1. Web Appendix A

Distribution of Elapsed Times in States 1 and 2

Let \( Y(t) = 0, 1, 2, 3 \) denote a subject’s state at time \( t \). Then for \( 0 < y < L_1^* \),

\[
F_1(y) = P[Y_1 < y \mid Y(t) = 1] = P[T > t - y \mid T < t < T + L_1],
\]

so that

\[
\pi_1(t)F_1(y) = P[\max(t - y, t - L_1) < T < t] = P[t - \min(y, L_1) < T < t]
\]

\[
= \int_0^y g_1(v)[F(t) - F(v)]dv + \int_y^{L_1^*} g_1(v)[F(t) - F(t - y)]dv
\]

\[
= F(t)G_1(y) - \int_0^y g_1(v)F(t - v)dv + [F(t) - F(t - y)][1 - G_1(y)].
\]

Thus

\[
\pi_1(t)f_1(y) = F(t)g_1(y) - g_1(y) - g_1(y)F(t-y) - [F(t)-F(t-y)]g_1(y) + [1-G_1(y)]f(t-y),
\]

and hence

\[
f_1(y) = \frac{[1 - G_1(y)]f(t-y)}{\pi_1(t)} \quad \text{for} \quad 0 < y < L_1^*.
\]

The expression for \( f_2(t) \) follows using similar arguments.
2. Web Appendix B

Analytical expressions for the elements of the Hessian matrix

2.1 All subjects tested at same time, but with different combinations of tests

Let \( \bm{\mu} = (\theta, f) \), where \( \hat{\bm{\mu}} \) denotes the MLE of \((\theta, f)\). Let the variance of \( \hat{\bm{\mu}} \) be denoted by:

\[
\text{Var}(\hat{\bm{\mu}}) = -E \left[ \begin{array}{cc} h_{11} & h_{12} \\ h_{21} & h_{22} \end{array} \right]^{-1}.
\]

When \( f(u) = f \) for \( u \geq 0 \), the \( h_{ij} \)s can be expressed as follows:

\[
h_{11} = \frac{n_{-0} + n_{00}}{(1 - \theta + f k_1)^2} - \frac{n_{00} + n_{00} + n_{00}}{(1 - \theta)^2} - \frac{n_{-0}}{(\theta - f k_1)^2} - \frac{n_{-0} + n_{-00}}{(1 - \theta + f k_1 + f k_2)^2} - \frac{n_{1-}}{\theta^2} - \frac{n_{-1} + n_{-1} + n_{111} + n_{111}}{(\theta - f k_1 - f k_2)^2}
\]

\[
h_{22} = \frac{(n_{-1-} + n_{-0}) k_1^2}{(1 - \theta + f k_1)^2} - \frac{n_{1-0} (k_1 + k_2)^2}{(f k_1 + f k_2)^2} - \frac{n_{00}}{(1 - \theta + f k_1 + f k_2)^2} - \frac{n_{-0} (k_1 + k_2)^2}{(\theta - f k_1 - f k_2)^2} - \frac{(n_{-10} + n_{11})}{f^2} - \frac{n_{1-0} + n_{100}}{f^2} - \frac{(n_{-1} + n_{-11} + n_{1-1} + n_{111})(-k_1 - k_2)^2}{f^2} - \frac{n_{1-0} (k_1 + k_2)^2}{(\theta - f k_1 - f k_2)^2} - \frac{(n_{-1-} + n_{-0}) k_1^2}{(1 - \theta + f k_1)^2} - \frac{n_{00}}{(1 - \theta + f k_1 + f k_2)^2} - \frac{n_{-0} (k_1 + k_2)^2}{(\theta - f k_1 - f k_2)^2} - \frac{(n_{-1} + n_{-11} + n_{1-1} + n_{111})(-k_1 - k_2)^2}{f^2} - \frac{n_{1-0} (k_1 + k_2)^2}{(\theta - f k_1 - f k_2)^2}
\]

