts 02472, U.S.A.

² Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, 3107-E McGavran-Greenberg Hall, Chapel Hill, North Carolina 27599, U.S.A.

Summary

We derive the operating characteristics of three-dimensional array-based testing algorithms for case identification in the presence of testing error. The operating characteristics investigated include efficiency (i.e., expected number of tests per specimen) and error rates (e.g., sensitivity, specificity, positive, and negative predictive values). The methods are illustrated by comparing the proposed algorithms with previously studied hierarchical and two-dimensional array algorithms for detecting recent HIV infections in North Carolina. Our results indicate that three-dimensional array-based algorithms can be more efficient and accurate than previously proposed algorithms in settings with test error and low prevalence.

Keywords

Array; Group testing; Hierarchical; HIV; Predictive value; Sensitivity; Specificity

1. Introduction

Pooling of serum samples for screening was first employed to detect syphilis in U.S. soldiers during the Second World War (Dorfman, 1943). It has subsequently been used as a method for reducing the costs of screening tests for many other infectious diseases; for a recent discussion see Kim et al. (2007). In addition to increasing efficiency, specimen pooling (or group testing) has been shown to reduce rates of misclassification. For example, in programs designed to detect “acute” or recent HIV infections, nucleic acid amplification tests in conjunction with specimen pooling have been shown empirically (Quinn et al., 2000; Pilcher et al., 2005) and theoretically (Kim et al., 2007) to substantially improve efficiency, specificity, and positive predictive value over individual testing. In this article, we consider the utility of array-based group testing algorithms in similar settings, where selecting an appropriate algorithm requires consideration of efficiency as well as rates of misclassification. This work is motivated by North Carolina Screening and Tracing Active Transmission (NC STAT), an acute HIV detection program employed by the North Carolina Department of Public Health (Pilcher et al., 2005). NC STAT uses robotic pooling to process approximately 120,000 specimens per year. The availability of automated pooling makes the implementation of array-based group testing algorithms feasible in this setting. In such high-throughput settings, even slight improvements in efficiency can lead to substantial cost savings.

*hkim@neriscience.com.

⁷Supplementary Materials

Web Appendices, Table, and Figures referenced in Sections 3, 4, and 5 are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

Array-based specimen pooling is a group testing algorithm that uses overlapping pools. Historically, this approach has been employed in genetics more than in the infectious disease setting. In the simple two-dimensional form, n^2 specimens are placed on an $n \times n$ matrix. Pools of size n are constructed from all samples in the same row or in the same column. These $2n$ pools are then tested such that all positive specimens will lie at the intersection of a positive row pool and a positive column pool (assuming no false negative tests). Specimens at these intersections are then tested to resolve any ambiguities. Phatarfod and Sudbury (1994), Hedt and Pagano (2008a,b), and Kim et al. (2007) derived the operating characteristics of two-dimensional array-based testing algorithms. Berger, Mandell, and Subrahmanya (2000) considered higher-dimensional arrays in the absence of test error. The focus of this article is to research aspects of three-dimensional array-based group testing algorithms for case identification in the presence of test error. This work can be considered an extension of Kim et al. (2007) to three dimensions and of Berger et al. (2000) to allow for imperfect testing.

The outline of this article is as follows. In Section 2, we define notation, model assumptions, and operating characteristics to be derived. In Section 3, we present four different three-dimensional array-based pooling algorithms; operating characteristics of these algorithms are derived in the Web Appendix. In Section 4, comparisons are made between the proposed algorithms and previously studied hierarchical and two-dimensional array algorithms for detecting recent HIV infections. In Section 5, we present results indicating arrays of dimensions greater than three do not lead to further appreciable gains in efficiency in low prevalence settings where there are restrictions on the number of specimens available. We conclude with a short discussion in Section 6.

