POOLED TESTING FOR HIV SCREENING: CAPTURING THE DILUTION EFFECT

LAWRENCE M. WEIN and STEFANOS A. ZENIOS

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received February 1994; revisions received April 1995, June 1995; accepted July 1995)

We study pooled (or group) testing as a cost-effective alternative for screening donated blood products (sera) for HIV; rather than test each sample individually, this method combines various samples into a pool, and then tests the pool. A group testing policy specifies an initial pool size, and based on the HIV test result, either releases all samples in the pool for transfusion, discards all samples in the pool, or divides the pool into subpools for further testing. We develop a hierarchical statistical model that relates the HIV test output to the antibody concentration in the pool, thereby capturing the effect of pooling together different samples. The model is validated using data from a variety of field studies. The model is embedded into a dynamic programming algorithm that derives a group testing policy to minimize the expected cost due to false negatives, false positives, and testing. Because the implementation of the dynamic programming algorithm is cumbersome, a simplified version of the model is used to develop near optimal heuristic policies. A simulation study shows that significant cost savings can be achieved without compromising the accuracy of the test. However, the efficacy of group testing depends upon the use of a classification rule (that is, discard the samples in the pool, transfuse them or test them further) that is dependent on pool size, a characteristic that is lacking in currently implemented pooled testing procedures.

In the first years of the AIDS epidemic, numerous instances of AIDS infection caused by blood transfusion were reported to the Centers for Disease Control. The incidence indicated that the blood supply provides an efficient pathway for spreading the epidemic, and the extent of the epidemic dictated that screening at the individual level should be adopted. As a consequence, all infected blood donors would be identified, and a measurably safer blood supply would be attained. Nevertheless, the cost for such a screening program is substantial, and many developing countries, particularly in Africa where the epidemic is spreading rapidly, are struggling to fight the disease on limited budgets.

Pooled testing is one potential way to reduce the monetary cost without compromising the accuracy of the tests. The rationale behind pooled testing is simple and intuitive: suppose we can pool the sera from ten (for example) individuals and test the pool using a single test. If the *seroprevalence* of HIV, which is the fraction of the population that is infected, is low enough, then there is a high probability that all ten individuals in the pool are HIV negative; in this case, we would learn from a single test what otherwise would be learned from ten individual tests. If, on the other hand, the test outcome is positive, then additional tests (either pooled or individual) would need to be carried out.

However, pooled testing has a possible shortcoming, the *dilution effect*: there is a serious concern that if the pool size is too large, then any HIV positive sera will be sufficiently diluted so as to become undetectable by the test. These *false negatives* can be extremely costly, particularly when pooled testing is employed to protect the blood supply. Moreover, infected individuals exhibiting an unusually low level of antibody concentration are less likely to be detected when screened in pools. Consequently, the *sensitivity* of the test can be seriously affected. (*Sensitivity* is the probability of detecting a diseased individual, whereas *specificity* is the probability of detecting a healthy individual.)

Pooling methods have been evaluated in blood banking systems in several developing countries, including Zaire, Zimbabwe, and Ecuador (see, for example, Cahoon-Young et al. 1988, Emmanuel et al. 1988, Kline et al. 1989, Behets et al. 1990, and Ledro-Monroy et al. 1990). These field studies suggest that pooling methods (with group sizes as large as 80) may be as sensitive and specific as individual testing, and can result in cost savings from 5% to 80%, depending on the actual seroprevalence. On the other hand, the World Health Organization (WHO), concerned with the dilution effect and its consequences on the sensitivity of the test, is more conservative in their proposal: they recommend the use of pools of size no greater than five in Tamashiro et al. (1993). The discrepancy between the field

Subject classifications: Dynamic programming: applications. Health care: screening for HIV. Probability: stochastic model applications.
Area of review: MANUFACTURING, OPERATIONS, AND SCHEDULING.

0030-364X/96/4404-0543 $01.25

studies and the recommendations of the WHO are indications that the dilution effect is not well understood, and the bloodbanking community is in danger of either underestimating or overestimating its importance.

In his seminal paper, Dorfman (1943) showed how pooled testing, which is called *group testing* in the statistical literature, can be employed to efficiently eliminate all defective items from certain large populations. The method found an immediate application in screening World War II draftees for syphilis, where it resulted in considerable savings. The group testing problem was researched aggressively in the 1950s and 60s, (see, for example, Sobel and Groll 1959), and a large literature now exists on this topic; readers are referred to Johnson et al. (1991) for a survey and some recent developments, and to Litvak et al. (1994) for recent work that is motivated by HIV testing. However, nearly all existing studies concentrate on either perfect tests (i.e., tests with no misclassification errors) or imperfect tests with errors that are independent of the group size; two exceptions are Hwang (1976) and Burns and Mauro (1987), who assume that test sensitivity is a specified function of the group size. In addition, all studies neglect the actual test mechanism and, except for Arnold (1977), assume that the test outcome is binary rather than continuous.

In contrast, we attempt to explicitly model both the dilution effect and the continuous nature of the test outcome. Our task is greatly complicated by the fact that the HIV test outcome, which is a continuous quantity called the optical density level, is only an indirect measurement of the unobservable antibody concentration. To model this effect, we propose a two-level hierarchical model whose form is motivated in large part by a probabilistic model that is developed from first principles. Our hierarchical model explicitly captures the physical pooling of sera, and is validated using data from two existing dilution studies and an existing pooling study.

Traditional group testing problems consider a binary test outcome; if the test is HIV negative then the pool is released for transfusion, and if the test is HIV positive then the pool is divided into subpools for further testing. If the pool consists of a single sample, then the sample is discarded if the test outcome is HIV positive. Hence, traditional group testing policies are characterized by the initial pool size and the resulting subpool configuration. Because we explicitly consider the continuous nature of the test outcome, our group testing policy must also incorporate a three-way *classification rule* that is based on the test outcome: the pool is deemed HIV positive (each sample in the pool is discarded), HIV negative (each sample in the pool is released for transfusion) or the pool is divided into subpools for further testing. Our hierarchical pooling model is embedded into a dynamic programming framework that derives the group testing policy that minimizes the expected cost due to testing, false positives and false negatives. Because the dynamic programming algorithm is computationally intensive, we also develop an alternative optimization procedure that uses a simplified pooling model. Our policies are tested on a Monte Carlo simulation model that is driven by the hierarchical model (as opposed to the simplified pooling model), and the results indicate that pooled testing, with a classification rule that explicitly depends on the pool size, can achieve significant cost savings over individual testing.

The paper is organized as follows. A preliminary description of ELISA, the biological assay used for HIV testing, is given in Section 1. The data used in the paper are described in Section 2. The two-level hierarchical model is described in Section 3, and its second level is validated in Section 4 using the data of Section 2. The parameters of the hierarchical model are estimated in Section 5 using the data of Section 2. A dynamic programming framework for the group testing problem is developed in Section 6. Section 7 contains the simplified pooling model, along with an optimization procedure that derives pooled testing policies. Our computational results are reported in Section 8 and concluding remarks appear in Section 9. The Appendix contains the probabilistic model that leads to our use of the hierarchical model.

1. SEROLOGICAL TESTS FOR AIDS

The human body reacts to microbial agents (viruses, bacteria, parasites, etc.) by producing antibodies. The antibodies recognize the infectious agents and bind to them. These agents are called *antigens* (*anti*body *gen*erators), and the regions of the antigens where the antibodies can bind are called *epitopes*. Various immunological tests are designed to detect antibodies, thereby identifying the serological status of the individual.

The Human Immunodefficiency Virus (HIV) is the pathological agent of the Acquired Immune Deficiency Syndrome (AIDS). Enzyme Linked ImmunoSorbent Assays (ELISA) for the HIV virus detect the anti-HIV antibodies and are frequently used for HIV screening. This section contains a brief nontechnical description of ELISAs.

A common configuration of ELISAs for HIV is the *indirect assay* pictured in Figure 1 (see George and Schochetman 1985 for more details). Antigens to HIV are attached to a solid phase support (usually wells). The patient's serum (or plasma) is diluted (at a dilution fixed by the manufacturer), added to the solid phase support and incubated for a time period. By the end of the incubation period, any antibodies to HIV that are present in the sample are attached to the antigens on the solid phase support. The well is then washed so that all unattached material is removed. The next stage of the test is designed to detect all attached HIV antibodies (immunoglobulins). Because there is no known biochemical mechanism that *directly* detects the immunoglobulins, enzyme labeled secondary antibodies are then added to the well. These antibodies will bind to all immunoglobulins and can be detected through an enzymatic reaction. The result of this reaction is a color change

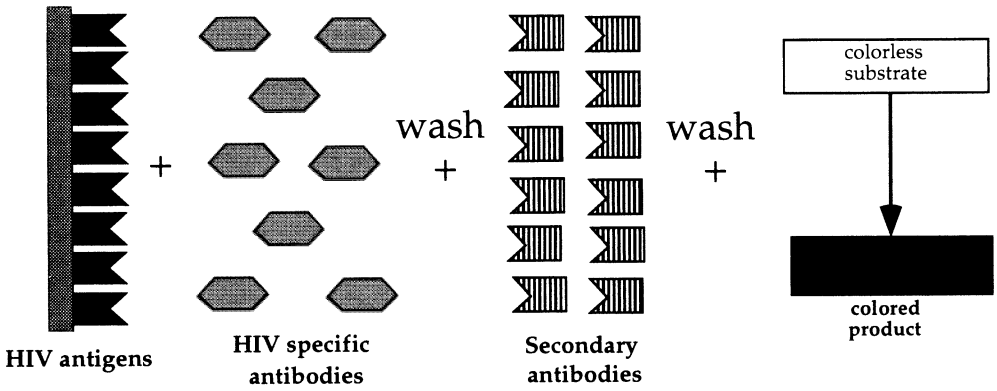

Figure 1. A schematic representation of the indirect ELISA.

proportional to the amount of human HIV antibodies present. The color change is measured using an automated spectrometer, and the outcome is an *optical density* (OD) level; hence, the OD reading is determined by the concentration of the antibodies. If the OD recorded at the end of the process exceeds the critical value, or cutoff, recommended by the manufacturer, then the patient is declared HIV positive; otherwise the patient is declared HIV negative.

An alternative configuration of ELISAs is the competitive assay. Although the same types of antigens used in indirect ELISAs are attached to the solid phase support, this method differs from the indirect one in the detection mechanism. Enzyme labeled HIV antibodies compete with the patient's antibodies for binding sites. The color change observed is inversely proportional to the concentration of HIV antibodies in the serum. If the recorded OD level exceeds the critical value set by the manufacturer, then the sample is declared negative, otherwise positive. We concentrate on indirect assays in this paper, because most of the commercially available antibody detection kits are based on the indirect configuration of ELISAS. Nevertheless, the study of competitive assays is not any more difficult, and only minor modifications of our models are required.

ELISAs are inexpensive, easy to administer, and very accurate; however, they have a shortcoming that stems from the test's *indirect* detection of HIV via the presence of antibodies. The patient's time of infection is followed by a *window period* during which the antibody concentration in the patient's serum is virtually undetectable. This period usually extends from two to twelve weeks (Courcouce 1987) and results in false negatives. Assays for detecting HIV antibodies cannot identify such individuals; therefore, whenever individuals are referred to as positive or negative, we are actually alluding to the presence or absence of HIV antibodies.

2. DESCRIPTION OF THE DATA

We use individual testing data, dilution series data, and pooled testing data obtained from three independent sources.

Individual Testing Data

OD readings for 4000 HIV negative and 3000 HIV positive individuals screened using four different assays were provided by the National HIV Reference Laboratory of Australia (Dax 1993). In this paper, we concentrate on the data from assay A (an indirect ELISA).

Every precaution was taken to avoid biases of any form, and the samples are expected to be representative of the general population. More specifically, the panel of HIV positive samples consists of subjects representing a wide range of clinical manifestations of the infection, including recent seroconverters. The panel of HIV negative samples is representative of the population of noninfected blood donors.

It is convenient to normalize the OD readings according to the equation $x = (OD - A_0)/(A_m - A_0)$, so that they fall between zero and one; A_0 and A_m are the minimum and maximum OD readings, respectively, recorded by the assay. The values of A_0 and A_m vary by assay, and were chosen based on an analysis of the data and discussions with the data providers. For assay A, we set $A_0 = 0$ and $A_m = 20$.

The empirical distributions for the normalized OD readings are given in Figure 2(a). We observe that both the mean and variance are smaller for HIV negative than for HIV positive individuals. The relatively large spread in the HIV positive distribution is to be expected, since an individual's antibody concentration tends to systematically vary as the disease progresses; see George and Schochetman for details. The two populations are well separated, and therefore a critical value separating the OD outcomes into HIV positive and HIV negative can be selected. The cutoff is determined by the manufacturer of the assays and is 0.05 in the normalized OD scale for assay A.

For reasons that will become clear in Section 7, we also consider the *logit* transformation of the normalized OD readings: $x \to \ln(x/(1 - x))$, which will be referred to as the LOD (logit OD) readings. The empirical densities of the LODs for the two populations are given in Figure 2(b). The sample mean and standard deviation are, respectively, $\mu_- = -4.82$ and $\sigma_- = 0.42$ for the HIV negative population, and $\mu_+ = 0.80$ and $\sigma_+ = 1.08$ for the HIV positive population.

Figures 2(c) and (d) displays the normal quantile plot for the empirical distributions in Figure 2(b); that is, the LOD readings are ranked in magnitude and are plotted

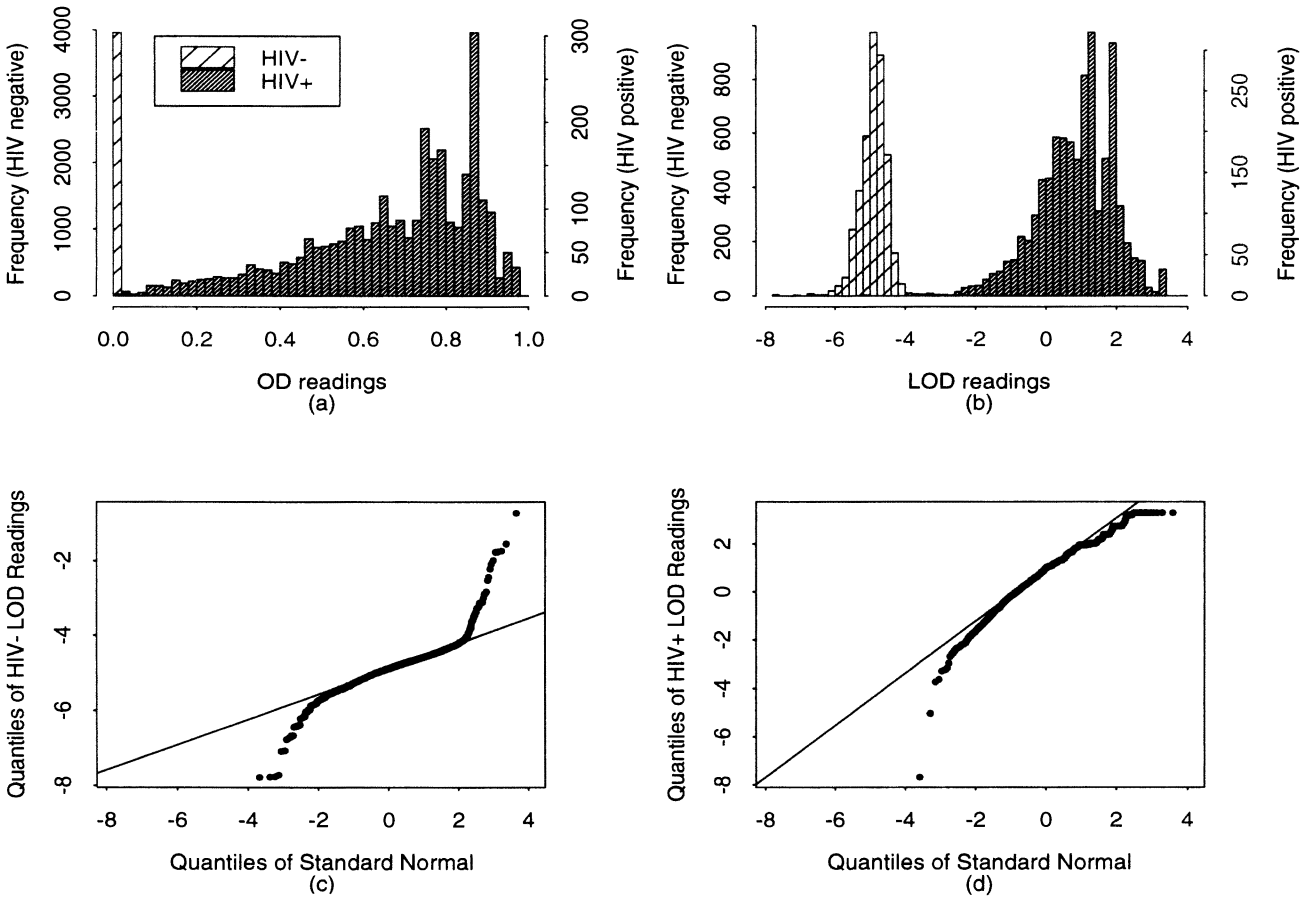

Figure 2. (a) Empirical densities for the OD levels of 4000 HIV negative and 3000 HIV positive individuals, (b) densities for the corresponding LOD values, (c) normal quantile plot for HIV negative individuals, and (d) normal quantile plot for HIV positive individuals.

against the standard normal quantiles. A straight line indicates normality of the data points. The quantile plot of the LOD readings is approximately linear for both populations. Substantial deviations from normality are observed in both tails of the HIV negative population and the right tail of the HIV positive population. Most importantly, the normal approximation captures the left tail of the HIV positive distribution, which contains the low OD readings that might become undetectable under pooled testing. On the other hand, the normal approximation to the HIV negative population underestimates the proportion of negative individuals with a relatively high OD reading, which might lead to an underestimation of false positives. Nevertheless, the false negatives, which are the overriding concern in pooled testing, will not be affected.

In the simplified pooling model in Section 7.2, we assume that the LOD readings for the HIV negative and positive individuals are normally distributed with respective means $\mu_- = -4.82$ and $\mu_+ = 0.80$, and respective standard deviations $\sigma_- = 0.42$ and $\sigma_+ = 1.08$.

Dilution Series Data

Dilution series data were obtained from the Caribbean Epidemiology Center (Hull 1991 and de Gourville 1992) and the National HIV Reference Laboratory of Australia (Dax). The purpose of both of these studies was to investigate the effect of dilution on the ability of ELISAs to detect reactive sera.