\[
h_{12} = \frac{(n_{-0} + n_{00}) k_1}{(1 - \theta + f k_1)^2} + \frac{(n_{11} + n_{-1}) k_1}{(\theta - f k_1)^2} + \frac{n_{-0} (k_1 + k_2)}{(1 - \theta + f k_1 + f k_2)^2} \frac{(n_{-1} + n_{-11} + n_{1-1} + n_{111})(-k_1 - k_2)}{(\theta - f k_1 - f k_2)^2} - \frac{n_{1-0} (k_1 + k_2)^2}{(\theta - f k_1 - f k_2)^2} - \frac{(n_{-1} + n_{-11} + n_{1-1} + n_{111})(-k_1 - k_2)^2}{f^2} - \frac{n_{1-0} (k_1 + k_2)^2}{(\theta - f k_1 - f k_2)^2}
\]
2.2 *All subjects given the same battery of tests at the same time*

Let $\boldsymbol{\mu} = (\theta, f)$, where $\hat{\boldsymbol{\mu}}$ denotes the MLE of $(\theta, f)$. Let the variance of $\hat{\boldsymbol{\mu}}$ be denoted by:

$$Var(\hat{\boldsymbol{\mu}}) = -E \left[ \begin{array}{cc} h_{11} & h_{12} \\ h_{21} & h_{22} \end{array} \right]^{-1}.$$

When $f(u)$ is assumed to be a constant density, that is $f(u) = f$ for $u \geq 0$ and when all subjects are tested at the same time using a battery of tests consisting of the detuned ELISA assay, standard ELISA and antigen test, the $h_{ij}$s can be expressed as follows:

$$h_{11} = -\frac{n_{000}}{(1-\theta)^2} - \frac{n_{111}}{(\theta - f k_1 - f k_2)^2}$$  \hspace{1cm} (4)

$$h_{22} = -\frac{n_{100}}{f^2} - \frac{n_{110}}{f^2} - \frac{n_{111}(-k_1 - k_2)^2}{(\theta - f k_1 - f k_2)^2}$$  \hspace{1cm} (5)

$$h_{21} = -\frac{n_{111}(-k_1 - k_2)}{(\theta - f k_1 - f k_2)^2}$$  \hspace{1cm} (6)

If all subjects are tested by a battery of two tests consisting of the detuned ELISA and standard ELISA assays, the $h_{ij}$s are equal to:

$$h_{11} = -\frac{n_{00}}{(1-\theta + f k_1)^2} - \frac{n_{11}}{(\theta - f k_1 - f k_2)^2}$$  \hspace{1cm} (7)

$$h_{22} = -\frac{n_{00}k_1^2}{(1-\theta + f k_1)^2} - \frac{n_{10}}{f^2} - \frac{n_{11}(-k_2 - k_1)^2}{(\theta - f k_1 - f k_2)^2}$$  \hspace{1cm} (8)

$$h_{21} = -\frac{n_{00}k_1}{(1-\theta + f k_1)^2} - \frac{n_{11}(-k_1 - k_2)}{(\theta - f k_1 - f k_2)^2}$$  \hspace{1cm} (9)

If all subjects are tested by a battery of two tests consisting of the antigen and standard
ELISA assays, the $h_{ij}$s are equal to:

$$h_{11} = -\frac{n_{00}}{(1-\theta)^2} - \frac{n_{11}}{(\theta - fk_1)^2}$$  \hspace{1cm} (10)

$$h_{22} = -\frac{n_{10}}{f^2} - \frac{n_{11}k_1^2}{(\theta - fk_1)^2}$$  \hspace{1cm} (11)

$$h_{21} = \frac{n_{11}k_1}{(\theta - fk_1)^2}$$  \hspace{1cm} (12)

If all subjects are tested by a battery of two tests consisting of the antigen and detuned ELISA assays, the $h_{ij}$s are equal to:

$$h_{11} = -\frac{n_{00}}{(1-\theta)^2} - \frac{n_{11}}{(\theta - fk_1-fk_2)^2}$$  \hspace{1cm} (13)

$$h_{22} = -\frac{n_{10}(k_1+k_2)^2}{(fk_1+fk_2)^2} - \frac{n_{11}(-k_1-k_2)^2}{(\theta - fk_1-fk_2)^2}$$  \hspace{1cm} (14)

$$h_{21} = -\frac{n_{11}(-k_1-k_2)}{(\theta - fk_1-fk_2)^2}$$  \hspace{1cm} (15)

3. Web Appendix C

Expressions for $\pi$ and likelihood functions under parametric assumptions

In this section, we present expressions for $\pi$ and the simplified likelihood functions based on the assumption that the density for someone born at $t_0$ is piece-wise constant as follows:

$f(u \mid t_0) = f_1$ for $u \in (t - L_1^*, t]$ and $f(u \mid t_0) = f_2$ for $u \in (t - L_1^* - L_2^*, t - L_1^*]$.