2. Preliminaries

2.1 Notation

Suppose that we have $L \times M \times N$ specimens where L , M , and N are positive integers. Let X_{i_1, i_2, i_3} be the random variable corresponding to the test outcome if specimen (i_1, i_2, i_3) is tested individually (i.e., without pooling) for $i_1 = 1, \dots, L$; $i_2 = 1, \dots, M$; and $i_3 = 1, \dots, N$. Let $X_{i_1, i_2, i_3} = 1$ if the specimen would test positive and 0 otherwise. Likewise let Y_{i_1, i_2, i_3} indicate the true status of specimen (i_1, i_2, i_3) , i.e., if tested individually in the absence of testing error. Now imagine the specimens have been arranged in an $L \times M \times N$ cube. For $i_1 = 1, \dots, L$, let X_{i_1++} denote the test outcome for the pool of size MN corresponding to the i_1 th planar slice from front to back. Define X_{+i_2+} for $i_2 = 1, \dots, M$ and X_{++i_3} for $i_3 = 1, \dots, N$ similarly. Denote the corresponding true values by Y_{i_1++} , Y_{+i_2+} , and Y_{++i_3} .

2.2 Assumptions

Here we define the key assumptions used to derive operating characteristics of the three-dimensional array-based pooling algorithms considered. These assumptions are analogous to those used by Kim et al. (2007) in the two-dimensional array setting.

Assumption 1—All specimens are independent and identically distributed with probability p of being positive.

We refer to p as the *prevalence* and let $q = 1 - p$.

Assumption 2—Given a pool $\mathcal{P}$ containing at least one positive specimen is tested, the probability $\mathcal{P}$ tests positive equals S_e .

We refer to S_e as the *test sensitivity*. Assumption 2 implies that the test sensitivity is independent of the number of specimens within a pool and the number of positive specimens therein.

Assumption 3—Given a pool $\mathcal{P}$ containing no positive specimen is tested, the probability $\mathcal{P}$ tests positive equals $1 - S_p$.

We refer to S_p as the *test specificity*. Assumption 3 implies test specificity is independent of pool size.

Assumption 4—Given the true status of any pool $\mathcal{P}$, the test result for $\mathcal{P}$ is independent of the true status and test result of any other pool.

For instance, under Assumption 4, $\Pr[X_{i_1++} = x_1, X_{i_2+} = x_2 | Y_{i_1++} = y_1, Y_{i_2+} = y_2] = \Pr[X_{i_1++} = x_1 | Y_{i_1++} = y_1] \Pr[X_{i_2+} = x_2 | Y_{i_2+} = y_2]$ for any $x_1, x_2, y_1, y_2 \in \{0, 1\}$.

2.3 Operating Characteristics

As in Kim et al. (2007), we define the following operating characteristics (in *italics*) for an arbitrary testing algorithm $\mathcal{A}$. *Efficiency*, denoted $E(\mathcal{A})$, is the expected number of tests per specimen for algorithm $\mathcal{A}$ to classify all specimens as positive or negative. *Pooling sensitivity*, denoted $S_e(\mathcal{A})$, is the probability an individual is categorized as positive by $\mathcal{A}$ given that individual is truly positive. *Pooling specificity*, denoted $S_p(\mathcal{A})$, is the probability an individual is categorized as negative by $\mathcal{A}$ given that individual is truly negative. *Pooling positive predictive value*, denoted $PPV(\mathcal{A})$, is the probability an individual is truly positive given they are categorized as positive by $\mathcal{A}$. *Pooling negative predictive value*, denoted $NPV(\mathcal{A})$, is the probability an individual is truly negative given they are categorized as negative by $\mathcal{A}$. The predictive values $PPV(\mathcal{A})$ and $NPV(\mathcal{A})$ can be expressed as simple functions of p , $S_p(\mathcal{A})$ and $S_e(\mathcal{A})$ (Kim et al., 2007).

3. Three-Dimensional Array-Based Algorithms

In this section, we present four different three-dimensional array-based pooling algorithms. The first two algorithms ($A3P$ and $A3P_2$) are two-stage algorithms, with planar slice pools tested in the first stage and individual specimens tested at the second stage. The other two algorithms ($A3PM$ and $A3PM_2$) are three-stage algorithms, which are analogous to $A3P$ and $A3P_2$, but entail first testing a master pool. Derivations of algorithm efficiencies, sensitivities, and specificities are presented in the Web Appendix.