In the Caribbean Epidemiology Center (CAREC) study, ten positive sera were diluted sequentially in a fixed negative serum to produce a series of 12 four-fold dilutions in the ratios $1{:}4^1, \ldots, 1{:}4^{12}$, where a $1{:}n$ ratio means that $1/n$ of the pool consists of the positive sample. Each dilution was tested by two indirect ELISAs according to the manufacturer's instructions. Since the data from both assays yielded similar results, we only report the results from one of them. Some data points are missing, and the raw data consist of 79 OD readings. We used $A_0 = 0$ and $A_m = 15$ to normalize the OD readings.

The National HIV Reference Laboratory of Australia (NRL) study sequentially diluted ten positive sera in a fixed negative serum to produce a series of ten two-fold dilutions, with ratios 1:16, 1:32, 1:64, ..., 1:8192. These dilutions were tested in duplicate on eleven different assays, denoted NRL1–NRL11. Assay NRL3 is a competitive assay, so we do not include it in the statistical analysis. For NRL4, the samples were tested in duplicate at four different laboratories, giving a total of eight readings per diluted

sample. The OD readings were normalized using values of A_0 and A_m obtained from an exploratory analysis of the data and after a discussion with the data providers.

Unfortunately, this data set is right censored because ELISA cannot report OD levels greater than A_m. More specifically, the data for NRL2, NRL5, NRL7 and NRL11 contain, respectively, 12, 54, 51, and 54 (out of 200) data points that give the maximum OD level. Because the models of Sections 3 and 4 assume that A_m is only asymptotically attainable, these data points will not be included in the model validation. This omission may introduce a bias into the model for OD levels close to A_m. However, such a bias should not affect any of our pooled testing results because pooled OD levels are expected to be substantially smaller than A_m, due to the dilution effect.

Pooled Testing Data

Cahoon-Young et al. (1992), which will be referred to hereafter as Cahoon-Young, tested 1280 specimens individually and in a series of nested pools. More specifically, the individual specimens were pooled to generate 128 pools of size 10; the pools of size 10 were then combined to form 64 pools of size 20, then 32 pools of size 40, and finally 16 pools of size 80. Twelve individuals were found to be HIV positive, and no more than one positive sample was found in any of the pools of size 80. The OD readings at every stage of this nested testing procedure were recorded.

Note that the dilution series studies by CAREC and NRL differ from Cahoon Young's pooling study in one important respect: positive sera are diluted with varying amounts of the *same* negative sera in the dilution series studies, whereas individual sera are combined with a varying number of *different individual's* sera in Cahoon-Young's pooling study. Hence, although the two dilution series can be used to assess the effect of dilution, the Cahoon-Young study exactly mimics the pooling that would take place under group testing.

3. THE HIERARCHICAL POOLING MODEL

Recall that the outcome of an HIV test is the OD reading, which is an indirect continuous measurement of the unobservable antibody concentration. To capture the continuity of the OD readings we propose a *two-level hierarchical* statistical model that predicts the OD level of an individual sample as a function of its antibody concentration. The first level specifies the probability density of the antibody concentration, and the second level determines the conditional density of the OD reading *given the antibody concentration*. In addition, to capture the dilution effect we assume that *the antibody concentration in a pool equals the average of the individual concentrations.*

Consider a pool consisting of m individual blood samples, and define p to be the HIV prevalence. Let $X^{(m)}$ and $Y^{(m)}$ denote, respectively, the OD reading of the pool and its antibody concentration, and let $f_X^{(m)}(x; p, \phi, \gamma)$ and $\pi_Y^{(m)}(y; p)$ be their respective probability densities (ϕ and γ are nuisance parameters that appear in the second level of the hierarchical model). In addition and without loss of generality, assume that the OD reading is normalized so as to fall between 0 and 1.

The density $\pi_Y^{(m)}(y; p)$ is specified at the model's first level. Let $\pi_+(y)$ and $\pi_-(y)$ denote the probability density for the antibody concentration of infected and noninfected individuals, respectively, and define $\pi^{*(k,m)} = \pi_+^{*k} * \pi_-^{*(m-k)}$, where $*$ is the convolution operator. Let $\pi_Y^{(k,m)}(y)$ be the conditional density of $Y^{(m)}$ *given that the pool consists of k infected and m − k noninfected individuals*. Because the antibody concentration of a pool equals the average of the individual concentrations, a change of variables implies that

$$\pi_Y^{(k,m)}(y) = m\pi^{*\{k,m\}}(my)\,. \tag{1}$$

The law of total probability yields

$$\pi_Y^{(m)}(y; p) = \sum_{k=0}^{m} \binom{m}{k} p^k (1-p)^{m-k} \pi_Y^{(k,m)}(y)\,. \tag{2}$$

At the model's second level, the conditional density of the OD reading given the antibody concentration is specified by

$$f(x; \phi, \gamma | y) = \frac{1}{\sqrt{2\pi\phi \dfrac{y^\gamma}{(1+y^\gamma)^2}}} \cdot \exp\left[-\frac{1}{2}\frac{(1+y^\gamma)^2}{\phi y^\gamma}\left(x - \frac{y^\gamma}{1+y^\gamma}\right)^2\right], \tag{3}$$

where the nuisance parameters ϕ and γ depend upon the test kit employed. Combining the two levels, we conclude that the density of $X^{(m)}$ is

$$f_X^{(m)}(x; p, \phi, \gamma) = \sum_{k=0}^{m} \binom{m}{k} p^k (1-p)^{m-k} f_X^{(k,m)}(x; \phi, \gamma)\,, \tag{4}$$

where

$$f_X^{(k,m)}(x; \phi, \gamma) = \int_0^\infty \pi_Y^{(k,m)}(y) f(x; \phi, \gamma | y)\, dy\,. \tag{5}$$

The model is characterized by the parameters ϕ and γ and the densities $\pi_+(y)$ and $\pi_-(y)$. In Section 5, these quantities will be estimated from the data described in Section 2.

Equation (3), which forms the crux of our model, can be restated as

$$(X|Y = y) \sim N\left(\frac{y^\gamma}{1+y^\gamma}, \phi\frac{y^\gamma}{(1+y^\gamma)^2}\right).$$

We were led to this relationship by a combination of probabilistic analysis and empirical considerations; details can be found in Appendix A.

Notice that the first level of the hierarchical model captures the *between sample* variability and the second level,

by allowing the same sample to give different OD levels when tested in duplicate, captures the *within sample* variability that is due to measurement error. The *dispersion parameter* ϕ measures the magnitude of the within sample variability, and the *shape parameter* γ determines the precision of the test. More specifically, the difference between the OD readings of two samples with similar antibody concentrations is increasing in γ.

A shortcoming of our hierarchical model is that the second level assumes that the within sample variability depends only on the antibody concentration. Although this may not hold in practice, it is an assumption that is frequently used in the statistical analysis of bioassays (Davidian and Giltinan 1994). In addition, this assumption explains the state-dependent variability of the OD readings that is present in the data; see Section 4.

Equations (3) and (4) provide the building blocks for analyzing efficient pooled testing procedures. In Section 6, we embed the hierarchical pooling model in a dynamic programming framework that obtains the optimal pooled testing procedures, and in Section 7 we start from the hierarchical pooling model to develop a simplified pooling model. The latter is used to derive simple but effective pooled testing procedures.

4. MODEL VALIDATION

In this section, we attempt to validate the hierarchical pooling model (hpm) described in Section 3. More specifically, we test Equation (3), as well as the assumption that the antibody concentration in a pool equals the average of the individual antibody concentrations. To perform this task, we develop two generalized linear models, one for the dilution series data of CAREC and NRL, and one for the pooling data of Cahoon-Young.

4.1. Dilution Series Data

We now develop a generalized linear model (GLM) for the dilution series data. Let X_{ij} denote the OD reading of sample i at dilution j, and let y_{ij} denote the unobserved antibody concentration of the diluted sample (because we view the antibody concentration as an unobservable constant, we use the notation y_{ij} instead of Y_{ij}; the latter would be more appropriate if the antibody concentration was considered to be a random variable). This sample consists of $1/d^j$ parts from the positive sample i with antibody concentration y_{i0}, and $(d^j - 1)/d^j$ parts a fixed negative sample with antibody concentration y_∞ ($d = 2$ for NRL and 4 for CAREC). For brevity of notation define $\mu_{ij} = E[X_{ij}]$.

Equation (3) implies that $\mu_{ij} = y_{ij}^\gamma/(1 + y_{ij}^\gamma)$, or

$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \gamma \ln(y_{ij}) . \tag{6}$$

Because the sample consists of the positive sample i diluted to the fixed negative sample, it follows that

$$y_{ij} = \frac{y_{i0} + (d^j - 1)y_\infty}{d^j} . \tag{7}$$

Since $y_{i0} \gg y_\infty$, if j is not very large then

$$y_{ij} \approx \frac{y_{i0}}{d^j} , \tag{8}$$

and substituting (8) into (6) gives the nonlinear model

$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \gamma \ln(y_{i0}) - \gamma \ln(d)j . \tag{9}$$

However, if j is large enough, then (8) may not hold. In this case we introduce a correction term ϵj in (9) to get

$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \gamma \ln(y_{i0}) - (\gamma \ln(d) - \epsilon)j . \tag{10}$$

Combining Equations (3), (9), and (10) give the piecewise GLM

$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \alpha_i - (\beta + \epsilon I_{\{j>k\}})j, \quad \text{and} \tag{11}$$

$$\text{Var}(X_{ij}) = \phi\mu_{ij}(1 - \mu_{ij}) , \tag{12}$$

where $I_{\{x\}}$ is the indicator function of event x, $\alpha_i = \gamma \ln(y_{i0})$, $\beta = \gamma \ln(d)$ and k is a constant determined in advance. The purpose of this constant is to introduce the correction term ϵ when positive samples are excessively diluted into the fixed negative sample.

Our GLM is similar to the logistic regression model that is traditionally used to analyze dilution series data (Tijssen 1985, Chapter 15). This model hypothesizes that

$$\ln\left(\frac{X_{ij}}{1 - X_{ij}}\right) = \alpha_i + \beta j + e_j , \tag{13}$$

where e_j are iid normal random variables with zero mean. Although this model typically fits the dilution data very well, it exhibits considerable heteroscedasticity (state dependent noise), and hence one of the basic assumptions of linear regression is violated. In contrast, our GLM assumes state-dependent noise, and consequently is expected to stabilize the noise present in the data.

The theory of *quasilikelihood functions* (see McCullagh and Nelder 1989, p. 323–352) can be used to estimate the parameters of the GLM from the NRL and CAREC data, and test the goodness of fit. This theory applies under the following four conditions that are satisfied by the GLM: (i) the range of possible normalized OD values X_{ij} is known, (ii) the mean normalized OD level is specified as a function of the dilution level j, (iii) the variance of the normalized OD is specified as a function of the mean OD, and (iv) the errors in the observations are statistically independent. Let us fix sample i, and assume that the X_{ij}s are independent with mean μ_{ij} and variance $V(\mu_{ij})$. Then the log-likelihood function for μ_{ij} can be replaced by the quasilikelihood function

$$\int_{x_{ij}}^{\mu_{ij}} \frac{x_{ij} - y}{V(y)}\, dy , \tag{14}$$

where x_{ij} is the realization of X_{ij}. The estimates for the model parameters are then obtained by maximizing the quasilikelihood function

Table I
GLM Parameter Estimates.

Assay	$\hat{\beta}$	$\hat{\beta} + \hat{\epsilon}$	$\hat{\phi}$	Deviance	d.f.
CAREC	−0.76 ± 0.210	−0.97 ± 0.210	0.035	1.92	58
NRL1	−0.46 ± 0.035	−0.33 ± 0.035	0.060	10.93	178
NRL2	−0.77 ± 0.051	−0.56 ± 0.051	0.023	3.42	156
NRL4	−0.41 ± 0.020	−0.38 ± 0.020	0.059	43.03	778
NRL5	−0.06 ± 0.120	−0.16 ± 0.120	0.250	34.59	124
NRL6	−0.17 ± 0.035	−0.17 ± 0.035	0.059	9.92	178
NRL7	−0.67 ± 0.099	−0.66 ± 0.099	0.066	8.70	128
NRL8	−0.43 ± 0.024	−0.36 ± 0.024	0.032	6.61	178
NRL9	−0.21 ± 0.035	−0.16 ± 0.035	0.084	14.78	178
NRL10	−0.24 ± 0.022	−0.32 ± 0.022	0.038	6.87	178
NRL11	−0.29 ± 0.130	−0.34 ± 0.130	0.147	18.83	124

$$Q(\mu, x) = \sum_{i=1}^{10} \sum_{j=n_1}^{n_2} \int_{x_{ij}}^{\mu_{ij}} \frac{x_{ij} - y}{V(y)} dy , \tag{15}$$

for each study, where $n_1 = 1$ and $n_2 = 12$ for CAREC, and $n_1 = 4$ and $n_2 = 14$ for NRL. We use the *glm* routine of S-plus (see Hastie and Pregibon 1992) to estimate the dispersion parameter ϕ, the slope β, the correction term ϵ, and y-intercept α_i, $i = 1, \ldots 10$ for each study. The estimates are denoted by $\hat{\alpha}_i$, $\hat{\beta}$, $\hat{\epsilon}$, and $\hat{\phi}$, and the *predicted values* $\hat{\mu}_{ij}$, which are the mean response values predicted by the model, are given by

$$\ln\left(\frac{\hat{\mu}_{ij}}{1 - \hat{\mu}_{ij}}\right) = \hat{\alpha}_i + \hat{\beta} j + \hat{\epsilon} I_{\{j>k\}} . \tag{16}$$

The Pearson residuals are defined as $(x_{ij} - \hat{\mu}_{ij})/V(\hat{\mu}_{ij})$. We considered two values for k, $n_1 + 5$ and $n_1 + 6$, and the latter value yielded better results.

The results are displayed in Table I and Figure 3. Table I gives the estimates for the GLM. The second and third columns give 95% confidence intervals for β and $\beta + \epsilon$, the fifth column gives the residual deviance and the sixth column the degrees of freedom. The residual deviance is the sum of the deviance residuals and measures the dispersion of the GLM; see McCullagh and Nelder (p. 33–38). The degrees of freedom are equal to the difference between the sample size and the number of parameters. Figure 3 gives the residuals vs. fitted values (RVSF) plots together with a lowess smooth that is added to help the interpretation.

We start with an analysis of the RVSF plots. The panels for six assays, CAREC, NRL5, NRL7–8, and NRL10–11, give an approximately null plot. In contrast, the panels for NRL1–2, NRL6, and NRL9 possess U-shaped residuals, and the panel for NRL4 has fan-shaped residuals. It is worth noting that the logistic regression model (13) was also fitted to the data, and the RVSF plots gave characteristic fan-shaped residuals. We also fitted to the data a GLM without the correction term ϵ, and the RVSF plots were U-shaped. This suggests that the proposed GLM is partially successful in stabilizing the state-dependent noise, and in capturing the OD level of excessively diluted samples. However, it appears that the exact form of the GLM may be assay dependent.

To assess the predictive power of the model, we use the OD readings at dilution levels $j = 1, 2, 6, 7, 11, 12$ from CAREC and $j = 4, 5, 8, 11$ from NRL1 to estimate the parameters of the GLM. The predicted values obtained for the remaining data (i.e., dilution levels $j = 3, 4, 5, 8, 9, 10$ for CAREC and $j = 6, 7, 9, 10, 12, 13, 14$ for NRL) are plotted against the observed values in Figure 4. The predictive power of the model is verified by observing that nearly all the observed points lie within the 95% confidence interval predicted by the model.

The informal analysis of the last two paragraphs can be supplemented by a more formal analysis based on the residual deviance. More specifically, it is possible to test the *goodness of fit* by comparing the proposed model to the *full model* that has as many parameters as data. Under the full model hypothesis, the residual deviance is asymptotically χ^2 with degrees of freedom given by the sixth column of Table I. The test statistic is significant for all assays (p-value > 0.995), thereby ascertaining the quality of the fit.

It is also possible to test statistically the homoscedasticity of the Pearson residuals using a Lagrange multiplier test; see Madansky (1988, p. 88). This test regresses the absolute Pearson residuals into the linear predictor and tests the hypothesis that the slope of the linear regression is zero. The homoscedasticity assumption is rejected if the test statistic is significant. Applying this test to the residuals from the GLM, we found that the test statistic was significant for nine out of the eleven assays. The p-values for NRL10 and NRL11 were, respectively, 0.39 and 0.28. Therefore, the GLM does not stabilize the state-dependent noise in all assays. Nevertheless, the results suggest that the GLM provides a good fit to the data from NRL10 and NRL11, and an adequate fit to the data from the remaining assays.