3.1 Expressions for $\pi$

The expressions for $\pi$ under the piece-wise constant density assumption are:
\[ \pi_0(t \mid t_0) = 1 - F(t \mid t_0) = 1 - \theta \quad (16) \]

\[ \pi_1(t \mid t_0) = f_1 \int_{\max(0,t-L_1^*)}^{t} [1 - G_1(t-u)] du \quad (17) \]

\[ \pi_2(t \mid t_0) = f_2 \int_{\max(0,t-L_1^*-L_2^*)}^{t-L_1^*} \int_{\max(0,t-u-L_2^*)}^{L_1^*} g_1(v)[1 - G_2(t-u-v \mid v)] dv du \quad (18) \]

\[ = f_2 k_2^* + f_1 k_3^* \]

\[ \pi_3(t \mid t_0) = \theta - f_1 k_1 - f_1 k_3^* - f_2 k_2^* \quad (19) \]

3.2 Likelihood function: Subjects tested with different batteries of tests

The resulting likelihood can be expressed as a product of terms that are linear combinations of \( f_1, f_2, \) and \( \theta \) as follows:

\[ L = [1 - \theta + f_1 k_1]^{n_{-0+}+n_{-00}} \times [1 - \theta]^{n_{-0+}+n_{00}+n_{000}+n_{000}} \times [f_1 k_1 + f_2 k_2^* + f_1 k_3^*]^{n_{1-0}} \times [\theta - f_1 k_1]^{n_{11-}+n_{-1-}} \times [1 - \theta + f_1 k_1 + f_2 k_2^* + f_1 k_3^*]^{n_{-0-}} \times [f_2 k_2^* + f_1 k_3^*]^{n_{10+}+n_{110}} \times [f_1 k_1]^{n_{10-}+n_{100}} \times [\theta]^{n_{1--}} \times [\theta - f_1 k_1 - f_1 k_3^* - f_2 k_2^*]^{n_{-1-1}+n_{-1-1}+n_{1-1}+n_{111}} \quad (20) \]

and where \( k_1, k_2^*, k_3^* \) are functions of \( t, G_1(\cdot), G_2(\cdot) \) (see equations (17) - (18)). For given
$k_1, k_2^*, k_3^*$, the maximum likelihood estimator of $\theta = F(t \mid t_0)$, $f_1$ and $f_2$ can be obtained from (20) by joint maximization of the likelihood using numerical techniques. Approximate confidence intervals and standard errors for MLEs of $\theta, f_1, f_2$, or functions of these components, can be obtained from the Hessian of the log-likelihood function (Cox and Hinkley, 1974). Closed form solutions for the elements of the Hessian matrix can be obtained using standard methods (details available upon request). Under the assumption that $f_1 = f_2 = f$ and since $k_2^* + k_3^* = k_2$, the likelihood in (20) reduces to the likelihood obtained under the constant density assumption (i.e. equation (4) in Section 3.1 of the paper).