3.1 A3P: Without Master Pool

First we consider a two-stage three-dimensional array-based testing algorithm. This algorithm entails planar slices of a three-dimensional array and thus is denoted $A3P$ ($[L, M, N]: 1$). For this method, in stage 1, each of the L planar slices from front to back, the M planar slices from top to bottom, and the N planar slices from left to right are tested. In stage 2, a specimen is tested individually if either (i) all three planar slices containing that specimen test positive or (ii) two of the three planar slices containing that specimen test positive and all planar slices in the remaining dimension test negative. If we let $T_{(i_1, i_2, i_3)}^{A3P}$ be an indicator variable for when the (i_1, i_2, i_3) specimen is tested individually under $A3P$, then

$$T_{(i_1, i_2, i_3)}^{A3P} = \begin{cases} 1 & \text{if } X_{i_1++} = 1 \text{ and } X_{+i_2+} = 1 \text{ and } X_{++i_3} = 1 \\ 1 & \text{if } X_{i_1++} = 1 \text{ and } X_{+i_2+} = 1 \text{ and } \sum_{i=1}^N X_{++i} = 0 \\ 1 & \text{if } X_{i_1++} = 1 \text{ and } \sum_{i=1}^M X_{+i+} = 0 \text{ and } X_{++i_3} = 1 \\ 1 & \text{if } \sum_{i=1}^L X_{i++} = 0 \text{ and } X_{+i_2+} = 1 \text{ and } X_{++i_3} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

In the absence of test error, two intersecting planar slices that test positive imply the existence of a third planar slice which is positive and intersects with the first two planar slices. Thus scenario (ii) will only arise in the setting where there is test error. In this case, two intersecting planar slices testing positive provides evidence suggesting at least one specimen is positive, whereas all planar slices in the remaining dimension testing negative suggests all specimens are negative. To resolve this contradictory information, under scenario (ii) algorithm $A3P$ tests all specimens at the intersection of any two planar slices that test positive.

3.2 $A3P_2$: Alternative to $A3P$

We also consider an alternative algorithm to $A3P$, which we denote as $A3P_2$. The difference between $A3P$ and $A3P_2$ is in the second stage, i.e., in the circumstances under which an individual sample is tested. In particular, let $T_{(i_1, i_2, i_3)}^{A3P_2}$ be an indicator variable for when the (i_1, i_2, i_3) specimen is tested individually under $A3P_2$, defined as

$$T_{(i_1, i_2, i_3)}^{A3P_2} = \begin{cases} 1 & \text{if } T_{(i_1, i_2, i_3)}^{A3P} = 1 \\ 1 & \text{if } X_{i_1++} = 1 \text{ and } \sum_{i=1}^M X_{+i+} = 0 \text{ and } \sum_{i=1}^N X_{++i} = 0 \\ 1 & \text{if } \sum_{i=1}^L X_{i++} = 0 \text{ and } X_{+i_2+} = 1 \text{ and } \sum_{i=1}^N X_{++i} = 0 \\ 1 & \text{if } \sum_{i=1}^L X_{i++} = 0 \text{ and } \sum_{i=1}^M X_{+i+} = 0 \text{ and } X_{++i_3} = 1 \\ 0 & \text{otherwise} \end{cases}$$

In other words, for the (i_1, i_2, i_3) specimen to be tested individually under $A3P_2$ it is necessary that at least one planar slice containing that specimen be positive, whereas under $A3P$ it is necessary that at least two planar slices containing that specimen be positive. Intuitively, $A3P_2$ would be expected to be less efficient and specific but more sensitive than $A3P$, because individual specimens will be tested more often under $A3P_2$ than $A3P$.

3.3 $A3PM$ and $A3PM_2$: With Master Pool

We also consider three-stage three-dimensional array-based testing algorithms where we first test a master pool containing all LMN samples. First, we define algorithm $A3PM$ ($LMN: [L, M, N]: 1$) as follows. In the first stage, the master pool is tested. If the master pool tests negative, the procedures stop. Otherwise, the procedures continue as in $A3P$. Second, we define algorithm $A3PM_2$ ($LMN: [L, M, N]: 1$) to be the same as $A3PM$, except that if the master pool tests positive, the procedure continues as in $A3P_2$.