We also used the Lagrange multiplier test to assess the homoscedasticity of the traditional logistic regression model and found that the test statistics were always greater

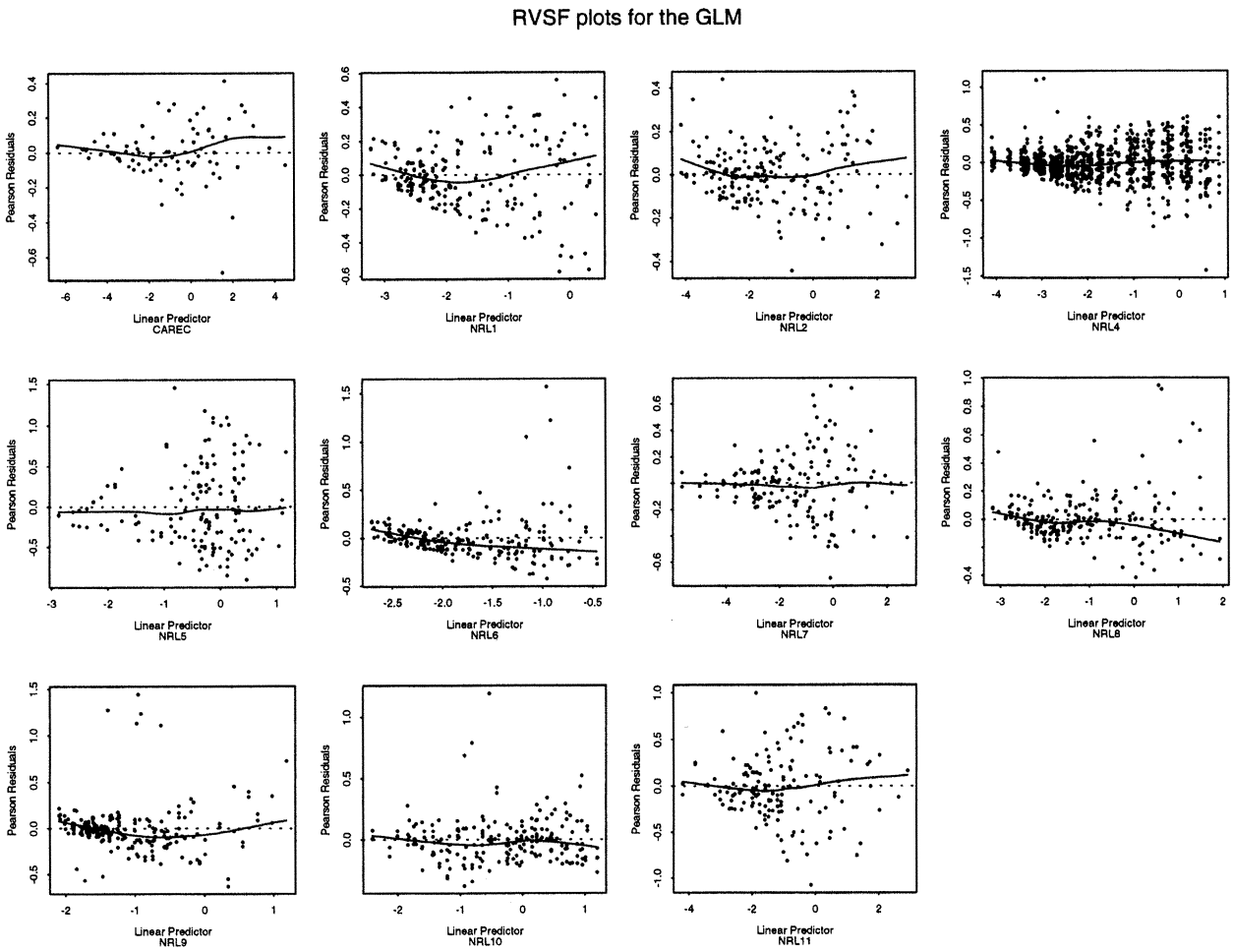

Figure 3. RVSF plots for dilution series data.

than the test statistics for the corresponding GLM. In summary, although the GLM is not successful in fully stabilizing the residuals, it reduces the heteroscedasticity of the traditional regression model.

The statistical analysis can also provide useful insights into the mechanism of ELISA. First, we observe that there is considerable intra-assay variability in the values of the parameters ϕ and γ. Second, we notice that for some assays the within sample variability is quite small; for example, $\phi = 0.038$ for NRL10. This has some important implications regarding the coefficient of variation (standard deviation divided by the mean) of the OD readings. More specifically, consider a sample with antibody concentration y. The coefficient of variation of its OD reading is

$$c_v = \sqrt{\phi \frac{1}{y^{\gamma}}}. \tag{17}$$

Hence, the within sample coefficient of variation should be quite small for HIV positive samples, but is expected to be much larger for HIV negative samples. For example, if $\phi = 0.038$ and $y^{\gamma} = 0.0086$ (Section 5 verifies that this is a typical HIV negative value) then $c_v = 2.10$. In contrast, if $y^{\gamma} = 3.79$ (a typical HIV positive sample) then $c_v = 0.10$. This observation will be instrumental when we estimate the densities $\pi_+(y)$ and $\pi_-(y)$ in Section 5.

4.2. Pooling Data

The GLM for the pooling data of Cahoon-Young is developed under the assumption that all HIV negative samples have the same antibody concentration. Although this may appear to be a bold assumption, we will see in Section 5 that it can be justified using the individual testing data of NRL.

Recall that Cahoon-Young individually tested 1280 samples, and then generated nested pools of size 10, 20, 40, and 80. The total sample contained 12 HIV positive individuals, and none of the pools contained more than one positive sample; hence, 12 of the 16 pools of size 80 contained exactly one positive sample. Let y_{i1} denote the unobserved antibody concentration of the positive sample i, and let y_{∞} denote the antibody concentration of the negative samples. In this subsection, let X_{is} denote the OD reading for the pool of size 10×2^s containing the positive sample i, where $s = 0, 1, 2, 3$, and let y_{is} denote its antibody concentration.

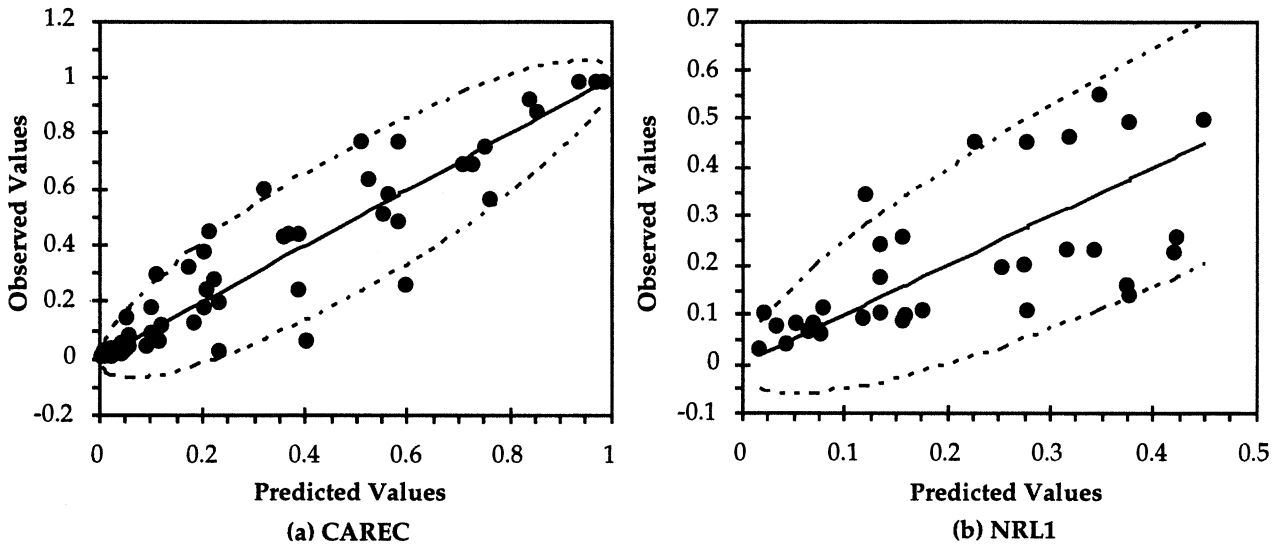

Figure 4. Verifying the predictive power of the GLM: the observed values lie within the 95% confidence interval predicted by the model.

Equation (3) implies that

$$\ln\left(\frac{E[X_{is}]}{1 - E[X_{is}]}\right) = \gamma \ln(y_{is}), \quad \text{where} \tag{18}$$

$$y_{is} = \frac{y_{i1} + (10 \times 2^s - 1)y_\infty}{10 \times 2^s}. \tag{19}$$

Because $y_{i1} \gg y_\infty$ and $s = 1, 2, 3$, it follows that

$$y_{is} \approx \frac{y_{i1}}{10 \times 2^s}. \tag{20}$$

Combining equations (3), (18) and (20) gives the GLM

$$\ln\left(\frac{E[X_{is}]}{1 - E[X_{is}]}\right) = \gamma \ln\left(\frac{y_{i1}}{10}\right) - s\gamma \ln(2), \quad \text{and} \tag{21}$$

$$\text{Var}(X_{is}) = \phi E[X_{is}](1 - E[X_{is}]). \tag{22}$$

We fit the GLM to the data, and the results suggest a very good fit. The residual deviance is 0.1686 at 31 degrees of freedom, and the RVSF plot is approximately null (Figure 5). However, the results may suggest overfitting because the model has 14 parameters and the sample size is only 44. Therefore, additional data is required to reach a stronger conclusion.

5. PARAMETER ESTIMATION FOR THE HPM

The HPM in Section 3 is characterized by the antibody concentration densities $\pi_+(y)$ and $\pi_-(y)$ and the parameters ϕ and γ. The purpose of this section is to specify these quantities using the data in Section 2. The form of the two densities are needed in Section 6 to derive pooled testing procedures, and all the parameter estimates are used to drive the simulation model that forms the basis of our computational study in Section 8.

The densities $\pi_+(y)$ and $\pi_-(y)$ will be estimated from the individual testing data in Section 2. The parameters ϕ and γ are solely dependent on the test kit employed. The parameter ϕ can be obtained as a by-product of estimating $\pi_+(y)$ and $\pi_-(y)$ (details are given below), but γ can only be estimated from dilution series data. Unfortunately, the test kits employed in the dilution series data in Section 2 are different than those used in the individual testing data, so we have no basis for finding an estimate for γ that is consistent with the individual testing data. Nevertheless, our analysis in Section 4.1 can be used to get a rough estimate of γ. Recall that $\beta = -\gamma \ln(d)$, where d is the known dilution level and β is the slope of the GLM model (11). Using the estimates in Table I, we find that the corresponding estimates $\hat{\gamma}$ range from 0.09 to 1.11, with a mean and median equal to 0.54. Consequently, we employ the value 0.54 in the simulation model. We also made several simulation runs with $\gamma = 1.0$ and detected no qualitative differences in our results between the $\gamma = 0.54$ and $\gamma = 1.0$ cases.

The densities $\pi_+(y)$ and $\pi_-(y)$ are difficult to estimate because the antibody concentration is not an observable quantity. We first develop a general methodology to estimate the densities $\pi_+(y)$ and $\pi_-(y)$ from generic individual testing data, and then use this methodology in conjunction with exploratory analysis of the NRL data to estimate the densities. Because we have assumed that the antibody concentration is related to the (observable) OD reading via (3), the estimation problem takes the form of the following inversion problem: determine densities $\pi_+(y)$ and $\pi_-(y)$ to "minimize" the distance between $\int \pi_+(y) f(x; \phi, \gamma|y)\, dy$ and the empirical OD density for infected individuals, as well as the distance between $\int \pi_-(y) f(x; \phi, \gamma|y)\, dy$, and the empirical OD density for HIV negative individuals. To solve this problem, we adopt a weakly parametric model that assumes that the antibody concentration takes discrete values $y_1, \ldots, y_k$. For concreteness, let us concentrate on infected individuals and define $p_j = P(Y = y_j | Y \in P_+)$ for $j = 1, \ldots, k$; Y denotes the antibody concentration of a random individual, and P_+

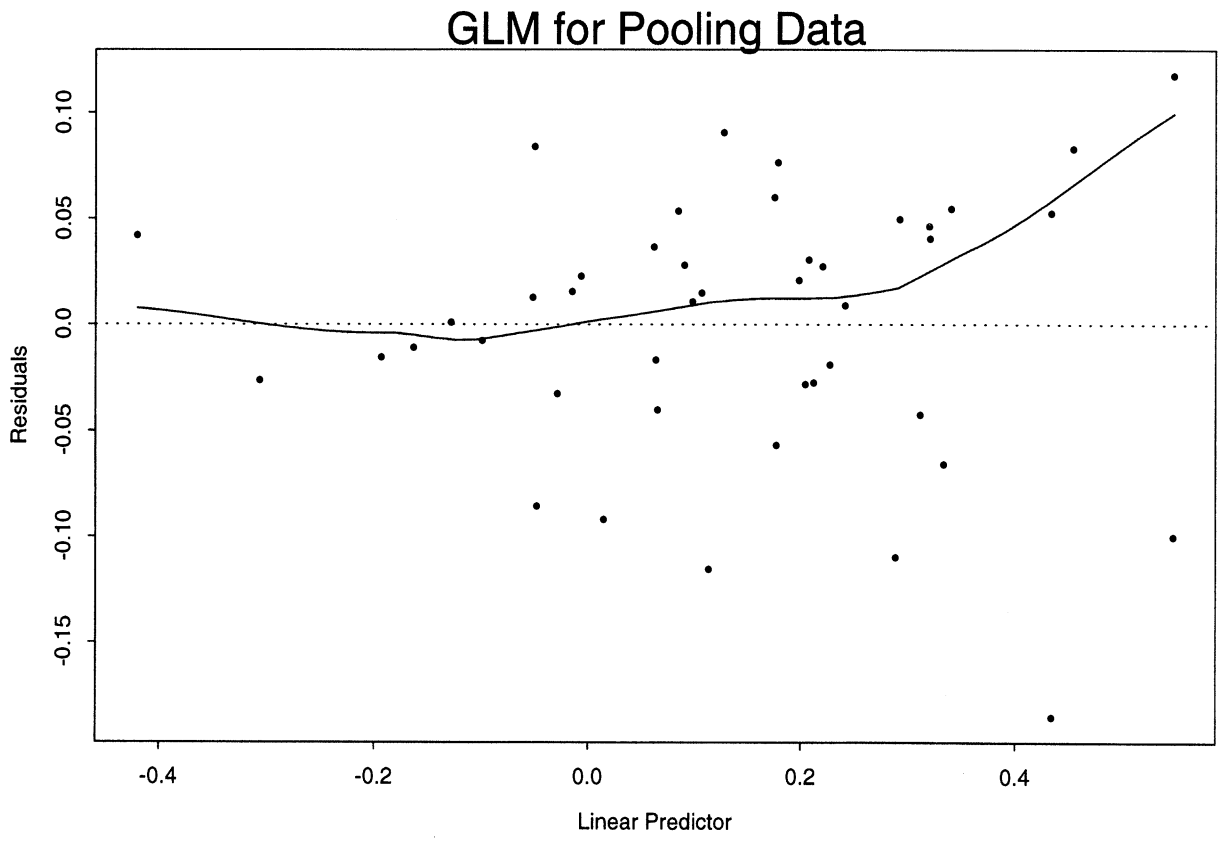

Figure 5. The RVSF plot for the pooling data.

denotes the set of infected individuals. Let the raw data be x_i, which is the OD reading for infected sample $i = 1, \ldots, n$ ($n = 3000$ from the individual data set of NRL).

We employ the EM algorithm (Dempster et al. 1977) to obtain the maximum likelihood estimates for $p_1, \ldots, p_k$, where δ is the desired tolerance level:

Step 1. Start the procedure with initial estimates $p_j^{(0)}$.

Step 2. For $i = 1, \ldots, n$ and $j = 1, \ldots, k$ calculate the conditional probability that sample i has antibody concentration y_j *given the OD reading* x_i:

$$\tau_{ij}^{(s)} = \frac{p_j^{(s)} f(x_i;\ \phi,\ \gamma|y_j)}{\sum_{l=1}^{k} p_l^{(s)} f(x_i;\ \phi,\ \gamma|y_l)}. \tag{23}$$

Step 3. Update the estimates $p_j^{(s)}$ using

$$p_j^{(s+1)} = \frac{\sum_{i=1}^{n} \tau_{ij}^{(s)}}{n}. \tag{24}$$

Step 4. Terminate if $\sum_{j=1}^{k} (p_j^{(s+1)} - p_j^{(s)})^2 \leq \delta$; otherwise, set $s \leftarrow s + 1$ and go to Step 2.

Although this algorithm assumes that ϕ and γ are known, it can be used even if ϕ is unknown. In this case, the following iteration step is added to Step 3:

$$\phi^{(s+1)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} \tau_{ij}^{(s)} \left(x_i - \dfrac{y_j}{1+y_j} \right)^2 \dfrac{(1+y_j)^2}{y_j}}{n}. \tag{25}$$

However, we do not recommend the use of this algorithm when γ is unknown because this parameter is not identifiable; for example, $\gamma = 1$ and $P(Y = 5) = 1$ is equivalent to $\gamma = 3$ and $P(Y = 5^{1/3}) = 1$.

It is possible to give an intuitive interpretation to the iteration Steps (24) and (25). Suppose that the antibody concentrations are observable, and define the hypothetically observable quantity $T_{ij} = I_{\{\text{sample } i \text{ has antibody concentration } y_j\}}$. Then, the mle of p_j is $n^{-1} \sum_{i=1}^{n} T_{ij}$ and the mle of ϕ is

$$\phi = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{k} T_{ij} \left(x_i - \frac{y_j}{1+y_j} \right)^2 \frac{(1+y_j)^2}{y_j}.$$

However, since the T_{ij}s are not observable, they are replaced by $\tau_{ij}^{(s)} = E[T_{ij}|x_i, p^{(s)}]$.

A variant of this algorithm was used by Becker et al. (1991) to estimate the HIV incidence curve from AIDS incidence data. Their results verify all the well-known advantages and shortcomings of the method. The procedure is easily implementable, but its convergence rate is very slow. It does not provide any information on the uncertainty of the estimates, and the estimates may depend on the initial estimate $p_j^{(0)}$. In addition, if some data points are substantially different from the majority of the observations (outliers) then the algorithm may produce unreliable estimates. Nevertheless, if the starting points are selected carefully and possible outliers filtered out, then this method can produce very reliable estimates.

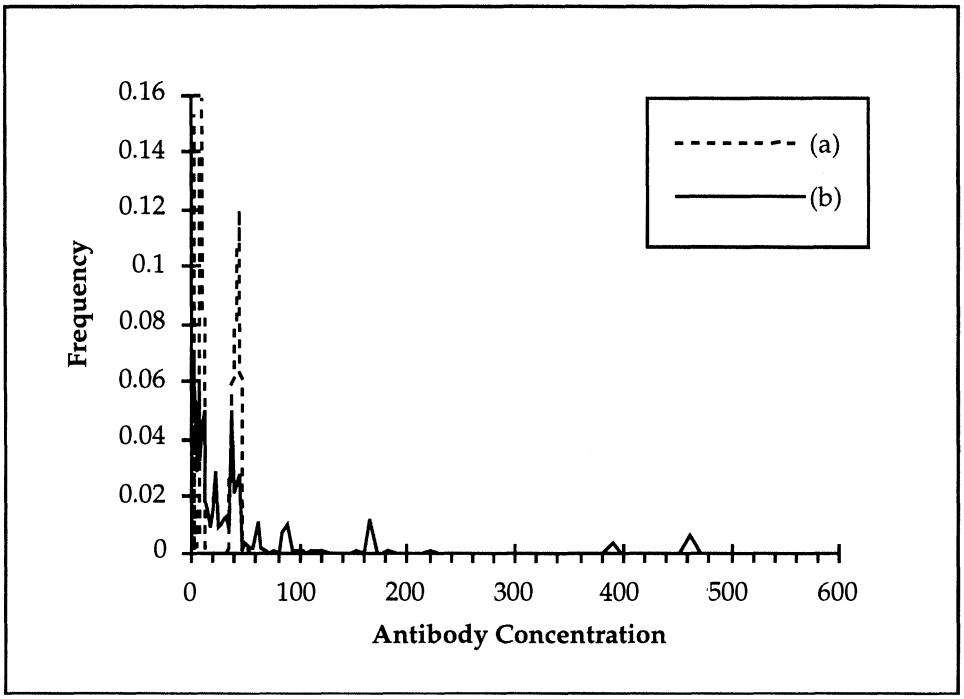

Figure 6. The density $\pi_+(y)$: (a) EM algorithm, and (b) exploratory data analysis.