3.3 Likelihood function: All subjects tested with the same battery of tests

The expressions for the likelihood when the battery consists of all 3 tests (AED), the antigen and standard ELISA (AE), the antigen and detuned ELISA (AD), and the standard and detuned ELISA (ED) are, respectively:

$$L_{AED} = p_{000}^{n_{000}} p_{100}^{n_{100}} p_{110}^{n_{110}} p_{111}^{n_{111}}$$

$$= [1 - \theta]^{n_{000}}$$

$$[f_1 k_1]^{n_{100}} [f_2 k_2^* + f_1 k_3^*]^{n_{110}}$$

$$[\theta - f_1 k_1 - f_2 k_2^* - f_1 k_3^*]^{n_{111}}$$

(21)

$$L_{AE} = p_{00-}^{n_{00-}} p_{10-}^{n_{10-}} p_{11-}^{n_{11-}}$$

$$= [1 - \theta]^{n_{00-}}$$

$$[f_1 k_1]^{n_{10-}}$$

$$[\theta - f_1 k_1]^{n_{11-}}$$

(22)
In each of the above equations (21) - (24), \( k_1, k^*_2, k^*_3 \) are integral functions of \( G_1(\cdot), G_2(\cdot) \) given in equations (17)-(18). For given \( k_1, k^*_2, k^*_3 \), the maximum likelihood estimators of \( \theta = F(t \mid t_0), f_1 \) and \( f_2 \) can be obtained by joint maximization of the likelihood using numerical techniques. Approximate confidence intervals and standard errors for MLEs of \( \theta, f_1, f_2 \), or functions of these components, can be obtained from the Hessian of the log-likelihood function (Cox and Hinkley, 1974). Analytical expressions for the Hessian matrix can be obtained using standard methods (details available upon request). Under the assumption that \( f_1 = f_2 = f \) and since \( k^*_2 + k^*_3 = k_2 \), the above likelihood expressions ((21) - (24)) reduce to the expressions obtained under the constant density assumption (i.e. equations (5) - (8) in Section 3.2 of the paper).

4. Web Appendix D

Allowance for imperfect diagnostic tests

The methods presented in the paper are derived under the assumption that all diagnostic tests have perfect sensitivity and specificity. However, in practice, both antigen
and ELISA tests may be prone to error. The likelihoods developed in Section 3 can be generalized to incorporate imperfect sensitivity and/or specificity of any or all of the tests in a given battery.

To illustrate, consider the setting where the (E,D) test battery in given to all subjects at time $t$ and that $f(u) = f$ for $u \in [t - L_1^* - L_2^*, t]$. Let the sensitivity of the standard ELISA test (E) be denoted by $p_E$ and that of the detuned assay (D) be denoted by $p_D$. That is,

$$\Pr(E+ | S_2, S_3) = p_E$$
$$\Pr(D+ | S_3) = p_D$$

Suppose that both the tests have perfect specificities; that is,

$$\Pr(E- | S_0, S_1) = 1$$
$$\Pr(D- | S_0, S_1, S_2) = 1$$

Under the assumption that the test results are independent conditional on an individual’s disease state, the log likelihood can be expressed as

$$L = n_{-00} \log[(1 - \theta + f k_1) + f k_2 (1 - p_E) + (\theta - f k_1 - f k_2)(1 - p_E)(1 - p_D)]$$
$$+ n_{-10} \log[f k_2 p_E + (\theta - f k_1 - f k_2)p_E(1 - p_D)]$$
$$+ n_{-11} \log[(\theta - f k_1 - f k_2)p_E p_D]$$
$$+ n_{-01} \log[(\theta - f k_1 - f k_2)(1 - p_E)p_D]$$
The maximum likelihood estimates based on the above likelihood are:

\[
\hat{f} = \frac{(n_{-01} + n_{-11})p_Ep_D - (n_{-01} + n_{-11})p_E + n_{-10}p_D}{p_Ep_Dk_2n} \\
\hat{\theta} = \frac{k_1p_Ep_D(n_{-01} + n_{-11}) - k_1p_E(n_{-01} + n_{-11}) + k_2p_Ep_D(n_{-01} + n_{-11}) + n_{-10}p_D(k_1 + k_2)}{p_Ep_Dk_2n} \\
\hat{\lambda} = \frac{(n_{-01} + n_{-11})(p_D - 1)p_E + n_{-10}p_D}{k_2p_Ep_D(n_{-00} + n_{-10}) + k_1p_E(n_{-01} + n_{-11})(1 - p_D) - n_{-10}p_D(k_1 + k_2)}
\]

, where \( n \) denotes the total number of subjects screened. When the tests are assumed to be perfect (i.e. \( p_E = p_D = 1 \) and \( n_{-01} = 0 \)), the above expressions for the MLEs of \( f, \theta \) and \( \lambda \) reduce to the corresponding expressions presented for the \( (E,D) \) test battery in Table 2 of the paper.