4. Application

In this section, we compare the operating characteristics of the three-dimensional array-based algorithms described above with previously studied group testing algorithms for identification of acute HIV. The comparative algorithms include the two-dimensional array algorithms $A2P$ and $A2PM$ and a three-stage hierarchical algorithms $D3$ as described in Kim et al. (2007). Briefly, $A2P$ and $A2PM$ are the two-dimensional analogues of $A3P$ and $A3PM$. $D3$ ($n_1:n_2:1$) corresponds to a three-stage hierarchical algorithm where first a master pool of size n_1 is tested; if the master pool tests positive, n_1/n_2 non-overlapping subpools of size n_2 are tested; finally all specimens within subpools that test positive are tested individually.

As a motivating example, we consider a setting similar to NC STAT. We assume prevalence of acute HIV is $p = 0.0002$ (Pilcher et al., 2005) and nucleic acid amplification test has a 99% test specificity (Hecht et al., 2002) and 90% test sensitivity. We also assume that the number of specimens required in the first stage of any algorithm can be no more than 100. This restriction on the maximum allowable batch size serves two purposes. First, in detection of acute HIV, it is important that relatively small batches of specimens be processed to allow timely identification of cases. Second, for algorithms with master pool testing ($D3$, $A2PM$, $A3PM$, $A3PM_2$), this assumptions also guards against dilution effects, i.e., decreases in sensitivity to detect positive specimens in large pools composed primarily of negative specimens.

Under these assumptions, for each algorithm considered the optimal configuration was selected that minimizes the expected number of tests per specimen. For instance, the optimal configuration of $A3PM$ was determined by computing the efficiency for all possible positive integers (L, M, N) such that $8 \leq LMN \leq 100$. For $A3PM$, the most efficient configuration is $(L, M, N) = (4, 5, 5)$; this is also the optimal configuration for $A3PM_2$. Similarly, $D3(100:10:1)$ and $A2PM(100:[10, 10]:1)$ are the optimal configurations of $D3$ and $A2PM$.

Table 1 shows the operating characteristics of the optimal configurations of each algorithm as well as $D3(90:10:1)$, the algorithm employed by Pilcher et al. (2005). These results suggest moving from $D3(90:10:1)$ to $A3PM(100:[4,5,5]:1)$ or $A3PM_2(100:[4, 5, 5]:1)$ would improve efficiency, pooling specificity, sensitivity, PPV and NPV of the NC STAT HIV detection program. Moving from $D3(90:10:1)$ to $A3PM(100:[4,5,5]:1)$ would decrease the average number of tests per specimen from 0.016 to 0.014. Given that the NC STAT program processes 120,000 specimens per year, this improvement in efficiency would translate into a decrease of 240 tests per year, a considerable savings given individual tests can cost \$60. This change in pooling algorithm would also result in a 10% increase in PPV , reducing the cost of retesting individual specimens identified as positive by the pooling algorithm. Likewise, the expected number of false negative classifications (or type II per-family error rate [Kim et al., 2007]) would decrease from 6.5 to 5.3, indicating on average 1.2 additional acute HIV cases would be detected each year in North Carolina.

Next we investigate the effect of the assumed values of p , S_p , S_e , and maximum allowable pool size on the relative performance of the different algorithms under consideration. First we consider p in the range of 2×10^{-5} to 2×10^{-3} for fixed $S_e = 0.9$, $S_p = 0.99$, and maximum allowable batch size equal to 100. This range of prevalence includes settings where the prevalence is an order of magnitude higher or lower than that observed by the NC STAT program. For each prevalence we found the optimally efficient configuration of the algorithms. The expected numbers of tests per specimen, pooling sensitivity, pooling PPV , and pooling NPV of the optimal configurations of $D3$, $A2PM$, $A3PM$, and $A3PM_2$ are depicted in Figure 1. $A2P$, $A3P$, and $A3P_2$ are not shown because these algorithms tend to be substantially less efficient than those displayed. Pooling specificities are not displayed as these values tend to

be very close to 1. Similar to Figure 1, the effects of varying S_e, S_p , and the maximum allowable pool size are demonstrated in Figures 2–4. In total, these results support *A3PM* as the preferred algorithm with regard to efficiency and *PPV* in settings similar to the NC STAT program. In such settings, *A2PM*, and *A3PM*₂ tend to have slightly higher pooling sensitivity than *A3PM* while all three algorithms have pooling *NPV* near 1.