We used Equations (23)–(25) to estimate the densities $\pi_+(y)$ and $\pi_-(y)$ for the populations underlying the NRL individual testing data; the parameter γ was set equal to 0.54, and we also found that it was more convenient to discretize the (antibody concentration)$^\gamma$ values than to discretize the antibody concentration values.

The algorithm was implemented in C, on a Sun Sparc Station 10. For the noninfected individuals we used the discretization $\{0.01, 0.03, \ldots, 2.01\}$. The algorithm converged to the degenerate density $p_1 = 0.9999$ and $\phi = 0.00817$ after 220 CPU minutes. For the infected individuals we used the discretization $\{0.3, 0.6, \ldots, 30.0\}$. The algorithm converged after 2770 CPU minutes, yielding $\phi = 0.0149$ and the density for $\pi_+(y)$ pictured in Figure 6(a).

Unfortunately, the two estimates for ϕ are very different, suggesting that the estimates for $\pi_+(y)$ and $\pi_-(y)$ may not be reliable. To understand the source of this discrepancy, we now obtain alternative estimates based on an exploratory analysis of the data.

Let η_- and η_+ denote, respectively, the mean antibody concentration of noninfected and infected individuals, and let ν_- and ν_+ be the respective standard deviations. To estimate the density $\pi_-(y)$ we make the simplifying assumption that $\nu_- = 0$, which implies that the variability of OD readings for HIV negative individuals is due entirely to within sample variability and contains no between sample variability.

We justify this assumption in two ways. First, the crude computations following (17) reveal that the within sample coefficient of variation of the OD readings for HIV negative individuals is roughly equal to two. In addition, the OD readings for the 4,000 HIV negative individuals of the NRL data have coefficient of variation 1.016. Therefore, we conclude that the variability of the OD readings of HIV negative individuals is mostly due to within sample variability.

Our second justification requires two steps, and also produces an estimate for the dispersion parameter ϕ. First, we use the hierarchical model to obtain the "feasibility" region of ϕ. Then, we obtain the subset of the feasibility region that is consistent with the empirical observations of the previous paragraph. If we let X and Y denote, respectively, the OD reading and antibody concentration of a typical HIV negative individual, then the hierarchical model implies that $X = T + e$, where $T = Y^\gamma/(1 + Y^\gamma)$ and e is a mean zero Gaussian random variable satisfying $E(e^2|T) = \phi T(1 - T)$. In Appendix B, we show that $\mathrm{Var}(T) \geq 0$ if and only if

$$\phi \leq \frac{\mathrm{Var}(X)}{E(X)(1 - E(X))}. \tag{26}$$

This inequality gives the feasibility region of ϕ; if ϕ does not satisfy (26), the model is misspecified.

Because the OD readings for the NRL individual testing data have mean 0.0086, the feasibility region is $\phi \leq 0.0088$. We now argue that $\phi \approx 0.0088$. Suppose that $\phi \ll 0.0088$ and consider a "typical" HIV negative blood sample with antibody concentration y, satisfying $y^\gamma/(1 + y^\gamma) = E(T)$. From (17), the coefficient of variation of its OD reading is $c_v = \sqrt{\phi(1 - E(T))/E(T)}$. Since $E(X) = E(T)$, and $E(X)$ is estimated to be 0.0086 from the individual testing data, we have $c_v \ll 1$. However, this contradicts the empirical

evidence that $c_v \approx 2$, and hence it must be that $\phi \approx 0.0088$.

In summary, the antibody concentration for HIV negative individuals is approximated by a degenerate random variable that is equal to η_- with probability 1; moreover, we have obtained the estimate 0.0088 for the parameter ϕ. To complete the description of π_- it remains to estimate η_-. Because $E(X) = E(T)$ and T and Y are degenerate, it follows that $E(T) = \eta_-^\gamma/(1 + \eta_-^\gamma)$. Therefore, $\eta_- = (E(X)/(1 - E(X)))^{1/\gamma}$. Substituting our estimates for $E(X)$ and γ yields $\eta_- = 0.00015$. Notice that the results of the informal analysis for HIV negative individuals are consistent with the results from the deconvolution method (where $P(Y = 0.000198) = 0.9999$); consequently, in the simulation model in Section 8 we use $\phi = 0.0088$ and for $\pi_-(y)$ we use the degenerate density $\eta_- = 0.00015$, $\nu_- = 0$.

Finally, we turn to the density $\pi_+(y)$. By the discussion following (17), the within sample coefficient of variation of the OD readings for HIV positive individuals is approximately 0.1. Furthermore, the OD readings for the 3000 HIV positive individuals from NRL have coefficient of variation 0.31. Hence, the within sample variability appears to be smaller than the between sample variability for HIV positive individuals, and we make the simplifying assumption that the within sample variability is zero for infected individuals. Thus, $X|Y = E(X|Y)$ for infected individuals, and combining this with Equation (3) we obtain

$$Y = \left(\frac{X|Y}{1 - X|Y}\right)^{1/\gamma}. \tag{27}$$

Since our estimate for γ is 0.54, if we let $x_1, \ldots, x_{3000}$ denote the observed OD readings for the 3000 infected individuals of the NRL data set, then the density $\pi_+(y)$ is specified by the empirical density of

$$\left(\frac{x_1}{1 - x_1}\right)^{1.85}, \ldots, \left(\frac{x_{3000}}{1 - x_{3000}}\right)^{1.85}.$$

The results are given in Figure 6(b). For completeness we report that $\eta_+ = 21.48$ and $\nu_+ = 53.41$.

A careful examination of Figures 6(a) and 6(b) provides a partial explanation for the inconsistency of the estimates derived from the EM algorithm. Intuitively, the derived density provides the "best" approximation to the heuristic density of Figure 6(b). However, because so few points contribute to the far right tail of the heuristic density (87 out of 3000 have concentrations greater than or equal to 15.0), the EM algorithm is unable to capture this tail. Consequently, the derived estimates for the tail of $\pi_+(y)$ are expected to be unreliable, and from (25) the estimate for ϕ will be very uncertain. It is possible to obtain a more precise estimate for ϕ by re-running the EM algorithm, but with the tail data points omitted. The new estimate is $\phi = 0.0096$, which is comparable to the estimates from the noninfected individuals.

This resolves the discrepancy between the two estimates for ϕ. However, it reinforces our concerns regarding the reliability of the density for $\pi_+(y)$ derived from the EM algorithm, and consequently we do not adopt this density in our analysis. Because there are no compelling reasons against the possibility of few samples exhibiting very large antibody concentrations, in the simulation model we use the empirical density for $\pi_+(y)$ obtained from the NRL data via (27). We do not claim that this is the exact density $\pi_+(y)$; however, we believe that the proposed density provides a *plausible* model that can be used to develop and evaluate several pooled testing policies.

Although the empirical density $\pi_+(y)$ is useful for generating random samples in the simulation study in Section 8, it is difficult to manipulate for purposes of finding an optimal pooled testing strategy. Thus, in the next section we assume that $\pi_+(y)$ has a particular distributional form. More specifically, recall that in Section 2 the probability density of the HIV positive LOD readings is approximately normal with mean $\mu_+ = 0.80$ and standard deviation $\sigma_+ = 1.08$. Because Y is related to X via (27), it follows that $\log(Y)$ is approximately normal, or Y is log-normal. Consequently, in Section 6 we assume that $\pi_+(y)$ is log-normal with mean parameter 0.8/0.54 and standard deviation parameter 1.08/0.54.

6. THE DERIVATION OF POOLED TESTING POLICIES

In this section, we embed the hierarchical pooling model into an optimization framework to find efficient pooled testing policies. Our objective is to minimize the expected weighted cost due to testing, false positives and false negatives. Suppose a pool of a specified size is tested and its OD reading is determined. The decision maker has three options: stop testing and classify all individuals in the pool as HIV negative (and transfuse these samples), stop testing and classify all individuals in the pool as HIV positive (and discard these samples), or divide the pool into subpools for further testing. There are many possible ways to subdivide the pool under the third option, and we consider a quite general class of multistage policies employed by Arnold, where each subpool is of identical size. Our procedure can be modified slightly to allow unequal subpool sizes, as in the sequential procedure in Hwang (1984) and the pooling procedure in Litvak et al.

Unfortunately, the proposed cost function may not be appropriate in this setting because the false negative cost is difficult to quantify (and may be arbitrarily large) when blood samples are tested for transfusion. Hence, the cost function should be used with care, and the false negative cost should be interpreted as an artificial parameter that imposes a tradeoff between testing cost and test accuracy. More specifically, in a limited resources environment it may be necessary to adopt a procedure with sensitivity lower than that of individual testing, so as to meet the imposed budget constraint. Our approach can optimally

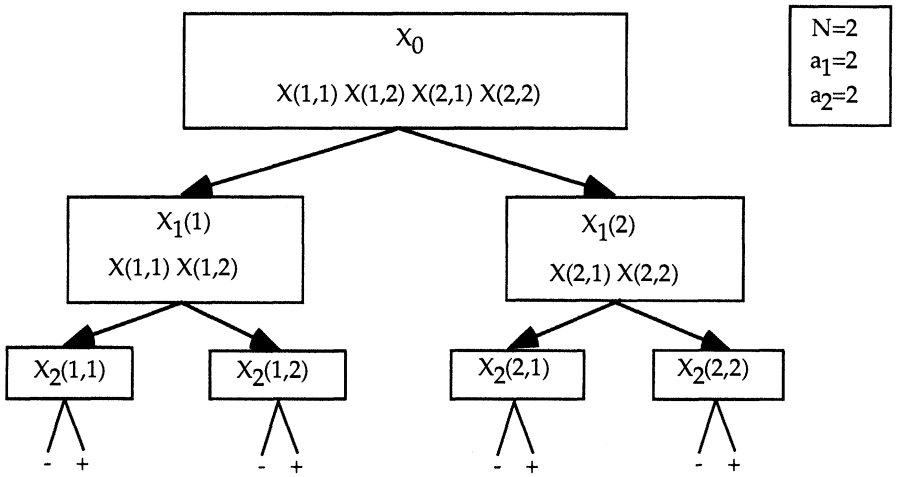

Figure 7. A simple example of a multistage group testing procedure.

balance the necessary tradeoff by considering different values for the false negative cost.

It is worth noting that a more appropriate objective in this setting may be the amount of money spent per HIV cases prevented. However, this objective makes the analysis intractable. Nevertheless, it is expected that a pooling procedure that minimizes the proposed cost function will also reduce the amount spent per cases prevented; this intuition is confirmed in Section 8.

This section and the next are devoted to the analysis of the optimization problem. For a given initial pool size and subpool configuration, a dynamic programming algorithm is developed in Section 6.1 for finding the optimal policy within the class of multistage policies under consideration. Exhaustive search among alternative initial pool sizes and subpool configurations is required to find the cost minimizing policy. Structural properties of the optimal policy are investigated in Section 6.2. Because the dynamic programming algorithm is computationally intensive and the resulting policy difficult to implement, a simpler procedure for deriving Dorfman policies is described in Section 7.

6.1. The Dynamic Programming Formulation

We assume that the blood donor population is composed of two subpopulations: HIV negative (denoted P_-) and HIV positive (P_+). The antibody concentrations of P_- (P_+, respectively) are assumed to be iid log-normal random variables with mean η_- (η_+, respectively) and standard deviation ν_- (ν_+, respectively); see the end of Section 5 for a justification of this assumption. The known *seroprevalence p* is the probability that a random donor is HIV positive. Arnold's notation will be adopted to describe the multistage testing procedure. Consider a random sample of $m_0 = \prod_{j=1}^{N} a_j$ individuals from the donor population; the a_j's dictate the subpool configuration. Blood sera is collected from all m_1 individuals and is indexed in such a way that the OD reading for every sample is denoted by

$$\{X(i_1, \ldots, i_N),\ 1 \leq i_1 \leq a_1, \ldots, 1 \leq i_N \leq a_N\}.$$

The individuals are tested and classified according to the following multistage screening procedure (see Figure 7 for a simple example): start by obtaining X_0, the OD reading of the pool composed of all m_0 individuals. Based on X_0, decide whether all individuals in the pool can be classified as HIV negative or HIV positive; if so, stop testing. If not, then subdivide the population into a_1 subpopulations of size $m_1 = \prod_{j=2}^{N} a_j$, with the first subpopulation consisting of all individuals with $i_1 = 1$, the second with $i_1 = 2$ and so on. Obtain the OD readings $X_1(1), \ldots, X_1(a_1)$ for all subpopulations. For each subpopulation, decide whether all individuals should be classified as HIV negative or positive based on the pair $(X_0, X_1(j))$; if so, stop testing. If not, subdivide those subpopulations that require further testing into z_2 subpopulations of size $m_2 = \prod_{j=3}^{N} a_j$. Continue in this vein until either all pools are deemed HIV negative or positive, or stage N, where individual testing is used, is reached. The testing scheme can be equivalently described by a rooted tree, where the nodes of the tree correspond to the different subgroups formed during the procedure.

The state of the system at any stage of the screening process can be described by the OD readings obtained thus far. If the pool that is currently being screened is composed of all individuals with the first j indices given by $i_1, \ldots, i_j$, then the state is given by $(X_0, X_1(i_1), \ldots, X_j(i_1, \ldots, i_j))$. To simplify notation, we shorten $X_j(i_1, \ldots, i_j)$ to X_j for $j = 1, \ldots, N$, and denote the state of the system by $S_j = (X_0, \ldots, X_j)$ for $j = 0, \ldots, N$. If the current state of the system is S_j and $j \leq N - 2$, then three decisions are possible: either declare all individuals in the pool as negative and stop testing, declare all individuals in the pool as positive and stop testing, or subdivide the pool into a_{j+1} subpools of size m_{j+1} and continue testing. Under the first decision, a false negative cost c_{FN} is incurred for each HIV positive individual in the pool, and under the second decision, a false positive cost c_{FP} is incurred for each HIV negative individual in the pool. Under the third decision, a testing cost $c(m_{j+1})$ is incurred for each of the a_{j+1}

subpools. By defining $a_{N+1} = 1$, we can adopt the same notation for the decision at stage $N - 1$. At stage N (individual testing), the individuals are classified as HIV positive or negative, and the false positive cost c_{FP} or the false negative cost c_{FN} is incurred for any individuals that are misclassified.

If we let $J_j(S_j)$ denote the optimal cost for stages j through N when in state S_j at stage j, then the dynamic programming algorithm is defined inductively by

$$J_N(S_N) = \min\{c_{FP}P(X_N \in P_-|S_N), c_{FN}P(X_N \in P_+|S_{NR})\}, \quad (28)$$

$$J_j(S_j) = \min\{a_{j+1}(c(m_{j+1}) + E[J_{j+1}(S_{j+1})|S_j]), c_{FN}\sum_{i_{j+1}=1}^{a_{j+1}}\cdots\sum_{i_N=1}^{a_N} P(X_N(i_1, \ldots, i_N) \in P_+|S_j), c_{FP}\sum_{i_{j+1}=1}^{a_{j+1}}\cdots\sum_{i_N=1}^{a_N} P(X_N(i_1, \ldots, i_N) \in P_-|S_j)\}, \quad (29)$$

for $j = 0, \ldots, N - 1$. Because the individual OD readings of each sample in a pool are iid random variables, Equation (29) can be simplified to

$$J_j(S_j) = \min\{a_{j+1}(c(m_{j+1}) + E[J_{j+1}(S_{j+1})|S_j]), m_{j+1}c_{FN}P(\tilde{X}_j \in P_+|S_j), m_{j+1}c_{FP}P(\tilde{X}_j \in P_-|S_j)\} \quad \text{for } j = 0, \ldots, N-1, \quad (30)$$

where $\tilde{X}_j$ is a random variable denoting the individual OD reading of a generic member of the pool at stage j.

Since the state of the system at stage j is given by the OD readings obtained through stage j, the dimensionality of the state space grows as the dynamic programming algorithm proceeds; hence, the algorithm in (28) and (30) cannot be efficiently used for numerical calculations. Furthermore, a policy that classifies the test outcomes using the complete state space is unlikely to be implemented in practice. For these reasons, we restrict our attention to *simple* policies that classify the test outcomes based only on the latest OD readings. Although these policies are, in general, suboptimal, they turn out to be optimal when $\phi = 0$ (see Zenios).

For simple policies, we can replace the state S_j by the latest OD reading X_j, and the optimal decision rule can be described by a set of critical regions R_j^+ and R_j^- for $j = 0, \ldots, N$ such that if $X_j \in R_j^-$ then all samples in the pool are classified as HIV negative and released for transfusion, if $X_j \in R_j^+$ then all samples are classified as HIV positive and discarded, and otherwise additional tests are carried out. The critical regions are defined by

$$R_N^+ = \left\{ x: \frac{f_X^{(0,1)}(x;\, \phi,\, \gamma)}{f_X^{(1,1)}(x;\, \phi,\, \gamma)} < \frac{c_{FN}p}{c_{FP}(1-p)} \right\}, \quad (31)$$

$$R_N^- = \left\{ x: \frac{f_X^{(0,1)}(x;\, \phi,\, \gamma)}{f_X^{(1,1)}(x;\, \phi,\, \gamma)} \geqslant \frac{c_{FN}p}{c_{FP}(1-p)} \right\}, \quad (32)$$

$$R_j^+ = \{x_j: c_{FP}m_{j+1}P(\tilde{X}_j \in P_-|X_j = x_j) \leqslant \min\{c_{FN}m_{j+1}P(\tilde{X}_j \in P_+|X_j = x_j), c(m_{j+1}) + E[J_{j+1}(X_{j+1})|X_j = x_j]\}\}, \quad (33)$$

$$R_j^- = \{x_j: c_{FN}m_{j+1}P(\tilde{X}_j \in P_+|X_j = x_j) \leqslant \min\{c_{FP}m_{j+1}P(\tilde{X}_j \in P_-|X_j = x_j), c(m_{j+1}) + E[J_{j+1}(X_{j+1})|X_j = x_j]\}\}, \quad (34)$$

where $f_X^{(0,\,1)}(x;\, \phi,\, \gamma)$ and $f_X^{(1,\,1)}(x;\, \phi,\, \gamma)$ (see Equation (5) for an earlier use of this notation) denote the log-normal OD densities for the HIV negative and HIV positive populations, respectively. Notice that the critical region R_N^- maximizes the power for a simple hypothesis test. Therefore, by the Neyman-Pearson lemma, the proposed classification policy at the individual testing stage not only minimizes the cost for the particular choices of c_{FN} and c_{FP}, it also minimizes the type II error (false positive) for a fixed level of type I error (false negative).