Suppose that the standard ELISA test is perfect but that the detuned ELISA may have both imperfect sensitivity and specificity. That is,

\[
\Pr(E^- | S_0, S_1) = 1 \\
\Pr(E^+ | S_2, S_3) = 1
\]

and

\[
\Pr(D^- | S_0, S_1, S_2) = q_D \\
\Pr(D^+ | S_3) = p_D
\]

Under the assumption that the test results are independent conditional on an individ-
ual’s disease state, the log likelihood can be expressed as

\[
L = \log [(1 - \theta + f k_1)q_D] + n_{-11} \log [fk_2(1 - q_D) + (\theta - f k_1 - f k_2)p_D] + n_{-01} \log [(1 - \theta + f k_1)(1 - q_D)]
\]

The maximum likelihood estimates based on the above likelihood are:

\[
\hat{f} = \frac{p_D(n_{-10} + n_{-11}) - n_{-11}}{k_2 n(q_D + p_D - 1)}
\]

\[
\hat{\theta} = \frac{(n_{-10} + n_{-11})[p_D(k_1 + k_2) - k_2(1 - q_D)] - k_1 n_{-11}}{k_2 n(q_D + p_D - 1)}
\]

\[
\hat{\lambda} = \frac{p_D(n_{-10} + n_{-11}) - n_{-11}}{k_2 n(q_D + p_D - 1) - (n_{-10} + n_{-11})[p_D(k_1 + k_2) - k_2(1 - q_D)] + k_1 n_{-11}}
\]

In addition, if we assume that the detuned ELISA has perfect specificity \(q_D = 1\) but possibly imperfect sensitivity \(p_D \leq 1\), then the maximum likelihood estimates are:

\[
\hat{f} = \frac{p_D(n_{-10} + n_{-11}) - n_{-11}}{k_2 p_D n}
\]

\[
\hat{\theta} = \frac{(n_{-10} + n_{-11})(k_1 p_D + k_2 p_D) - k_1 n_{-11}}{k_2 p_D n}
\]

\[
\hat{\lambda} = \frac{p_D(n_{-10} + n_{-11}) - n_{-11}}{k_2 p_D n - p_D(n_{-10} + n_{-11})(k_1 + k_2) + k_1 n_{-11}}
\]

In the special case where both the standard and detuned ELISA tests are assumed to be perfect (i.e. \(q_D = p_D = 1\)), the above expressions for the MLEs of \(f, \theta\) and \(\lambda\) reduce to the corresponding expressions presented for the (E,D) test battery in Table 2 of the paper.
Other approaches for handling misclassification errors have been suggested by Parekh et al. (2002), McDougal et al. (2006), and Hargrove et al. (2008).

5. Web Appendix E

Expanded model incorporating risk of death

We consider the impact of the risk of death on the prevalence formula developed in Section 3 by expanding the 4-state model considered in Section 2 to a competing risks framework, in which (a) the subject is at risk of death at any time point (with hazard function denoted $\lambda_d(t)$), (b) following the entrance into State 3 (positive detuned ELISA), a subject is at risk of entering a 'symptomatic' disease state (denoted State 4), (c) the risk of death is not elevated by entrance into State 1 (HIV infection), State 2 (ELISA positivity), or State 3 (detuned ELISA positivity), and (d) upon entrance into State 4 (symptomatic disease), the hazard for death shifts from $\lambda_d(t)$ to $\lambda^*_d(t)$. We then assume that sampling is from an asymptomatic (apparently healthy) population and determine the state prevalence probabilities and investigate the properties of the incidence rate estimators developed in Section 3 under this expanded model.