The results above are motivated by the NC STAT program where the prevalence of acute HIV is relatively low. The Web Appendix includes analogous results for a high-prevalence setting. In particular, we let $p = 0.03$, the prevalence of acute HIV among antibody negative men observed in Malawi by Pilcher, Price, et al. (2004). The results given in Web Table 1 and Web Figures 1–4 indicate that in higher-prevalence settings three dimensional array-based algorithms are still relatively efficient but may be of less utility due to lower pooling sensitivities compared to *D3* or *A2P*.

5. Higher-Dimensional Arrays

It is natural to consider whether extending the algorithms defined in Section 3 to higher dimensions (i.e., of dimensions greater than three) leads to additional improvement in efficiency or rates of misclassification. Deriving the operating characteristics of higher-dimensional array-based algorithms in the presence of test error, although tedious, should be straightforward. In the absence of test error, one can write down the efficiencies immediately. In particular, consider *ASPM*, the S -dimensional generalization of *A3PM*, where S is some positive integer greater than one. Let L be a vector of length S , where the i th component L_i denotes the size of the i th dimension of the array. The first stage of *ASPM* entails testing a

collection of $\prod_{i=1}^S L_i$ specimens in a master pool; if the master pool tests positive, pools from the $\sum_{i=1}^S L_i$ hyperplanar slices are tested and subsequently specimens at the intersection of S positive hyperplanar slice pools are tested individually. The efficiency of *ASPM* can be found using the method of inclusion and exclusion (Feller, 1968). For example, the derivation of *E* (*A4PM*) is given in Web Appendix C.

Therefore, we can conduct efficiency comparisons of higher-dimensional arrays in the absence of test error in settings similar to the NC STAT program, i.e., where the prevalence of disease is low ($p \in [2 \times 10^{-5}, 2 \times 10^{-3}]$) and the maximum allowable batch size (due to dilution effects and throughput considerations) is in the range of 75 to 125. Similar to Figure 1, for a given prevalence and maximum allowable batch size, the configuration of a particular algorithm was determined that minimized expected number of tests per specimen. For example, when $p = 0.0002$ and pool sizes are limited to no more than 100, the optimal configurations of *A3PM* and *A4PM* are 100:[4,5,5]:1 and 100:[2,2,5,5]:1, both with efficiencies of 0.013. The expected number of tests per specimen for optimally configured array-based algorithms are given in Web Figure 5 as a function of prevalence and maximum allowable batch size. Because the maximum allowable batch size is no more than 125, arrays of dimension greater than six were not considered (because $2^7 > 125$). Web Figure 5 demonstrates that *A3PM* and *A4PM* are nearly equivalent and that either would be preferred over *A2PM*, *A5PM*, and *A6PM*. These results suggest that higher-dimensional arrays will not lead to appreciable gains in efficiency over *A3PM* in settings similar NC STAT.

6. Discussion

This article adds to a growing body of work indicating that array-based group-testing algorithms are an attractive alternative to traditional hierarchical algorithms. That array-based algorithms are efficient has been established previously. In particular, Berger et al. (2000) show that multidimensional array-based algorithms are optimally efficient among all two-stage

algorithms. However, the practical implications of their results are not immediately clear because (i) they assume no test error and (ii) achieving optimality using their algorithm can require an infeasible number of specimens. For example, at a prevalence of $p = 0.01$, their optimally efficient algorithm requires $74^6 \approx 164$ billion specimens at the first stage. While they indicate (without proof) that the same efficiency can be achieved using 4600 specimens, the utility of array-based algorithms where the maximum allowable batch size is on the order of 100 is not clear from their work. More recently, Kim et al. (2007) show two-dimensional arrays can be as efficient and more accurate than three-stage hierarchical algorithms in settings where prevalence is low, test error is present, and the batch size is limited (e.g., as in the NC STAT program). In this work, we show that in such settings additional gains in efficiency and accuracy are possible by moving to three-dimensional arrays. Our findings also indicate arrays of dimension greater than three do not lead to further appreciable gains in efficiency.