To complete the derivation of the DP algorithm, we must calculate the conditional density of X_j *given* X_{j-1}, denoted by $f_{X_j|X_{j-1}}(x_j|x_{j-1})$, and the conditional probability that $\tilde{X}_j \in P_+$ *given that* $X_j = x_j$. These are given by the following proposition (see Appendix C for proof).

Proposition 1.

$$f_{X_j|X_{j-1}}(x_j|x_{j-1}) = \frac{\left[\int_0^\infty\int_0^\infty \pi_Y^{(m_j)}(y_j)\, f(x_j;\, \phi,\, \gamma|y_j)\, \pi_Y^{(a_j-1)m_j}(y_{j-1})\, f\left(x_{j-1};\, \phi,\, \gamma\middle|\frac{(a_j-1)y_{j-1}+y_j}{a_j}\right) dy_j\, dy_{j-1}\right]}{f_X^{(m_{j-1})}(x_{j-1};\, p,\, \phi,\, \gamma)}, \quad (35)$$

and

$$P(\tilde{X}_j \in P_+|X_j = x_j) = \sum_{k=0}^{m_j} \frac{k}{m_j} \frac{\int_0^\infty \pi_Y^{(k,m_j)}(y)\, f(x_j;\, \phi,\, \gamma|y)\, dy}{f_X^{(m_j)}(x_j;\, p,\, \phi,\, \gamma)}. \quad (36)$$

Proposition 1 shows that in order to numerically implement the dynamic programming algorithm, it is necessary to compute $\pi_Y^{(k,m)}(y)$, for $k = 0, 1, \ldots, m$, as well as the double integral in the numerator of (35). The numerical calculation of $\pi_Y^{(k,m)}(y)$ uses characteristic functions and fast Fourier transforms and is very efficient. In contrast, the numerical calculation of the double integral is cumbersome. Specifically, a very efficient algorithm that calculates the double integral using a two dimensional grid of 256 × 256 takes 20 seconds of CPU time on a Sun Sparc Station 10. Because in the implementation of the DP algorithm it is necessary to repeatedly calculate such double integrals, we conclude that it is practically infeasible to use the DP algorithm to obtain efficient generalized group testing procedures. However, it is possible to significantly speed up

the algorithm if $N = 1$ (i.e., the procedure has two stages) and if the classification rule takes a simple form. This is discussed in the next subsection.

6.2. Structural Properties of the Optimal Policy

Intuitively, one might expect that the optimal classification policy could be characterized by a set of constants $\{c_j^-, c_j^+ : 0 \leq j \leq N\}$ (where $c_N^- = c_N^+$) such that $R_j^- = \{x_j : x_j \leq c_j^-\}$ and $R_j^+ = \{x_j : x_j \geq c_j^+\}$. Such a classification policy for a generalized group testing procedure will be called a *cutoff policy.*

Cutoff policies are particularly convenient, not only because they are easy to implement, but also because they can be exploited to substantially improve the performance of the DP algorithm. More specifically, suppose that to implement the DP algorithm we discretize the continuous state space to n states,

$$0 = x^{(0)} < x^{(1)} < \cdots < x^{(n)} = 1 . \tag{37}$$

In this case, the cutoffs c_j^- and c_j^+ satisfy the following conditions,

$$\max_{0 \leq s \leq n} (x^{(s)} : x^{(s)} \in R_j^-) \leq c_j^- \leq \min_{0 \leq s \leq n} (x^{(s)} : x^{(s)} \notin R_j^-) , \tag{38}$$

$$\max_{0 \leq s \leq n} (x^{(s)} : x^{(s)} \notin R_j^+) \leq c_j^+ \leq \min_{0 \leq s \leq n} (x^{(s)} : x^{(s)} \in R_j^+) . \tag{39}$$

These conditions imply that to obtain the optimal classification rule, it suffices to start from $x^{(0)}$ and progress to higher states until $x^{(s)} \notin R_j^+$, and to start from $x^{(n)}$ and progress to lower states until $x^{(s)} \notin R_j^+$. Thus, we can avoid the cumbersome double integration for all intermediate states $x^{(s)} \notin R_j^+ \cap R_j^-$.

Arnold obtained sufficient conditions ensuring the optimality of a cutoff policy for a simpler group testing problem that possesses only two possible classifications, and assumes a linear pooling model that does not capture within sample variability. Here, we extend his results to the model of Section 6.l. To do that, we need the following monotonicity notion that was introduced by Arnold: the density $f_X^{(m_j)}(x; p, \phi\ \gamma)$ of the OD reading X_j has the Mon(j) property if for all nonincreasing functions $h(x)$, the conditional expectation $E(h(X_j)|X_{j-1} = x)$ is monotone nonincreasing in x. The following proposition provides sufficient conditions for the optimality of the cutoff policy; the proofs to both propositions in this section can be found in Appendix C.

Proposition 2. *A cutoff policy is optimal if the likelihood ratio* $f_X^{(1,1)}(x; \phi, \gamma)/f_X^{(0,1)}(x; \phi, \gamma)$ *is monotone nondecreasing and the density* $f_X^{(m_j)}(x; \phi, \gamma, p)$ *of* X_j *has the Mon*(j) *property for all* j.

The definition of Mon(j) cannot be used for testing whether a density has the required property. Instead, the following proposition can be employed.

Proposition 3. *If for all* $s_1 > s_2$ *the likelihood ratio* $f x_j|x_{j-1}(x|s_1)/f x_j|x_{j-1}(x|s_2)$ *is monotone nondecreasing in* x, *then the density* $f_X^{(m_j)}(x; \phi, \gamma, p)$ *has the Mon*(j) *property.*

It turns out that neither of the sufficient conditions of Propositions 2 and 3 are satisfied by our model. In fact, it can be shown that if $\pi_+(y)$ is an arbitrary discrete density then $f_X^{(1,1)}(x; \phi, \gamma)$ and $f_{X_j|X_{j-1}}(x_j|x_{j-1})$ are complex normal mixtures that do not satisfy the monotone likelihood ratio property. Because in the implementation of the DP algorithm it is necessary to discretize the density $\pi_+(y)$, it follows that the optimal classification rule may not be a cutoff policy.

Nevertheless, in the implementation of the DP algorithm we assume the contrary in order to speed up the performance of the algorithm. Specifically, if we restrict ourselves to cutoff policies then the computation time for a two-stage ($N = 1$) group testing procedure drops from 20 hours of CPU time to only 1 hour! It is worth noting that in an extensive computational study with $\phi = 0$ (see Zenios), the optimal cutoff policy performed nearly as well as the overall optimal policy.

7. HEURISTIC DORFMAN POLICIES

In the Dorfman procedure, a pool of a specified size is tested, after which either every sample in the pool is deemed HIV negative, or every sample in the pool undergoes individual testing. Due to its simplicity and effectiveness, this procedure is frequently used in practice for mass screening programs. In particular, recent field studies (e.g., Behets et al., Emmanuel et al. and Kline et al.) in developing countries demonstrate that such procedures can be used to reduce the cost of HIV screening. The complexity of general group testing strategies, such as the one described in Section 6.1, renders them more vulnerable to human error. Therefore, the improvement achieved by the more complex testing strategy could be offset by the human errors incurred during implementation.

Using the dynamic program of Section 6.1, we can obtain the optimal decision rule for a Dorfman procedure with pool size m by setting $N = 1$ and $n_1 = a_1 = m$, and disallowing the option of discarding a pool that contains more than one sample. However, even this streamlined version of the DP algorithm is computationally intensive. Therefore, we propose a relatively simple method for obtaining an effective Dorfman policy. Our method involves two simplifications: the first replaces the hierarchical pooling model by a simplified pooling model, and the second makes the calculation of the conditional density (35) straightforward by assuming that the outcome of the pooled test is used to calculate the posterior seroprevalence, but not the posterior conditional densities. Clearly, this assumption is not making the most efficient use of the pooled OD reading.

7.1. The Simplified Pooling Model

We propose a simplified pooling model that replaces the two-level hierarchical model by a nonlinear stochastic relationship between the OD reading and the antibody concentration. This model leads to a simple analytical

approximation to the conditional density of the pooled LOD reading *given that the pool consists of k infected and m − k noninfected individuals*. This expression is used in Section 7.2 to develop an algorithm that generates pooling procedures.

Recall that Equation (3) of the hierarchical pooling model (HPM) implies that

$$\ln\left(\frac{E[X|Y]}{1-E[X|Y]}\right) = \gamma \ln(Y)\,. \tag{40}$$

Motivated by (40), we propose the following approximate relationship between the LOD reading, $Z = \ln(x/(1-X))$, and the antibody concentration Y:

$$Z = \gamma \ln(Y) + e\,, \tag{41}$$

where e is a normal random variable with mean zero and variance σ^2. Equation (41) will be called the simplified pooling model SPM.

Although the SPM is not mathematically equivalent to the HPM, it preserves the key characteristics of the HPM: it not only captures the within sample variability (via e) and the between sample variability, but it also predicts a sigmoid relation between the OD reading and the antibody concentration.

Even though the homoscedastic assumption (i.e., σ^2 is not a function of Y) may appear to be overly restrictive, especially in light of the discussion in Section 4, we will see in Section 8 that the pooled testing policies that arise from the SPM are very robust, evidently because the cost function is relatively flat in the neighborhood of the global minimum. Consequently, minor model misspecifications such as this do not appear to adversely affect the performance of the resulting policies.

Now we use (41) to derive the conditional density of the pooled LOD reading *given that the pool consists of k infected and m − k noninfected individuals.* Consider a pool of size m and let $Y_1, \ldots, Y_m$ denote the random antibody concentrations and $Z^{(m)}$ the pooled LOD reading. It is known that the first k individuals are infected and the last $m - k$ individuals are not infected.

The SPM implies that

$$Z^{(m)} = \gamma \ln\left(\frac{Y_1 + \cdots + Y_m}{m}\right) + e\,. \tag{42}$$

The asymptotic density of $Z^{(m)}$ will be derived with the aid of Proposition 1 of Zenios and Wein 1995, which states that as $k, m \to \infty$ such that $k/m \to r$, then

$$\sqrt{m}\left(\frac{Y_1 + \cdots + Y_m}{m} - r\eta_+ - (1-r)\eta_-\right) \to N(0,\, r\nu_+^2 + (1-r)\nu_-^2)\,. \tag{43}$$

From (43) and Cramer's delta method (Billingsley 1987),

$$\sqrt{m}\left(\ln\left(\frac{Y_1 + \cdots + Y_m}{m}\right) - \ln(r\eta_+ + (1-r)\eta_-)\right) \to N\left(0,\, \frac{r\nu_+^2 + (1-r)\nu_-^2}{r\eta_+ + (1-r)\eta_-}\right)^2. \tag{44}$$

Substituting (44) into (42) shows that the conditional density of $Z^{(m)}$, *given that the pool consists of k infected and m − k noninfected individuals*, can be approximated by

$$Z^{(m)}|A_{km} \sim N\left(\gamma \ln\left(\frac{k}{m}\eta_+ + \left(1-\frac{k}{m}\right)\eta_-\right),\; \gamma\left(m\,\frac{k\nu_+^2 + (m-k)\nu_-^2}{(k\eta_+ + (m-k)\eta_-)^2} + \sigma^2\right)\right), \tag{45}$$

where A_{km} is the conditioning event.

The remainder of this subsection is devoted to estimating the parameters η_+, η_-, ν_+, ν_- and σ appearing in (45). We begin with η_+ and ν_+. Recall that the empirical density derived from the individual testing data via (27) generated the estimates $\eta_+ = 21.48$ and $\nu_+ = 53.41$. Although these appear to be the natural choices to adopt in (45), we now explain our reasons for using different estimates.

Equation (45) is expected to provide an accurate approximation when k and m are very large. However, in practice, k and m are expected to be of moderate, if not small, magnitude. Jensen's inequality shows that (45) *underestimates* the dilution effect:

$$\begin{aligned} E[Z^{(m)}|A_{km}] &= \gamma E\left[\ln\left(\frac{Y_1 + \cdots + Y_m}{m}\right)\Big|\, A_{km}\right] \\ &\leq \gamma \ln E\left(\left[\frac{Y_1 + \cdots + Y_m}{m}\Big|\, A_{km}\right]\right) \\ &= \gamma \ln\left(\frac{k}{m}\eta_+ + \left(1-\frac{k}{m}\right)\eta_-\right). \end{aligned} \tag{46}$$

This underestimation may lead to policies with high false negative rates, and we propose a judicious choice of the estimates for η_+ and ν_+ to counteract this effect.

Figure 6 shows that the empirical density for $\pi_+(y)$ has a long right tail and can be roughly fit to a log-normal (see the end of Section 5). This tail leads to the large variance ($\nu_+ = 53.41$) and the inflated mean ($\eta_+ = 21.48$). However, because false negatives are so costly in this setting, we are primarily interested in the left tail of $\pi_+(y)$; individuals in the left tail, when tested in pools, are expected to cause most of the false negatives. Moreover, to counteract (45), if our estimate of the left tail is to err, we prefer it to overestimate the width of the left tail (i.e., overestimate the proportion of HIV positive individuals with low antibody concentrations). Hence, to obtain a conservative (in the above sense) estimate of the left tail of $\pi_+(y)$, we use a normal distribution, which has a wider left tail than a corresponding log-normal. More specifically, we take the mean η_+ to be the median of $\pi_+(y)$ and take ν_+ to be the difference between the median and the 2.5th percentile of $\pi+(y)$, divided by 1.96. For the NRL data, we obtain $\eta_+ = 6.429$ and $\nu_+ = 3.25$.

Now we turn to the parameters η_-, ν_- and σ. In Section 5 we derive the estimates $\eta_- = 0.00015$ and $\nu_- = 0$. Moreover, since $\nu_- = 0$, it follows that

$$Z^{(m)}|A_{0m} \sim N(\gamma \ln(\eta_-), \sigma^2), \quad (47)$$

for all values of m. Thus, $\text{var}(Z^{(m)}|A_{0m}) = \text{var}(Z|A_{01})$ and so $\sigma = \sigma_- = 0.42$. It is worth noting that (47) implies that $E[Z^{(m)}|A_{0m}] = E[Z|A_{01}]$. Hence, $\eta_- = e^{1/\gamma\mu_-} = 0.000133$, which is consistent with the estimate from Section 5. In summary, we use $\eta_- = 0.00015$ $\nu_- = 0$ and $\sigma = 0.42$ in (45).

7.2. The Algorithm

With the aid of (45), we calculate the approximate expected cost of an arbitrary Dorfman procedure, and then find the cost-minimizing policy within this class of procedures.

Consider a Dorfman policy of pool size m applied in a seroprevalence p population, and let Z_i denote the LOD reading of individual i, and $Z^{(m)}$ denote the pooled LOD reading. Suppose that u is the cutoff employed for individual testing and v is the cutoff for group testing.

The expected cost incurred at the group testing stage of the process is

$$\begin{aligned} C_g(v) &= c(n) + c_{FN} \sum_{k=1}^{m} P(A_{km}) P(Z^{(m)} \leq v|A_{km}) \\ &= c(n) + c_{FN} \sum_{k=1}^{m} P(A_{km}) \\ &\quad \Phi\left(\frac{v - \gamma \ln\left(\frac{k}{m}\eta_+ + \left(1-\frac{k}{m}\right)\eta_-\right)}{\sqrt{\gamma^2 \frac{k\nu_+^2 + (m-k)\nu_-^2}{(k\eta_+ + (m-k)\eta_-)^2} + \sigma^2}} \right)^2 k, \end{aligned} \quad (48)$$

where the first term is the testing cost and the second is the misclassification cost of false negatives. The expression for $P(Z^{(m)} \leq v|A_{km})$ follows from (45).

If we assume that the individual LOD readings are a mixture of normals (see Section 2) then the expected cost of an individual test is

$$\begin{aligned} C_i(p, u) &= c(1) + c_{FP}(1-p)P(Z_i \geq u) + c_{FN} p P(Z_i < u) \\ &= c(1) + c_{FP}(1-p)\left(1 - \Phi\left(\frac{u-\mu_-}{\sigma_-}\right)\right) \\ &\quad + c_{FN} p \Phi\left(\frac{u-\mu_+}{\sigma_+}\right), \end{aligned} \quad (49)$$

where Φ is the cumulative standard normal distribution. Notice that $P(Z_i \in P_+|A_{km}) = k/m$ because $Z_1, \ldots, Z_m$ are iid. Therefore, according to our second simplification (i.e., the outcome of the pooled test is used to calculate the posterior seroprevalence, but not the posterior conditional densities) and equation (45), the approximate expected cost incurred at the individual testing stage of the Dorfman procedure is

$$\begin{aligned} C_{ig}(u, v) &= \sum_{k=0}^{m} P(A_{km}) P(Z^{(m)} \geq v|A_{km}) m C_i\left(\frac{k}{m}, u\right) \\ &= \sum_{k=0}^{m} P(A_{km}) \left(1 - \Phi\left(\frac{v - \gamma \ln\left(\frac{k}{m}\eta_+ + \left(1-\frac{k}{m}\right)\eta_-\right)}{\sqrt{\gamma^2 \frac{k\nu_+^2 + (m-k)\nu_-^2}{(k\eta_+ + (m-k)\eta_-)^2} + \sigma^2}} \right)\right) \\ &\quad \cdot m C_i\left(\frac{k}{m}, u\right). \end{aligned} \quad (50)$$

Combining (48) and (50) allows us to conclude that the approximate expected cost per individual is

$$C(m, u, v) = \frac{1}{m}[C_g(v) + C_{ig}(u, v)],$$

or

$$\begin{aligned} C(m, u, v) &= c(m) + c_{FN} \sum_{k=1}^{m} \binom{m}{k} p^k (1-p)^{m-k} k \\ &\quad \Phi\left(\frac{v - \gamma \ln\left(\frac{k}{m}\eta_+ + \left(1-\frac{k}{m}\right)\eta_-\right)}{\sqrt{\gamma^2 \frac{k\nu_+^2 + (m-k)\nu^2}{(k\eta_+ + (m-k)\eta_-)^2} + \sigma^2}} \right) \\ &\quad + \sum_{k=0}^{m} \binom{m}{k} p^k (1-p)^{m-k} \\ &\quad \cdot \left(1 - \Phi\left(\frac{v - \gamma \ln\left(\frac{k}{m}\eta_+ + \left(1-\frac{k}{m}\right)\eta_-\right)}{\sqrt{\gamma^2 \frac{k\nu_+^2 + (m-k)\nu^2}{(k\eta_+ + (m-k)\eta_-)^2} + \sigma^2}} \right)\right) \\ &\quad \cdot m C_i\left(\frac{k}{m}, u\right). \end{aligned} \quad (51)$$

To obtain the proposed Dorfman procedure, we must find m, u and v to minimize (51). This minimization is performed in two stages: First, obtain the optimal cutoffs u and v for every m by solving for the first-order and second-order optimality conditions, and then search among the integers for the optimal group size m. This procedure generates locally optimal solutions, which turn out to be globally optimal in our numerical studies.