To account for the competing risk of death, the hazard functions describing entrance into States 1, 2, and 3 in the original framework are now considered as cause-specific hazard functions. For simplicity in notation, we suppress the dependence of these functions on $t_0$, the time of birth, and take $t_0 = 0$, but otherwise use the same notation as in Section 2. In particular, we use $\pi_j(t)$ as before to denote the algebraic expressions developed in Section 2 and now use $\pi^*_j(t)$ to denote the state prevalence functions under the expanded model; that is, $\pi^*_j(t) = P(\text{in State } j \text{ at time } t)$. Then
\[ \pi_0^\ast(t) = e^{-\int_0^t \lambda(u) + \lambda_d(u) du} = e^{-\Lambda(t)} e^{-\Lambda_d(t)} = \pi_0(t) e^{-\Lambda_d(t)}, \]

where \( \Lambda(t) = \int_0^t \lambda(u) du \), \( \Lambda_d(t) = \int_0^t \lambda_d(u) du \), and where we have used the fact that \( \pi_0(t) = 1 - F(t) = e^{-\Lambda(t)} \). Similarly,

\[ \pi_1^\ast(t) = \int_{\text{max}(0, t-L_1^\ast)}^t \lambda(u)e^{-[\Lambda(u)+\Lambda_d(u)]}[1 - G_1(t-u)]e^{-\int_u^t \lambda_d(x) dx} du \]

\[ = \int_{\text{max}(0, t-L_1^\ast)}^t f(u)[1 - G_1(t-u)]du e^{-\Lambda_d(t)} = \pi_1(t) e^{-\Lambda_d(t)}, \]

where we have used the fact that \( f(u) = \lambda(u)e^{-\int_0^u \lambda(x) dx} \), and

\[ \pi_2^\ast(t) = \int_{\text{max}(0, t-L_1^\ast)}^t \int_{\text{max}(0, t-L_2^\ast)}^{\text{min}(t, u+L_1^\ast)} \{ \lambda(u)e^{-[\Lambda(u)+\Lambda_d(u)]} \} \{ g_1(v-u) e^{-\int_u^v \lambda_d(x) dx} \} \]

\[ \cdot \{ [1 - G_2(t - v | L_1 = v - u)] e^{-\int_v^t \lambda_d(x) dx} \} dv \]

\[ = \pi_2(t) e^{-\Lambda_d(t)}. \]

Finally,

\[ \pi_3^\ast(t) = \int_0^t \int_u^{\text{min}(t, u+L_1^\ast)} \int_v^{\text{min}(t, v+L_2^\ast)} \{ \lambda(u)e^{-[\Lambda(u)+\Lambda_d(u)]} \} \{ g_1(v-u) e^{-\int_u^v \lambda_d(x) dx} \} \]

\[ \cdot \{ g_2(w-v | L_1 = v-u) e^{-\int_v^w \lambda_d(x) dx} \} \]

\[ \cdot \{ e^{-\int_w^t \lambda_d(x) dx} [1 - G_s(t - w)] \} dw \]

\[ = e^{-\Lambda_d(t)} \int_0^t \int_u^{\text{min}(t, u+L_1^\ast)} \int_v^{\text{min}(t, v+L_2^\ast)} \{ \lambda(u)e^{-\Lambda(u)} g_1(v-u) \} \]

\[ \cdot g_2(w-v | L_1 = v-u) e^{-\int_v^w \lambda_d(x) dx} \cdot e^{-\int_w^t \lambda_d(x) dx} [1 - G_s(t - w)] \]
\[ g_2(w - v \mid L_1 = v - u)[1 - G_s(t - w)] \] \[ dw \, dv \, du, \]

\[ = e^{-\Lambda(t)} \int_0^t \int_u^{\min(t,u+L_1^*)} \int_v^{\min(t,v+L_2^*)} \{ \lambda(u)e^{-\Lambda(u)}g_1(v - u) \]

\[ \cdot g_2(w - v \mid L_1 = v - u)[1 - G_s(t - w)] \} \, dw \, dv \, du, \]

where \( G_s(x) \) is the c.d.f. for the time between entering State 3 and entering State 4, and \( L_1^*, L_2^* \) denote the supports of the distributions of \( L_1, L_2 \), respectively. Note that \( L_2^* \) is typically small relative to the time from infection until testing. Thus, we can write the above as

\[ \pi_3^*(t) = e^{-\Lambda(t)}[1 - G_s(t^*)] \int_0^t \int_u^{\min(t,u+L_1^*)} \int_v^{\min(t,v+L_2^*)} \{ \lambda(u)e^{-\Lambda(u)}g_1(v - u)g_2(w - v \mid L_1 = v - u) \} \, dw \, dv \, du, \]

\[ = e^{-\Lambda(t)}[1 - G_s(t^*)]\pi_3(t), \]

where \( t^* \in (t - L_2^*, t). \)