The accuracy of the predicted efficiencies and error rates presented here depend on the veracity of Assumptions 1–4. Assumption 1 is a standard assumption made in the group testing literature and should hold provided the individuals being tested approximate a random sample from a larger population. Assumptions 2 and 3 should hold provided the total number of specimens per pool is not too large. These assumptions can be tested by conducting specimen-pooling experiments using only negative and positive controls. Alternatively, one could relax these assumptions by considering models that allow test sensitivity or test specificity to depend on pool size, e.g., see Johnson, Kotz and Wu (1991). One could also test Assumption 4 by conducting experiments using negative and positive controls because both the true status of the planar slice pools as well as the corresponding test outcomes are observable in such experiments. In practice, this assumption might be unreasonable in settings where pooling error or contamination is possible.

In addition to the operating characteristics of different group testing algorithms, consideration must be given to the feasibility of adopting a particular algorithm. For detection of acute HIV in North Carolina, robotic pooling is employed (Pilcher et al., 2005), such that implementation of complex group testing algorithms is feasible. Automated pooling is also used in many other settings, such as in screening millions of blood donations in Europe and Japan for HIV and hepatitis (Roth et al., 2002; Mine et al., 2003). In settings where automated pooling is not possible, multidimensional array-based algorithms can still, in principle, be implemented by hand. However the complexity of these algorithms will likely increase the potential for human error compared to simpler algorithms. Thus, in manual pooling settings, the efficiency gains afforded by array-based algorithms must be weighed against this potential increase in error.

Further consideration might be given to extensions of the algorithms considered here where pools which are negative are retested. Litvak, Tu, and Pagano (1994) showed retesting negative pools can increase the sensitivity of hierarchical group-testing algorithms. Similarly, Hedt and Pagano (2008a,b) proposed a retesting extension of Phatarfod and Sudbury's (1994) matrix algorithm. There are two reasons why we do not consider such extensions in this work. First, retesting negative pools would increase the expected number of tests per specimen. As discussed above, efficiency is a critical parameter in high-throughput settings such as the NC STAT program. Second, retesting negative pools would result in an additional stage of testing, thus increasing the turn-around time in processing individual specimens. In the context of acute HIV detection, minimizing turn-around time is imperative (Pilcher, Eron, et al., 2004). In other contexts, the increased sensitivity afforded by retesting negatives may be worth the trade-off of decreased efficiency and increased turn-around time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health grants P30 AI50410-07 and R03 AI068450-01.

References

- Berger T, Mandell JW, Subrahmanya P. Maximally efficient two-stage screening. *Biometrics* 2000;56:833–840. [PubMed: 10985223]
- Dorfman R. The detection of defective numbers of large populations. *Annals of Mathematical Statistics* 1943;14:436–440.
- Feller, W. *An Introduction to Probability Theory and Its Applications*. 3. New York: John Wiley & Sons; 1968.
- Hecht F, Busch M, Rawal Webb M, Rosenberg E, Swanson M, Chesney M, Anderson J, Levy J, Kahn J. Use of laboratory tests and clinical symptoms for identification of primary HIV infection. *AIDS* 2002;16:1119–1129. [PubMed: 12004270]
- Hedt, BL.; Pagano, M. A matrix pooling algorithm for disease detection. Working Paper 57, Harvard University Biostatistics Working Paper Series. 2008a. Available from <http://www.bepress.com/harvardbiostat/paper57>
- Hedt, BL.; Pagano, P. Matrix pooling: An accurate and cost effective testing algorithm for detection of acute HIV infection. Working Paper 58, Harvard University Biostatistics Working Paper Series. 2008b. Available from <http://www.bepress.com/harvardbiostat/paper58>
- Johnson, NL.; Kotz, S.; Wu, X. *Inspection Errors for Attributes in Quality Control*. New York: Chapman and Hall; 1991.
- Kim HY, Hudgens MG, Dreyfuss J, Westreich D, Pilcher CD. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 2007;63:1152–1163. [PubMed: 17501946]
- Litvak E, Tu XM, Pagano M. Screening for the presence of a disease by pooling sera samples. *Journal of the American Statistical Association* 1994;89:424–434.
- Mine H, Emura H, Miyamoto M, Tomono T, Minegishi K, Murokawa H, Yamanaka R, Yoshikawa A, Nishioka K. Japanese Red Cross NAT Research Group. High throughput screening of 16 million serologically negative blood donors for hepatitis B virus, hepatitis C virus and human immunodeficiency virus type-1 by nucleic acid amplification testing with specific and sensitive multiplex reagent in Japan. *Journal of Virological Methods* 2003;112:145–151. [PubMed: 12951223]
- Phatarfod RM, Sudbury A. The use of a square array scheme in blood testing. *Statistics in Medicine* 1994;13:2337–2343. [PubMed: 7855467]
- Pilcher CD, Price MA, Hoffman IF, Galvin S, Martinson FE, Kazembe PN, Eron JJ, Miller WC, Fiscus SA, Cohen MS. Frequent detection of acute primary HIV infection in men in Malawi. *AIDS* 2004;18:517–524. [PubMed: 15090805]
- Pilcher CD, Eron JJ, Galvin S, Gay C, Cohen MS. Acute HIV revisited: New opportunities for treatment and prevention. *Journal of Clinical Investigation* 2004;113:937–945. [PubMed: 15057296]
- Pilcher CD, Fiscus SA, Nguyen TQ, Foust T, Wolf L, Williams D, Ashby R, O'Dowd JO, McPherson JT, Stalzer B, Hightow L, Miller WC, Eron JJ, Cohen MS. Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine* 2005;352:1873–1883. [PubMed: 15872202]
- Quinn TC, Brookmeyer R, Kline R, Shepherd M, Paranjape R, Mehendale S, Gadkari DA, Bollinger R. Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence. *AIDS* 2000;14:2751–2757. [PubMed: 11125894]
- Roth W, Weber M, Buhr S, Drosten C, Weichert W, Sireis W, Hedges D, Seifried E. Yield of HCV and HIV-1 NAT after screening of 3.6 million blood donations in central Europe. 2002