8. COMPUTATIONAL RESULTS

In this section we focus on the relative performance of three testing policies: the manufacturer-recommended individual testing policy, the heuristic Dorfman policy developed in Section 7.2, and the optimal Dorfman policy derived from the dynamic programming algorithm. A wide range of scenarios are considered by varying the seroprevalence and false negative cost, and the derived policies

are tested on a Monte Carlo simulation. The cost parameters are estimated in Section 8.1 and the computer implementation is briefly described in Section 8.2. The policies are tested on the simulation model in Section 8.3. In Section 8.4, we apply our model to the data from N'tita et al. (1991).

8.1. The Cost Parameters

We employ the detailed cost estimates contained in the field study of Behets et al. They estimated the material and labor cost of testing a single sample to be \$2.12, and the cost of testing a pool containing $m \geq 2$ samples to be \$2.87 + \$0.083m. Without loss of generality, we normalize these costs so that the cost of testing a single sample is $c(1) = 1$, and the cost of testing a pool of size m is $c(m) = 1.35 + 0.04m$ for $m \geq 2$.

The false positive cost c_{FP} and the false negative cost c_{FN} cannot be estimated as easily. To get a rough estimate for c_{FP}, we note that under the current Red Cross screening protocol, individuals that are found to be HIV positive during an initial ELISA must undergo two additional ELISAs. If at least one of the additional tests is positive, then a highly specific test (Western Blot) is used to verify the individual's serological status. Since ELISA's specificity is more than 0.99, the probability that a noninfected individual with a positive initial ELISA test requires a Western Blot test is approximately (assuming successive ELISA results are independent) $1 - 0.99^2$. The Western Blot test is approximately ten times as costly in materials and labor as an ELISA test. Hence, the expected false positive cost under the Red Cross protocol is $2 + 10(1 - 0.99^2) = 2.199$. This cost may underestimate the true false positive cost because successive ELISA results may be positively correlated, and the Western Blot test may not be available in developing countries; in the latter case, a human cost may be incurred, particularly if test results are reported to individuals. Hence, we choose the conservative estimate of $c_{FP} = 5$.

Since a false negative cost will contaminate the blood supply, these costs are extremely difficult to quantify and are sometimes taken to be arbitrarily high. However, as argued in Section 6, the false negative cost generates a tradeoff between testing cost and test efficiency. Consequently, we consider four different values for c_{FN} (100, 1,000, 5,000 and 10,000), and combine them with seven different values for the seroprevalence p (ranging from 0.001 to 0.15) to generate 28 different scenarios that span a broad range of possible settings.

8.2. Implementation

For each scenario, the simulation model randomly generates antibody levels and OD readings for both individual and pooled samples using the hierarchical model in Section 3 and the distributions and parameters specified in Section 5. The simulation terminated at the first time after 40,000 simulated pools when the standard error for the expected cost dropped below 0.05. To avoid the possibility of sequential dependencies due to any inherent deficiencies of the random number generator, the *ran0* routine described in Press et al. (1988, Chapter 7) was used. All policies were tested on the same random sequence of OD readings.

The heuristic Dorfman policy in Section 7.2 was derived using Maple on a Sun Sparc Station 10. The partial derivatives of $C(m, u, v)$ with respect to u and v were obtained using symbolic differentiation, and the stationary points were identified using the built-in nonlinear solver. Only stationary points lying in the rectangle $[\mu_-, \mu_+] \times [\mu_-, \mu_+]$ were considered, and these points satisfied both the first-order and second-order optimality conditions in all cases. A search over the integers was then employed to obtain the optimal group size m.

The implementation of the dynamic programming algorithm is considerably more complex. The continuous state space must be truncated and discretized: our state space consisted of 200 equally spaced points in [0, 1], with step size 0.05. Simpson's numerical integration rule with both fixed and varying interval size was employed to achieve four digit accuracy. Because it is not feasible to use the DP algorithm to search among the integers to obtain the optimal group size, we decided to adopt the group size of the heuristic Dorfman procedure and. use the DP algorithm to obtain the optimal classification rule. The derived policy is called the DP policy.

8.3. Simulation Results

Main Results. We begin by comparing the manufacturer's individual testing policy, heuristic Dorfman policy and DP policy; later in this subsection, we consider several other policies.

Our main results are reported in Table II, which describes the policies, and Table III, which displays their performance in the simulation study. The first column in Table II enumerates the 28 scenarios, and the next two columns characterize the scenarios. The final column gives the pool size for each scenario. The remaining columns give the LOD cutoff points for both stages (the first stage is the pooled testing stage and the second stage is the individual testing stage) of both Dorfman procedures. As a reference point, recall that the cutoff recommended by the manufacturer of the assays is 0.05, or -2.94 in the LOD scale. For each scenario, Table III gives the 95% confidence interval for the expected total cost of each policy, together with the amount spent per cases prevented.

The following four observations can be extracted from our numerical study:

(1) The DP and heuristic Dorfman procedures are quite similar. They both employ similar cutoff points, with the heuristic Dorfman being consistently more conservative. The performance of the two Dorfman procedures is comparable. The DP procedure outperforms the heuristic for 20 scenarios, and the average suboptimality of the heuristic procedure is only 6.1%. The strong performance of the

Table II
DP Dorfman and Heuristic Dorfman for the 28 Scenarios.

			Cutoff Values				
			Heuristic		DP		
	p	c_{FN}	Stage 1	Stage 2	Stage 1	Stage 2	Pool Size
1	0.001	100	−3.59	−3.12	−3.11	−2.75	35
2		1000	−3.74	−3.32	−3.40	−2.94	21
3		5000	−3.85	−3.45	−3.48	−3.06	15
4		10000	−3.78	−3.49	−3.18	−2.94	12
5	0.005	100	−3.56	−3.20	−3.18	−2.84	15
6		1000	−3.77	−3.40	−3.32	−3.06	12
7		5000	−3.87	−3.55	−3.48	−3.18	9
8		10000	−3.90	−3.61	−3.48	−3.18	8
9	0.01	100	−3.56	−3.22	−3.18	−2.94	12
10		1000	−3.76	−3.43	−3.32	−3.06	9
11		5000	−3.90	−3.57	−3.48	−3.18	8
12		10000	−3.93	−3.64	−3.66	−3.32	7
13	0.025	100	−3.56	−3.26	−3.18	−2.94	8
14		1000	−3.77	−3.46	−3.48	−3.18	7
15		5000	−3.89	−3.62	−3.66	−3.32	6
16		10000	−3.97	−3.76	−3.66	−3.32	6
17	0.05	100	−3.55	−3.29	−3.18	−3.06	6
18		1000	−3.72	−3.50	−3.32	−3.18	5
19		5000	−3.89	−3.64	−3.66	−3.32	5
20		10000	−3.97	−3.72	−3.89	−3.48	5
21	0.1	100	−3.56	−3.32	−3.18	−3.06	5
22		1000	−3.72	−3.55	−3.32	−3.32	4
23		5000	−3.88	−3.68	−3.89	−3.48	4
24		10000	−3.96	−3.75	−5.29	−3.48	4
25	0.15	100	−3.53	−3.35	−3.18	−3.18	4
26		1000	−3.76	−3.55	−3.48	−3.32	4
27		5000	−3.93	−3.70	−5.29	−3.48	4
28		10000	−3.88	−3.80	−5.29	−3.48	3

heuristic Dorfman procedure is noteworthy, since this policy is much easier to derive than the DP procedure.

(2) Group testing is optimal for all 28 scenarios, and significant savings over the manufacturer's procedure are achieved. The expected cost reduction for the DP procedure relative to individual testing ranges from 51.0% (scenario 17) to 89.6% (scenario 1), and the average cost reduction over the 28 scenarios in Table III is 64.8%. For the heuristic Dorfman policy, the expected cost reduction relative to individual testing ranges from 36.0% to 88.9% and averages 63.5%. The monetary testing cost is also significantly reduced; the average reduction relative to individual testing is 48.8% for the DP policy and 50.4% for the heuristic policy. Moreover, although we do not show the numbers in Table III, both Dorfman procedures are highly sensitive and specific. The average sensitivity and specificity over the 28 scenarios are 99.30% and 99.64% for the heuristic Dorfman procedure, and 99.10% and 99.82% for the DP procedure. In contrast, the sensitivity of the individual testing procedure is 98.9% and the specificity 97.5%. Furthermore, the sensitivity of the heuristic Dorfman procedures dropped below 98.75% only for scenario 1, and the sensitivity of the DP policy dropped below 98.75% only for scenarios 1, 2, and 4.

(3) The Dorfman procedures significantly reduce the amount spent per cases prevented: the average reduction relative to individual testing is 49.52% for the heuristic, and 47.78% for the DP procedure. The individual testing policy outperforms the DP policy in scenarios 23, 24, 27, and 28 with respect to this performance measure. The discrepancies in the results between the two performance measures in Table III stem from the fact that the amount spent per case prevented does not take into account false positives. The superior specificity of the pooled testing procedures leads to smaller relative improvements under the cost per cases prevented performance measure.

(4) In Table II, the optimal cutoff values for the pooled testing stage of the two Dorfman procedures are much lower than the cutoffs for the individual testing stage, ranging from −3.11 to −5.24. Hence, the Dorfman procedure is able to maintain its high test accuracy by a judicious choice of cutoffs at each stage; more specifically, *the cutoff level is more conservative at the pooled testing stage to compensate for the dilution effect*. In contrast, previous (field and statistical) researchers in pooled testing have assumed that the *the same cutoff level is used at both stages*; in particular, the cutoff level proposed by the test kit manufacturer is employed at both stages of the Dorfman procedure.

Traditional Dorfman Policy. Observation (4) motivates us to assess the performance of the *traditional* Dorfman policy that has been considered in previous studies. Here, we assume that the manufacturer's cutoff is employed at *both* stages of the procedure. Under this assumption, the optimal value of the pool size m is derived using the cost function (51). The average expected cost increase relative to the heuristic Dorfman is 42.37%, and 51.75% relative to the optimal Dorfman. Hence, the Dorfman policy is significantly compromised by using the manufacturer's cutoff at the pooled testing stage.

Optimal Individual Testing. We also compare our Dorfman policies to the optimal individual testing policy. The results are consistent with our previous conclusions, but the magnitude of the savings is somewhat smaller. The Dorfman policies outperform the individual policies for all scenarios except the high prevalence, high false negative cost scenarios: 20, 23, 24, 27, and 28. The average cost reduction was 26.76% for the heuristic policy and 31.37% for the DP policy.

Predictive Power. To illustrate the predictive power of our model with respect to the traditional Dorfman policy, we consider the study carried out by Behets et al. in Kinshasha, Zaire, where the seroprevalence of the 8,000 samples was 2.44%. The traditional Dorfman procedure with pools of size ten reduced the monetary cost of HIV screening by 56% relative to individual testing; however, six low

Table III
Simulated Performance of the Policies for the 28 Scenarios.

	Expected Total Cost			Cost per Cases Prevented		
	Individual	Heuristic	DP	Individual	Heuristic	DP
1	1.122 ± 0.000	0.123 ± 0.002	0.117 ± 0.002	988.947	125.280	117.757
2	1.132 ± 0.000	0.177 ± 0.008	0.155 ± 0.010	988.947	165.170	129.850
3	1.177 ± 0.001	0.297 ± 0.026	0.198 ± 0.019	988.947	215.278	157.361
4	1.232 ± 0.002	0.299 ± 0.042	0.289 ± 0.050	988.947	215.939	168.919
5	1.126 ± 0.000	0.219 ± 0.002	0.212 ± 0.002	197.789	42.971	41.239
6	1.176 ± 0.001	0.296 ± 0.009	0.270 ± 0.011	197.789	51.411	42.783
7	1.397 ± 0.004	0.419 ± 0.039	0.445 ± 0.050	197.789	60.754	49.395
8	1.673 ± 0.008	0.675 ± 0.050	0.695 ± 0.050	197.789	66.548	52.237
9	1.131 ± 0.000	0.287 ± 0.002	0.282 ± 0.002	98.895	27.885	26.877
10	1.231 ± 0.002	0.356 ± 0.011	0.375 ± 0.016	98.895	31.445	28.100
11	1.673 ± 0.008	0.705 ± 0.050	0.702 ± 0.050	98.895	36.696	29.820
12	2.225 ± 0.015	0.887 ± 0.050	0.960 ± 0.050	98.895	39.231	34.503
13	1.146 ± 0.001	0.425 ± 0.003	0.414 ± 0.003	39.558	16.351	15.737
14	1.395 ± 0.004	0.590 ± 0.023	0.547 ± 0.023	39.558	17.430	16.235
15	2.500 ± 0.019	0.962 ± 0.050	0.978 ± 0.050	39.558	19.037	17.851
16	3.882 ± 0.039	1.619 ± 0.050	1.579 ± 0.050	39.558	20.153	17.868
17	1.171 ± 0.001	0.567 ± 0.004	0.568 ± 0.005	19.779	10.794	10.672
18	1.668 ± 0.008	0.776 ± 0.032	0.788 ± 0.035	19.779	11.241	10.815
19	3.879 ± 0.039	1.537 ± 0.050	1.529 ± 0.050	19.779	12.018	11.451
20	6.642 ± 0.077	2.337 ± 0.050	2.194 ± 0.050	19.779	12.480	12.895
21	1.220 ± 0.002	0.779 ± 0.006	0.773 ± 0.006	9.889	7.296	7.214
22	2.215 ± 0.015	1.123 ± 0.048	1.159 ± 0.050	9.889	7.441	7.185
23	6.636 ± 0.077	2.377 ± 0.050	2.111 ± 0.050	9.889	7.744	8.121
24	12.162 ± 0.096	3.770 ± 0.050	2.984 ± 0.050	9.889	7.931	12.174
25	1.269 ± 0.002	0.929 ± 0.007	0.925 ± 0.007	6.593	5.780	5.723
26	2.761 ± 0.023	1.399 ± 0.050	1.322 ± 0.050	6.593	5.854	5.749
27	9.393 ± 0.096	3.085 ± 0.050	2.604 ± 0.050	6.593	6.047	8.335
28	17.682 ± 0.096	4.936 ± 0.050	4.011 ± 0.050	6.593	6.165	8.936

reactivity individuals were not detected. We used the Monte Carlo simulation model to calculate the performance of the traditional Dorfman procedure. The sensitivity of the procedure was 98.1 ± 0.8% and the expected monetary testing cost was 0.39 ± 0.002 (recall that our testing cost function $c(m)$ is based on the cost model in Behets et al.); hence, our analysis predicts a 61% reduction in monetary testing cost and 3.8 ± 0.4 false negatives. Therefore, the model captures both the magnitude of the cost savings and the extent of the dilution effect as manifested by the low reactivity individuals that are not detectable in pools. Under the heuristic Dorfman policy for scenario 16, the expected monetary testing cost is 0.50 ± 0.0005 and the sensitivity is 99.60 ± 0.03%, and hence the expected number of false negatives is 0.8 ± 0.06. Therefore, we predict that the heuristic Dorfman procedure would not have had any trouble detecting the low reactivity individuals.

The Optimal Group Size. Additional scenarios were considered to generate Figure 8, which provides switching curves depicting the optimal group size (as calculated by the heuristic Dorfman procedure) as a function of both the seroprevalence and the false negative cost. As expected, the group size is a decreasing function of both quantities; if the seroprevalence is high, then large group sizes will contain HIV positive individuals with high probability. Similarly, if the false negative cost is high, then smaller group sizes are required to diminish the impact of dilution. Notice that groups of size two (and sometimes three) are never optimal. This phenomenon is due to the pooled testing cost $c(m) = 1.35 + 0.04m$: the cost of constructing the pool is larger than the limited savings realized when the group size is two or three. Finally, it is worth noting that the *breakeven* (between individual and pooled testing) seroprevalence in Figure 8 ranges from 0.36 when the false negative cost is 100 to 0.30 when the cost is 10,000. In contrast, for the case of perfect binary tests, pooled testing is optimal if and only if the prevalence is less than $(3 - \sqrt{5})/2 \approx 0.382$ (see Unger 1960).

Robustness. Although the proposed Dorfman procedures perform quite well over the 28 scenarios, for implementation purposes it is important to establish their robustness. Here, we summarize the results from a simulation experiment that demonstrates that, indeed, the proposed policies are robust.

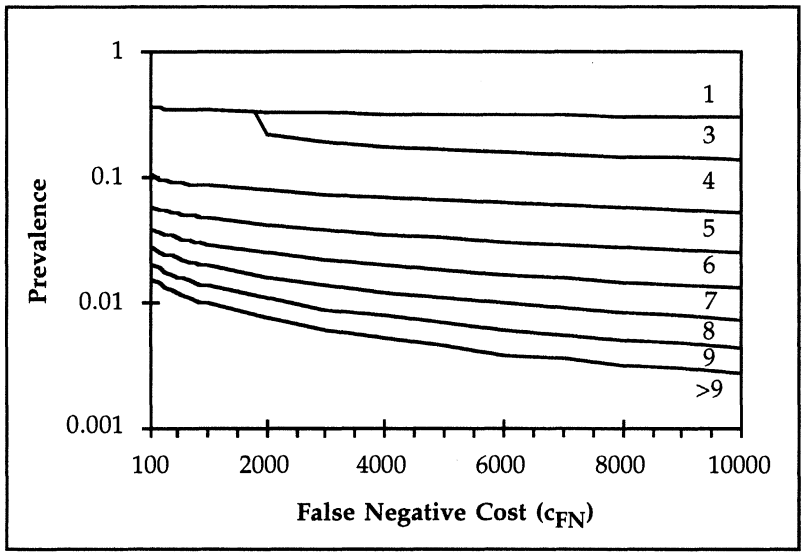

Figure 8. Optimal group size as a function of c_{FN} and p.