Now suppose we sample from an asymptomatic population (persons in States 0, 1, 2, or 3). The probability of finding a subject in state \( j \) equals

\[ P(\text{in State } j \text{ at time } t \mid \text{in State } S_0, S_1, S_2 \text{ or } S_3) = \frac{\pi_j^*(t)}{\sum_{i=0}^3 \pi_i^*(t)} \]

\[ = \frac{\pi_j(t) \, e^{-\Lambda_d(t)}}{\pi_0(t) \, e^{-\Lambda_d(t)} + \pi_1(t) \, e^{-\Lambda_d(t)} + \pi_2(t) \, e^{-\Lambda_d(t)} + \pi_3(t) \, e^{-\Lambda_d(t)}[1 - G_s(t^*)]} \]
\[
\pi_j(t) = \frac{\pi_j(t)}{\pi_0(t) + \pi_1(t) + \pi_2(t) + \pi_3(t)[1 - G_s(t^*)]},
\]

for \(j = 0, 1, 2\) and

\[
P(\text{in State 3 at time } t \mid \text{ in State } S_0, S_1, S_2, \text{ or } S_3) \quad (S1)
\]

\[
= \frac{\pi_3(t) e^{-\Lambda_d(t)}[1 - G_s(t^*)]}{\pi_0(t) e^{-\Lambda_d(t)} + \pi_1(t) e^{-\Lambda_d(t)} + \pi_2(t) e^{-\Lambda_d(t)} + \pi_3(t) e^{-\Lambda_d(t)}[1 - G_s(t^*)]}
\]

\[
= \frac{\pi_3(t) [1 - G_s(t^*)]}{\pi_0(t) + \pi_1(t) + \pi_2(t) + \pi_3(t)[1 - G_s(t^*)]} \quad (S2)
\]

Now let \(\pi_j^{**}(t)\) denote the relative frequencies of State \(j\) from among asymptomatic subjects, for \(j = 0, 1, 2, 3\); that is,

\[
\pi_j^{**}(t) = \frac{\pi_j(t)}{\sum_{i=0}^{3} \pi_i(t)}.
\]

From (S1) and (S2), it follows that

\[
\pi_j^{**}(t) = \frac{\pi_j(t)}{D} \quad \text{for } j = 0, 1, 2,
\]

\[
\pi_3^{**}(t) = \frac{\pi_3(t)[1 - G_s(t^*)]}{D},
\]

where \(D = \pi_0(t) + \pi_1(t) + \pi_2(t) + \pi_3(t)[1 - G_s(t^*)]\).
Note that \([1 - G_s(t^*)]\) represents the prevalence, at time \(t^*\), of asymptomatic persons from among all those born at time 0; that is, including both uninfected persons and HIV-infected individuals that have not yet developed symptoms by \(t^*\). Since \(\sum_{i=0}^{3} \pi_j(t) = 1\) and \(0 < 1 - G_s(t^*) \leq 1\), we see that in the expanded model, the true state prevalence probabilities \(\pi_j^*(t)\) are smaller than the quantities \(\pi_j(t)\) for \(j = 0, 1, 2, 3\), and in this sense the latter are distorted by the introduction of death into the model. Similarly, it is easily shown that \(\pi_j(t)\) is smaller than \(\pi_j^{**}(t)\) for \(j = 0, 1, 2\) and larger than \(\pi_j^{**}(t)\) for \(j = 3\).