Figure 1.

Operating characteristics for optimally efficient configurations of $D3$, $A2PM$, $A3PM$, and $A3PM_2$ as a function of prevalence p assuming a maximum allowable batch size of 100, test sensitivity $S_e = 0.9$, and test specificity $S_p = 0.99$.

Figure 2. Operating characteristics for optimally efficient configurations of $D3$, $A2PM$, $A3PM$, and $A3PM_2$ as a function of test sensitivity S_e assuming a maximum allowable batch size of 100, prevalence $p = 0.0002$, and test specificity $S_p = 0.99$.

Figure 3. Operating characteristics for optimally efficient configurations of $D3$, $A2PM$, $A3PM$, and $A3PM_2$ as a function of test specificity S_p assuming a maximum allowable batch size of 100, prevalence $p = 0.0002$, and test sensitivity $S_e = 0.9$.

Figure 4. Operating characteristics for optimally efficient configurations of $D3$, $A2PM$, $A3PM$, and $A3PM_2$ as a function of the maximum allowable batch size assuming prevalence $p = 0.0002$, test sensitivity $S_e = 0.9$, and test specificity $S_p = 0.99$.

Table 1

Comparison of operating characteristics for optimally efficient configurations of D3, A2P, A2PM, A3P, A3P₂, A3PM, and A3PM₂ assuming a maximum allowable batch size of 100, test sensitivity $S_e = 0.9$, test specificity $S_p = 0.99$, and prevalence $p = 0.0002$

$\mathcal{A}$	$E(\mathcal{A})$	$S_p(\mathcal{A})$	$S_e(\mathcal{A})$	$PPV(\mathcal{A})$	$NPV(\mathcal{A})$
$D3(90:10:1)^*$	0.016	0.999983	0.729	0.896	0.999946
$D3(100:10:1)$	0.015	0.999983	0.729	0.895	0.999946
$A2P([10, 10]:1)$	0.218	0.999817	0.875	0.488	0.999975
$A2PM(100:[10, 10]:1)$	0.016	0.999995	0.787	0.969	0.999957
$A3P([4, 5, 5]:1)$	0.141	0.999995	0.864	0.972	0.999973
$A3P_2([4, 5, 5]:1)$	0.168	0.999728	0.886	0.394	0.999977
$A3PM(100:[4, 5, 5]:1)$	0.014	0.999998	0.778	0.988	0.999956
$A3PM_2(100:[4, 5, 5]:1)$	0.015	0.999996	0.797	0.975	0.999960

* $D3(90:10:1)$ is the algorithm used in Pilcher et al. (2005).