Monte Carlo simulation is used to estimate the expected cost for 1694 different policies under each of four scenarios: 9, 10, 21, and 22. The policies include all possible group sizes between 2 and 15, and all possible combinations of cutoffs, in increments of 0.1, in the square grid $[-3, -4] \times [-3, -4]$ (the values of the cutoffs are reported in the logit scale).

A sample of these results are illustrated in Figure 9. This figure gives contour plots for the expected cost as a function of the two cutoffs, for four different combinations of pool sizes and scenarios. Although, the contour plots are fairly irregular because of the uncertainty in the Monte Carlo estimates, the cost surfaces appear to be quite flat in the neighborhood of the minimum. Moreover, if the cutoffs are perturbed by ± 0.2 from the optimal then the increase in the expected cost is at most 10%. These results establish the robustness of the Dorfman policies. Furthermore, they suggest that it is possible to obtain very efficient policies using only approximate models, thus explaining the excellent performance of our proposed policies.

Using the results from this simulation experiment, we can also test the suboptimality of our proposed policies. Specifically, using exhaustive search among the 1694 policies, we identified the best policy for each of the four scenarios. Then, we employed the z-test to compare the estimated cost for the best policy to the estimated cost for the DP and heuristic policies. The test concludes that there is not enough evidence to reject the hypothesis that the DP and heuristic policies are at least as good as the best policies obtained through exhaustive search.

Generalized Pooled Testing Policies. We have thus far focused on the Dorfman policy, due to its simplicity and ease of use. For the $\phi = 0$ case, Zenios investigates the performance of the generalized pooled testing policy for scenarios 6 and 19. In both scenarios, the expected total cost under several different subpool configurations for the generalized multistage procedure was higher than the cost of either of the proposed Dorfman procedures. In summary, although in theory it is possible to obtain generalized multistage policies that outperform the Dorfman procedures, the additional improvement is questionable and is expected to be offset by the difficulty in deriving and implementing these policies.

8.4. Application

The numerical results in Section 8.3 (for example, the switching curves in Figure 8) cannot be universally applied for several reasons. Our numerical results depend upon the HIV positive and HIV negative distributions, which may differ across the world, due to seroconversion rates (a larger seroconversion rate may lead to a fatter left tail of the OD readings for HIV positive individuals) or the particular strain of virus that is prevalent. Also, the testing cost $c(m)$ may depend upon various economic factors that are distinctive to each country. Nevertheless, we will loosely apply our results in a documented setting to obtain a rough estimate of the benefits that are achievable from group testing.

In Kinshasha, Zaire, of the 3,741 units of blood transfused in February 1990, 1,045 (27.9%) were not screened for HIV infection (see N'tita et al. 1991). Assuming that this was a consequence of budget constraints, we can propose an alternative strategy that will reallocate funds across the transfusion centers so that every blood donor can be tested for antibodies to HIV. Since 72.1% of the units were individually tested, the monetary testing cost of the currently implemented policy is 0.721. Seroprevalence in Zaire is estimated to be about 2.5% (see Behets et al.). Suppose that the individual testing policy was employed on 72.1% of the units. Since the sensitivity and specificity of this policy are 98.9% and 97.5%, respectively, the expected number of infected units that are transfused is

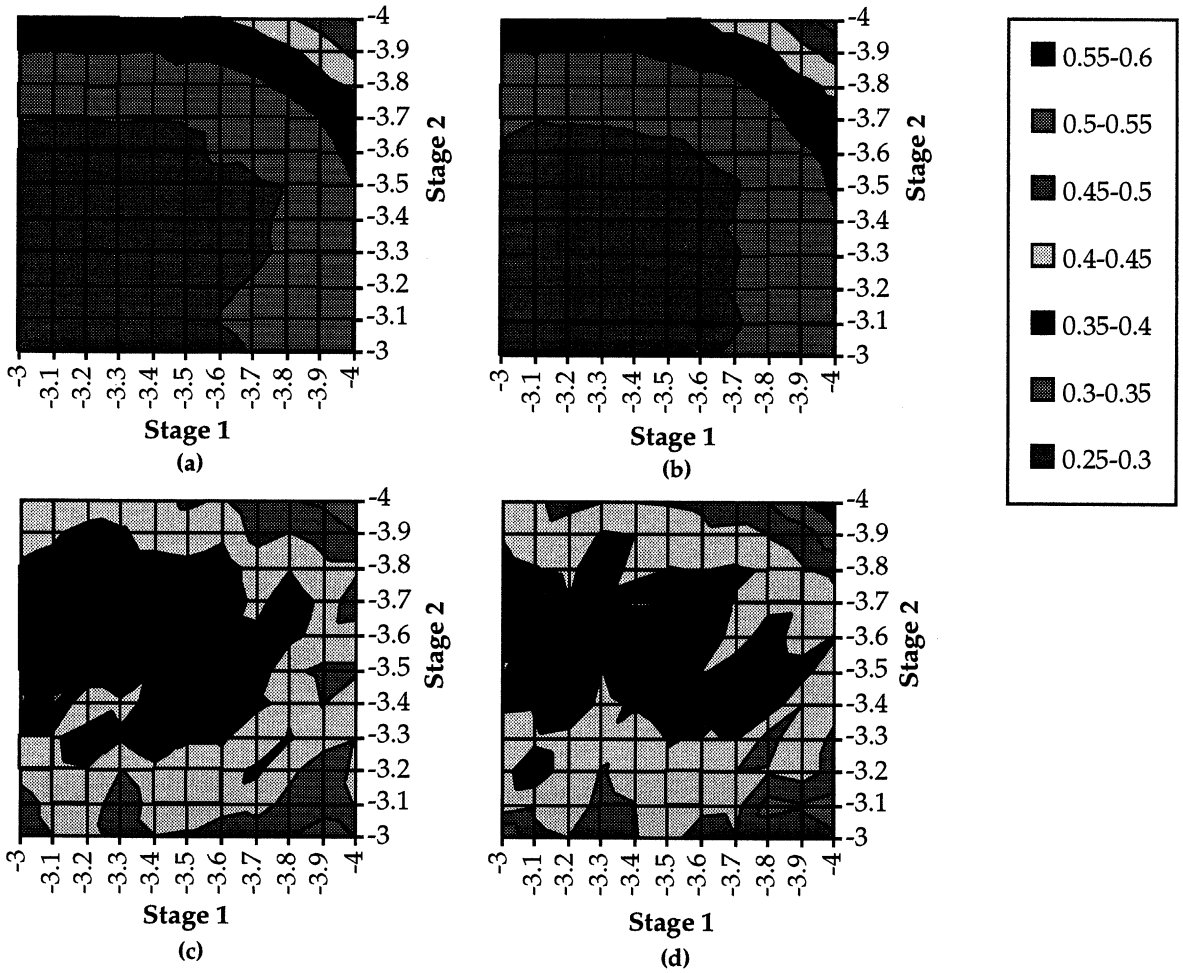

Figure 9. Contour plots for the cost function: (a) $p = 0.001$, $m = 11$, $c_{FN} = 100$, (b) $p = 0.001$, $m = 12$, $c_{FN} = 100$, (c) $p = 0.01$, $m = 9$, $c_{FN} = 1000$, (d) $p = 0.01$, $m = 10$, $c_{FN} = 1000$.

(0.025)(1045) + (0.025)(0.0011)(2696) = 26.87, and the expected number of false positives is (0.975)(0.025)(2696) = 65.72. If we use the heuristic Dorfman policy under scenario 15, then the monetary testing cost is 0.43, the sensitivity is 99.40%, and the specificity is 99.90%. Hence, the expected number of infected units that are transfused is (0.025)(0.0040)(3741) = 0.37, and the expected number of false positives is (0.975)(0.0001)(3741) = 0.36. In summary, pooled testing in this setting reduces the monetary testing cost by 40.3% *and* reduces the expected number of infected units transfused from 27 to essentially zero. It is clear that pooled testing, if used properly, can save hundreds of lives worldwide.

9. CONCLUDING REMARKS

We have developed a statistical model that captures the dilution effect that occurs when HIV positive sera are pooled with HIV negative sera. The hierarchical pooling model (2) and (3), and the generalized linear models (11) and (21) develop new insights into the nature of the dilution effect and reduce the heteroscedasticity problem that has plagued the traditional regression models obtained through a purely empirical approach. These models and the simplified pooling model (41) may be useful for other applications besides group testing, and our general approach may be applicable whenever pooled testing is used to identify a disease or contaminant in a liquid.

Our simulation results suggest that the heuristic Dorfman policy derived in Section 7.2 provides a cost-effective, accurate, and relatively simple alternative to currently implemented HIV screening protocols. While existing field studies and mathematical analyses assume that the same cutoff point is used to classify the pool at both stages of the Dorfman procedure, our analysis shows that only by selecting a different cutoff point at each stage can we ensure that the sensitivity of the test is not compromised. HIV testing is also extensively used for seroprevalence estimation; in a companion paper (Zenios and Wein), we show how the hierarchical pooling model developed here can be employed to derive efficient seroprevalence estimates.

Our analysis has two shortcomings. First, the models do not explicitly consider the effect of *window period* donors; these HIV-infected donors have not yet generated anti-HIV antibodies. In practice, these donors are the major source of transfusion related HIV infections in settings where individual testing is used. It is possible to modify

our analysis to study this effect. To do that, it is necessary to develop a time series model for the immune response of recently infected individuals, which can then be embedded into the hierarchical model. Intuitively, the dilution effect exacerbates the window period effect under pooled testing. However, our analysis shows that it is possible to counteract the window period effect by lowering the cutoff point. The time series model can be used to calculate the change in the cutoff that is necessary to counteract the dilution effect. The development of such a model is left for future research.

The second shortcoming pertains to our model validation efforts. The models were validated using data from a variety of field studies. Although these validation results are promising, our analysis in Section 4.2 may suffer from overfitting: the model has 14 parameters, and the sample size is 44. Consequently, additional pooling data is required to reach a stronger conclusion. We leave subsequent validation efforts for future research. However, if the realism of our hierarchical model can be substantiated by future validation efforts, then the simulation results in this paper provide strong evidence for the adoption of the proposed pooling policies to safeguard the integrity of the blood supply, and consequently reduce the spread of the AIDS epidemic, in areas of the world where individual testing is not a viable option.

APPENDIX

A. Analysis Leading to Equation (3)

Whereas the existing literature has taken a purely empirical approach to the modeling of ELISA, we start by developing a probabilistic model for the second level of the hierarchical model. This model, which is displayed in (55), is characterized by a hyperbolic relation between $E[X|Y = y]$ and the antibody concentration y, as well as state dependent nonlinear noise. The model validation phase led us to consider Equation (3), which retains the qualitative features of (55) but provides a better fit to the data.

Our model is influenced by Fisher (1922), who developed a probabilistic model to estimate the number of bacteria in a sample of water or soil. We begin by stating a set of assumptions, which are divided into four groups that correspond to the four stages of ELISA: (a) binding of anti-HIV antibodies, (b) binding of secondary antibodies, (c) enzymatic reaction, and (d) translation of color change to OD level.

Binding of Anti-HIV Antibodies. We hypothesize that these binding events follow a Poisson process. This involves the following assumptions:

A1. There is a very large number of antigens attached to the well. Typically, this number is greater than 10^6.

A2. Although more than one antibody can bind to a specific antigen, only one antibody can bind to a specific epitope of an antigen.

A3. The probability, q, that an antibody binds to a specific epitope is very small.

A4. Binding events in neighboring epitopes are independent.

A5. The probability q is *linearly* proportional to the antibody concentration in the blood serum. The proportionality constant, denoted by h, can vary among different individuals due to differences in the binding properties of the antibodies. The units for the antibody concentration are ng/ml, and the units for h are ml/ng.

Secondary Binding

A6. The concentration of secondary antibodies is sufficient to saturate *all* immunoglobulins.

Enzymatic Reaction. The mechanism of the enzymatic reaction is very complex and only partially understood, and we adopt a simplistic description. According to this description, the secondary antibodies carry an "imaginary" flag that is raised by the enzymatic reaction. The resulting color change is proportional to the total number of raised flags.

A7. The enzymatic reaction raises all flags.

OD reading. The spectrometer records an OD level by detecting the raised flags.

A8. Because of limited resolution, the spectrometer does not detect all flags. Instead, it subdivides the well into k hypothetical subwells, and measures the total number of subwells that carry flags. Intuitively, the resolution of the spectrometer is expected to be increasing in k.

A9. The OD level is linearly proportional to the number of detected flags.

A10. The maximum OD reading is obtained when all subwells carry a flag.

We conferred with several specialists and although some assumptions were accepted unequivocally, some others raised serious disputation. More specifically, assumptions A3, A5, A7, and A9 appear to raise the most serious objections:

Assumption A3: This assumption may suggest that the total number of antibodies is much smaller than the number of antigens, something that does not hold in practice. Nevertheless, A3 may hold even if the number of antibodies is greater than the number of antigens. In this case, A3 implies that the binding probability of the antibodies is *extremely* small, something that is apt to hold in practice.

Assumption A5: There are some concerns regarding the linearity assumption because the exact form of this relationship is not known. However, we later show that incorporating a specific nonlinear relationship between q and the antibody concentration leads to a model that is very similar to (3).

Assumption A7: The main objection is that this assumption is overly simplistic. In reality, the reaction is governed by complicated kinetics that we do not attempt to model here.

Assumption A9: Again, the objection concerns the linearity assumption. However, in practice the test is calibrated to give a linear relation for a subregion of the possible OD readings.

We should point out that although the final model (after generalizing assumption A5) provides a good fit to the data, this does not imply that the ten assumptions hold; it only suggests the plausibility of these assumptions. Moreover, because the subwells are a modeling artifice, the reader should not infer anything about k from the data.

Assuming that the antigens are uniformly bound on the well, and that the epitopes are uniformly distributed on the surface of the antigens, we let m denote the number of epitopes on every subwell. Suppose that a blood serum of antibody concentration y is added to the well, and let N_s represent the number of antibodies attached to the antigens on subwell $s = 1, \ldots, k$.

Our model development consists of three main steps. First, we find the probability distribution for

$$S^{(k)} = (I_{\{N_1 \geq 1\}} + \cdots + I_{\{N_k \geq 1\}})/k\,,$$

which is the fraction of subwells with immunoglobulins, in terms of the unknown binding probability q. Then we use assumption A5 to relate the unknown binding probability q to the unknown antibody concentration in the diluted serum. Finally, we use assumptions A6–A10 to relate $S^{(k)}$ to the normalized OD reading. Combining these three steps yields our basic model.

By our comments above, $N_1, \ldots, N_k$ are independent binomial random variables with size parameter m and success probability q. By A3 and the law of rare events, the binomial random variable can be approximated by a Poisson random variable with parameter $q_{(m)} = 1 - (1 - q)^m \approx mq$, which implies $P(N_s = 0) = e^{-q_{(m)}}$ and $P(N_s \geq 1) = 1 - e^{-q_{(m)}}$. Invoking the Central Limit Theorem, we obtain

$$S^{(k)} \overset{law}{\to} N\left(1 - e^{-q_{(m)}}, \frac{1}{k} e^{-q_{(m)}}(1 - e^{-q_{(m)}})\right) \quad \text{as } k \to \infty\,. \tag{52}$$

Since $\Sigma_{s=1}^{k} I_{\{N_s \geq 1\}}$ is a binomial random variable, the distribution of $S^{(k)}$ is well approximated by the normal distribution even for relatively small values of k (typically $k \geq 15$).

Assumption A5 implies that $q = hy$, and thus $q_{(m)} = hmy$, which relates the binding probability to the antibody concentration in the diluted serum. Because, neither h nor m nor y are observable, and because their net effect is multiplicative in nature, we will hereafter refer to their product as the antibody concentration y. With this new definition, we have

$$q_{(m)} = y\,. \tag{53}$$

Let X denote the normalized OD reading of the sample. By A6, all immunoglobulins will be detected by a secondary antibody, and by A7 all flags will be raised. Assumption A8 implies that $S^{(k)}$ gives the total number of detected flags, and by A9, $X = wS^{(k)}$, where w is the constant of proportionality. Assumption A10 implies that the maximum normalized OD reading, $X = 1$, is attained when antibodies are bound on all subwells. In this case, $S^{(k)} = 1$, and so $w = 1$ and

$$X = S^{(k)}\,. \tag{54}$$

Combining Equations (52)–(54) gives the basic probabilistic model relating the normalized OD reading to the antibody concentration:

$$(X|Y = y) \sim N(1 - e^{-y},\ \phi e^{-y}(1 - e^{-y}))\,, \tag{55}$$

where the dispersion parameter $\phi = 1/k$.

This model provides two major insights: the relation between y and $E[X|Y = y]$ is hyperbolic,

$$E[X|Y = y] = 1 - e^{-y}\,, \tag{56}$$

and the variance is state-dependent,

$$E[(X - E[X|Y])^2|Y = y] = \phi E[X|Y - y](1 - E[X|Y = y])\,. \tag{57}$$

We now explain how this model led us to the empirical model (3). First we attempted to validate (55) on the dilution series data. Repeating the analysis of Section 4.1, but with (3) replaced by (55), yields the generalized linear model

$$\log(-\log(1 - E[X_{ij}])) = \alpha_i - (\ln(2) - \epsilon I(j > k))j, \quad \text{and} \tag{58}$$

$$\mathrm{Var}(X_{ij}) = \phi E[X_{ij}](1 - E[X_{ij}])\,. \tag{59}$$

This model did not provide a good fit to the data. Although it appeared to stabilize the state dependent residuals, the linear predictor (58) was not adequate; see Zenios (1995) for details. These empirical results suggested the following improvement on (58):

$$\log(-\log(1 - E[X_{ij}])) = \alpha_i - (\gamma \ln(2) - \epsilon I(j > k))j\,, \tag{60}$$

which generalizes $E[X|Y = y]$ to include both sigmoid ($\gamma > 1$) and hyperbolic ($0 \leq \gamma \leq 1$) functions. No improvement on (59) was deemed necessary, because the proposed variance function appeared to stabilize the residuals. We also observe that instead of the complementary log-log link of (59) we can either adopt the logit link,

$$\log\left(\frac{E[X_{ij}]}{1 - E[X_{ij}]}\right) = \alpha_i - (\gamma \ln(2) - \epsilon I(j > k))j\,, \tag{61}$$

or the probit link,

$$\Phi^{-1}([E[X_{ij}]) = \alpha_i - (\gamma \ln(2) - \epsilon I(j > k))j\,. \tag{62}$$

These link functions are frequently used to model sigmoid or hyperbolic response curves and give consistent results.