To illustrate, suppose that by age \(t\) year, 5% of individuals have died, 70% are alive and HIV-uninfected, 15% are HIV-infected and asymptomatic, and 10% are HIV-infected and symptomatic, which would be typical in an endemic area. Then \(1 - G_s(t) = .85\). If the incidence rate is 2\% (.02), \(E(L_1) = 3/52\) (3 weeks) and \(E(L_2) = 26/52\) (26 weeks), it can be shown that \(\pi_1(t) = .0008\) and \(\pi_2(t) = .007\), so that \(\pi_j^*(t) = 0.95\pi_j(t)\) for \(j = 0, 1, 2\) and \(\pi_3^*(t) = .81\pi_3(t)\). Thus, under the expanded model, the original prevalence expressions \(\pi_j(t)\) understate \((j = 0, 1, 2, 3)\) the true state prevalence functions. The relative frequencies of States \(j\) from among the asymptomatic subjects are \(\pi_j^{**}(t) = 1.05\pi_j(t)\) for \(j = 0, 1, 2\) and \(\pi_3^{**}(t) = 0.89\pi_3(t)\).

Now suppose testing is done with the (E,D) battery and consider a random sample of asymptomatic persons. Then, analogous to the arguments leading to equation (8) in Section 3.2 of the paper, and dropping the dependency on \(t\) for notational simplicity, the likelihood function of the data is

\[
L = \left(\frac{\pi_0 + \pi_1}{D}\right)^{n-00}\left(\frac{\pi_2}{D}\right)^{n-10}\left(\frac{\pi_3[1 - G_s]}{D}\right)^{n-11} \propto \frac{(\pi_0 + \pi_1)^{n-00}\pi_2^{n-10}\pi_3^{n-11}}{D^n}
\]

Then using the fact that \(\pi_0 = 1 - \theta\), \(\pi_1 = f k_1\), and \(\pi_2 = f k_2\) (Section 2), we can write
\[ L \propto \frac{(1 - \theta + f k_1)^{n-00} (f k_2)^{n-10} \pi_3^{n-11}}{[1 - \theta + f k_1 + f k_2 + (1 - G_s) \pi_3]^n}. \]

Taking logs and partial derivatives gives

\[
\frac{\partial \ln L}{\partial \theta} = \frac{-n_{-00}}{1 - \theta + f k_1} + \frac{n}{1 - \theta + f(k_1 + k_2) + (1 - G_s) \pi_3}
\]

and

\[
\frac{\partial \ln L}{\partial f} = \frac{n_{-00} k_1}{1 - \theta + f k_1} + \frac{n_{-10}}{f} - \frac{n(k_1 + k_2)}{1 - \theta + f(k_1 + k_2) + (1 - G_s) \pi_3}.
\]

Setting both partials to zero yields

\[
\frac{n(k_1 + k_2)}{1 - \theta + f(k_1 + k_2) + (1 - G_s) \pi_3} = \frac{n_{-00}(k_1 + k_2)}{1 - \theta + f k_1}
\]

and

\[
\frac{n(k_1 + k_2)}{1 - \theta + f(k_1 + k_2) + (1 - G_s) \pi_3} = \frac{n_{-00} k_1}{1 - \theta + f k_1} + \frac{n_{-10}}{f}.
\]

Equating the right-hand sides of the equations and solving for \( f \) gives

\[ f = \frac{n_{-10}(1 - \theta)}{n_{-00} k_2 - n_{-10} k_1}, \]

and hence the MLE of the incidence rate \( \lambda = f/(1 - \theta) \) is

\[ \tilde{\lambda} = \frac{n_{-10}}{n_{-00} k_2 - n_{-10} k_1}, \]

which is identical to the MLE of \( \lambda \) under the original model. Thus, the incidence estimator developed in Section 3 remains valid under an expanded model in which sampling is from an apparently healthy (asymptomatic) population and the risk of death from HIV infection
is not elevated until persons become symptomatic. Similar arguments apply to the esti-
mators of $\lambda$ under different batteries of tests and to the Janssen et al estimator.

**References**


Hargrove, J., Humphrey, J., Mutasa, K., Parekh, B., McDougal, J., Ntozini, R., Chi-


Parekh, B., Kennedy, M., Dobbs, T., Pau, C., Byers, R., Green, T., Hu, D., Vanichseni, S., Young, N., Choopanya, K., Mast o, K., Mastro, T. and McDougal, J. (2002). Quan-