The final step was to fit these three models to the data. The statistical analysis revealed that the proposed models provide adequate fit to the data, and did not reveal any detectable differences between the three models. Solely for clarity of exposition, we decided to concentrate on the logit GLM (61) and (59), which is equivalent to the GLM (11) and (12). Tracing back the steps that lead to (11) and (12), we see that an appropriate model for $X|Y$ is

$$(X|Y=y) \sim N\left(\frac{y^\gamma}{1+y^\gamma}, \phi\frac{y^\gamma}{(1+y^\gamma)^2}\right), \tag{63}$$

which is equivalent to (3). Hence, the proposed improvement leading from (55) to (3) can be accomplished by altering assumption A5 so that the probability q is linearly proportional to y^γ, where the new proportionality constant $\bar{h}$ is related to the old proportionality constant h via $\bar{h} = h^{1/\gamma}$.

In summary, we developed the probabilistic model (55) from first principles. This model was validated on dilution series data and the analysis revealed possibilities for model improvement. The improved model coincides with model (3), and preserves the form of the state-dependent variance function (57), while retaining the hyperbolic form in (56).

B. Feasibility Region for ϕ

In this Appendix, Var(X) is calculated and Equation (26) is derived. We start by observing that $E(eT) = E(E(eT|T)) = E(TE(e|T)) = 0$. Hence, the covariance of T and e is zero, and Var(X) = Var(T) + Var(e). Moreover, since $E(e^2|T) = \phi T(1-T)$, it follows that

$$\text{Var}(X) = (1-\phi)\text{Var}(T) + \phi E(T)(1-E(T)). \tag{64}$$

Because $E(X) = E(T) + E(e)$ and $E(e) = E(E(e|T)) = 0$, we have

$$E(X) = E(T). \tag{65}$$

Substituting (65) into (64) yields

$$\text{Var}(X) = (1-\phi)\text{Var}(T) + \phi E(X)(1-E(X)). \tag{66}$$

Setting Var(T) ≥ 0 yields Equation (26).

C. Main Proofs

Proof of Proposition 1. Let $Y_j(i_1, \ldots, i_j)$ denote the antibody concentration in the pool consisting of the individuals with the first j indices given by $i_1, \ldots, i_j$, and let $T_j(i_1, \ldots, i_j)$ be the total number of infected individuals in that pool. For brevity of notation, shorten $Y_j(i_1, \ldots, i_j)$ to Y_j and $T_j(i_1, \ldots, i_j)$ to T_j. In addition, let $f_{X_j, X_{j-1}}(x_j, x_{j-1})$ denote the joint density of X_j and X_{j-1}.

Bayes' rule shows that

$$f_{X_j|X_{j-1}}(x_j|x_{j-1}) = \frac{f_{X_j,X_{j-1}}(x_j, x_{j-1})}{f_X^{(m_j)}(x_j; p, \phi, \gamma)}. \tag{67}$$

Let us express the numerator of (67) in terms of more elementary quantities. If we condition on the antibody concentration Y_j, then the law of total probability gives

$$f_{X_j,X_{j-1}}(x_j, x_{j-1}) = \int_0^\infty f_{X_j,X_{j-1}|Y_j}(x_j, x_{j-1}|y_j)\pi_Y^{(m_j)}(y_j)\,dy_j, \tag{68}$$

where $f_{X_j, X_{j-1}|Y_j}(x_j, x_{j-1}|y_j)$ is the conditional joint density of X_j and X_{j-1}, given $Y_j = y_j$. Bayes' rule shows that

$$f_{X_j, X_{j-1}|Y_j}(x_j, x_{j-1}|y_j) = f_{X_j|Y_j}(x_j|y_j) f_{X_{j-1}|Y_j, X_j}(x_{j-1}|y_j, x_j), \tag{69}$$

and because X_{j-1} depends on X_j only through the antibody concentration Y_j, it follows that

$$f_{X_j, X_{j-1}|Y_j}(x_j, x_{j-1}|y_j) = f(x_j; \phi, \gamma|y_j) f_{X_{j-1}|Y_j}(x_{j-1}|y_j). \tag{70}$$

It now remains to compute $f_{X_{j-1}|Y_j}(x_{j-1}|y_j)$, the conditional density of X_{j-1}, *given* Y_j. First, observe that

$$Y_{j-1}(i_1, \ldots, i_{j-1}) = \frac{Y_j(i_1, \ldots, i_{j-1}, 1)}{a_j} + \frac{a_{j-1}}{a_j}\frac{\sum_{i_j=2}^{a_j} Y_j(i_1, \ldots, i_j)}{a_{j-1}}. \tag{71}$$

If we let $\hat{Y}_{j-1}$ denote $\sum_{i_j=2}^{a_j} Y_j(i_1, \ldots, i_j)/a_{j-1}$ then we can condition on $\hat{Y}_{j-1}$ and use the law of total probability to obtain

$$f_{X_{j-1}|Y_j}(x_{j-1}|y_j) = \int_0^\infty \pi_Y^{(a_j-1)m_j}(y_{j-1}) f\left(x_{j-1}\,\middle|\,\frac{y_j + a_{j-1}y_{j-1}}{a_j}\right). \tag{72}$$

Substituting (72) into (70) and then (70) into (68) gives (35).

The quantity $P(\tilde{X}_j \in P_+|X_j)$ is derived as follows:

$$\begin{aligned} P(\tilde{X}_j \in P_+|X_j) &= \frac{E[T_j|X_j]}{m_j} \\ &= \sum_{k=0}^{m_j}\frac{k}{m_j}\frac{P(T_j = k|X_j)}{f_X^{(m_j)}(x_j; p, \phi, \gamma)} \\ &= \sum_{k=0}^{m_j}\frac{k}{m_j}\frac{\int_0^\infty \pi_Y^{(k,m_j)}(y) f(x_j; \phi, \gamma|y)\,dy}{f_X^{(m_j)}(x_j\, p, \phi, \gamma)}. \end{aligned} \tag{73}$$

Proof of Proposition 2. The proof is by induction. First observe that Equation (30) implies that the dynamic programming algorithm for simple policies takes the form

$$\begin{aligned} J_j(X_j) = \min\{&a_{j+1}(c(m_{j+1}) + E[J_{j+1}(x_{j+1})|X_j]), \\ &m_j c_{FN} P(\tilde{X}_j \in P_+|X_j), \\ &m_j c_{FP} P(\tilde{X}_j \in P_-|X_j)\} \\ &\text{for } j = 0, \ldots, N-1. \end{aligned} \tag{74}$$

Let $\hat{J}_j(X_j) = J_j(X_j) - m_j c_{FN} P(\tilde{X}_j \in P_+|X_j)$, and $F(X_N) = c_{FP}P(\tilde{X}_N \in P_-|X_N) - c_{FN}P(\tilde{X}_N \in P_+|X_N)$. By the law of

total probability and the fact that all individuals in the pool are indistinguishable,

$$\begin{aligned} &a_{j+1}E[J_{j+1}(X_{j+1})|Y_j] - m_j c_{FN} P(\tilde{X}_j \in P_+|X_j) \\ &= a_{j+1}E[J_{j+1}(X_{j+1}) - m_{j+1} c_{FN} P(\tilde{X}_{j+1} \in P_+|X_{j+1})|X_j] \\ &= a_{j+1}E[\hat{J}_{j+1}(X_{j+1})|X_j], \quad \text{and} \end{aligned} \tag{75}$$

$$\begin{aligned} &m_j(c_{FP}P(\tilde{X}_j \in P_-|X_j) - c_{FN}P(\tilde{X}_j \in P_+|X_j)) \\ &= m_j E[c_{FP}P(\tilde{X}_N \in P_-|X_N) - c_{FN}P(\tilde{X}_N \in P_+|X_N)|X_j] \\ &= m_j E[F(X_N)|X_j]. \end{aligned} \tag{76}$$

Subtracting $m_j c_{FN} P(\tilde{X}_j \in P_+|X_j)$ from both sides of (74) yields

$$\begin{aligned} \hat{J}_j(X_j) = \min\{&a_{j+1}(c(m_{j+1}) \\ &+ E[\hat{J}_{j+1}(X_{j+1})|X_j]),\ 0,\ m_j E[F(X_N)|X_j]\}. \end{aligned} \tag{77}$$

Straightforward algebraic manipulations show that

$$F(X_N) = \frac{c_{FP}(1-\pi) - c_{FP}p\dfrac{f_X^{(1,1)}(x;\ \phi,\ \gamma)}{f_X^{(0,1)}(x;\ \phi,\ \gamma)}}{(1-p) + p\dfrac{f_X^{(1,1)}(x;\ \phi,\ \gamma)}{f_X^{(0,1)}(x;\ \phi,\ \gamma)}}; \tag{78}$$

hence, $F(X_N)$ is monotone nonincreasing by the assumed monotonicity of $f_X^{(1,1)}(x;\ \phi,\ \gamma)/f_X^{(0,1)}(x;\ \phi,\ \gamma)$. Since $\hat{J}_N(X_N) = \min\{0, F(X_N)\}$, the function $\hat{J}_N(X_N)$ is monotone nonincreasing and the unique root of $F(X_N)$ gives the optimal cutoff c_N^- for stage N.

To prove inductively that $R_j^- = \{x{:}x \leq c_j^-\}$ for some c_j^-, let us assume that $\hat{J}_{j+1}(X_{j+1})$ is monotone nonincreasing. Then by the Mon(j) property, the nonzero terms on the right side of (77) are also monotone nonincreasing. Therefore, there exists c_j^- (the minimum of the roots of the two terms) such that $\hat{J}_j(X_j) = 0$ if and only if $X_j < c_j^-$; thus, $R_j^- = \{z{:}z \leq c_j^-\}$. Moreover, $\hat{J}_j(X_j)$ is monotone nonincreasing. This completes the first part of the proof. Similar arguments establish that $R_j^+ = \{z{:}z \leq c_j^+\}$.

Proof of Proposition 3. This is a corollary of some elementary stochastic ordering results. First, let $W = (X_j|X_{j-1} = s_1)$ and $V = (X_j|X_{j-1} = s_2)$. Proposition 8.4.3 of Ross (1982) implies that W is stochastically greater than V, and proposition 8.1.2 of Ross implies that $E[f(W)] \geq E[f(V)]$ for all nondecreasing functions $f{:}\Re \to \Re$. This implies that

$$E[f(X_j)|X_{j-1} = s_1] \geq E[f(X_j)|X_{j-1} = s_2]. \tag{79}$$

ACKNOWLEDGMENT

We are very grateful to Barbara Cahoon-Young, Elizabeth Dax, Esther de Gourville, and Richard George for providing data. We also thank Karla Ballman, Barbara Cahoon-Young, Elizabeth Dax, Richard George, David Heymann, Richard Kline, Eugene Litvak, Sheila Mitchell, Peter Page, Constantia Petrou, Chris Stowell, Hiko Tamashiro, and Guido van der Groen for helpful discussions about various aspects of pooled testing. Finally, we thank John Karon and Glenn Satton, whose numerous comments on an earlier version of this paper led to significant improvements. This research is supported by National Science Foundation grant DDM-9057297 and American Foundation for AIDS Research (AmFAR) grant 02100-15-RG.

REFERENCES

ARNOLD, S. F. 1977. Generalized Group Testing. *Ann. Statistics* **5,** 1170–1182.

BECKER, N. G., L. F. WATSON, AND J. B. CARLIN. 1991. A Method of Non-parametric Back-projection and Its Application to AIDS Data. *Stat. Medicine* **10,** 1527–1542.

BEHETS, F., S. BERTOZZI, M. KASALI, M. KASHAMUKA, L. ATIKALA, C. BROWN, R. W. RYDER, AND C. QUINN. 1990. Successful Use of Pooled Sera to Determine HIV-1 Seroprevalence in Zaire with Development of Cost-Efficiency Models. *AIDS* **4,** 737–741.

BILLINGSLEY, P. 1987. *Probability and Measure*. Wiley, New York.

BURNS, K. C. AND C. A. MAURO. 1987. Group Testing with Test Error as a Function of Concentration. *Comm. Stat.-Theory Meth.* **16,** 2821–2837.

CAHOON-YOUNG, B., A. CHANDLER, T. LIVERMORE, J. GAUDINO, AND R. BENJAMIN. 1989. Sensitivity and Specificity of Pooled Versus Individual Sera in Human Immunodeficiency Virus Antibody Prevalence Study. *J. Clinical Microbiology* **27,** 1893–1895.

CAHOON-YOUNG, B., A. CHANDLER, T. LIVERMORE, J. GAUDINO, AND R. BENJAMIN. 1992. Optimal Pool Size for Determination of HIV Prevalence in Low Risk Populations. Presented at the HIV/AIDS Surveillance Workshop, South San Francisco, CA.

COUROUCE, A. M. 1987. Latency Preceding Seroconversion in Sexually Transmitted HIV Infection. *Lancet* **2(8566),** 1025.

COX, D. R. AND D. V. HINKLEY. 1974. Theoretical Statistics. Chapman and Hall, London.

DAVIDIAN, M. D. AND D. M. GILTINAN. 1994. Assays for Recombinant Proteins: A Problem in Non-Linear Calibration. *Stat. Medicine* **13,** 1165–1180.

DAX, E. M. 1993. Director, National HIV Reference Laboratory, Melbourne, Australia. Private Correspondence.

DE GOURVILLE, E. 1992. Research Associate, CAREC, Trinidad W.I. Private Correspondence.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion). *J. Royal Stat. Soc. Series B* **339,** 1–22.

DORFMAN, R. 1943. The Detection of Defective Members of Large Populations. *Ann. Math. Stat.* **44,** 436–441.

EMMANUEL, J. C., M. T. BASSETT, H. J. SMITH, AND J. A. JACOBS. 1988. Pooling of Sera for Human Immunodeficiency Virus (HIV) Testing: An Economical Method for use in Developing Countries. *J. Clinical Pathology* **41,** 582–585.

FISHER, R. A. 1922. On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. R. Soc.* **222,** 309–368.

GEORGE, R. J. 1992. Chief. Dev. Technology Section, Division HIV/AIDS, Center for Disease Control, Atlanta, GA. Private Correspondence.

GEORGE, R. J. AND G. SCHOCHETMAN. 1985. Serological Tests for the Detection of HIV Infection. In *AIDS Testing,*

Methodology and Management Issues. G. Schochetman and J. R. George (eds.). Springer-Verlag, New York, 49–69.

HASTIE, T. J. AND D. PREGIBON. 1992. Generalized Linear Models. In *Statistical Models in S*, J. M. Chambers and T. J. Hastie (eds.). Wadsworth & Brooks/Cole Computer Science Series, CA 195–247.

HULL, B. 1991. Serum Pooling for HIV Screening in Trinidad and Tobago. Carribean Epidimology Center Technical Report.

HWANG, F. K. 1976. Group Testing with a Dilution Effect. *Biometrika* **63,** 671–673.

HWANG, F. K. 1984. Robust Group Testing. *J. Quality Technology* **16,** 189–195.

JOHNSON, N. L., S. KOTZ, AND X. WU. 1991. *Inspection Errors for Attributes in Quality Control*. Chapman and Hall, London.

KLINE, R. L., T. A. BROTHERS, R. BROOKMEYER, S. ZEGGER, AND T. C. QUINN. 1989. Evaluation of Human Immunodeficiency Virus Seroprevalence in Population Surveys using Pooled Sera. *J. Clinical Microbiology* **27,** 1449–1452.

LEDRO-MONROY, G. AND E. ARCHBOLD. 1990. HIV Serum Pooling Study. Cruz Roja Ecuatoriana.

LITVAK, E., X. M. TU, AND M. PAGANO. 1994. Screening for the Presence of a Disease by Pooling Sera Samples: Simplified Procedures. *J. American Stat. Assoc.* **89,** 424–434.

MADANSKY, A. 1988. *Prescriptions for Working Statisticians*. Springer-Verlag, New York.

MCCULLAGH, P. AND J. A. NELDER. 1989. *Generalized Linear Models, 2nd Edition*. Chapman and Hall, London.

N'TITA, I., K. MULUNGA, C. DULAT, D. LUSAMBA, T. REHLE, R. KORTE, AND H. JAGGER. 1991. Risk of Transfusion-Associated HIV Transmission in Kishasha, Zaire. *AIDS* **5,** 437–439.

PRESS, W. H., B. P. FLANNEY, S. A. TEUKOLSKY, AND W. T. VETTERLING. 1988. *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, Cambridge.

ROSS, S. M. 1982. *Stochastic Processes*. Wiley, New York.

SOBEL, M. AND P. A. GROLL. 1959. Group Testing to Eliminate Efficiently All Defectives in a Binomial Sample. *Bell System Tech. J.* **38**. 1179–1252.

TAMASHIRO, H., W. MASKILL, J. EMMANUEL, A. FAUQUEX, P. SATO, AND D. HEYMANN. 1993. Reducing the Cost of HIV Antibody Testing. *Lancet* **342,** 87–90.

THOMPSON, K. H. 1962. Estimation of the Proportion of Vectors in a Natural Population of Insects. *Biometrics* **18,** 568–578.

TIJSSEN, P. 1985. *Laboratory Techniques in Biochemistry and Molecular Biology*, Volume **15**. Elsevier, Amsterdam.

UNGER, P. 1960. The Cutoff Point for Group Testing. *Comm. Pure and Appl. Math.* **13,** 49–54.

WILLIAMS, D. 1991. *Probability with Martingales*. Cambridge University Press, Cambridge, England.

ZENIOS, S. A. 1996. Health Care Applications of Optimal Control. Ph.D. thesis, Operations Research Center, MIT, Cambridge, MA.

ZENIOS, S. A. AND L. M. WEIN. 1995. Pooled Testing for HIV Prevalence Estimation: Exploiting the Dilution Effect. Working Paper, Sloan School of Management, MIT, Cambridge, MA.