Supplementary Material for "Borrowing Information across Populations in Estimating Positive and Negative Predictive Values"

Ying Huang[†*], Youyi Fong[*], John Wei[$], and Ziding Feng[*]

[*] *Fred Hutchinson Cancer Research Center, Vaccine & Infectious Disease / Public Health Sciences, 1100 Fairview Avenue N., Seattle, WA 98109, USA.*

[$] *University of Michigan, Department of Urology, 2301 Commonwealth Blvd, Ann Arbor, MI 48105.*

[*] *Fred Hutchinson Cancer Research Center, Public Health Sciences, 1100 Fairview Avenue N., Seattle, WA 98109, USA.*

[†] *Corresponding author: yhuang124@gmail.com*

A: Proofs of Asymptotic Normality of the Weighted PPV Estimators

A1: Proof of Theorem 1

To find the asymptotic variance of $\widehat{\mathrm{PPV}}_w(y)$, we apply delta method,

$$
\mathrm{var}\left[\sqrt{n_{\bar{D}}}\left\{\widehat{\mathrm{PPV}}_w(y) - \mathrm{PPV}(y)\right\}\right]
$$

$$
\simeq \left(\begin{array}{c} \frac{\rho(1-\rho)S_{\bar{D}}(y)}{P(Y>y)^2} \\ \frac{-\rho(1-\rho)\mathrm{ROC}\left[S_{\bar{D}}(y)\right]}{P(Y>y)^2} \end{array}\right)^T \mathrm{var}\left(\begin{array}{c} \sqrt{n_{\bar{D}}}\left[\widehat{\mathrm{ROC}}_w\left\{\widetilde{S}_{\bar{D}}(y)\right\} - \mathrm{ROC}\{S_{\bar{D}}(y)\}\right] \\ \sqrt{n_{\bar{D}}}\left\{\widetilde{S}_{\bar{D}}(y) - S_{\bar{D}}(y)\right\} \end{array}\right) \left(\begin{array}{c} \frac{\rho(1-\rho)S_{\bar{D}}(y)}{P(Y>y)^2} \\ \frac{-\rho(1-\rho)\mathrm{ROC}\left[S_{\bar{D}}(y)\right]}{P(Y>y)^2} \end{array}\right)
$$

$$
= n_{\bar{D}}\left[\frac{-\rho(1-\rho)\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}}{P(Y>y)^2}\right]^2 \mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\} + n_{\bar{D}}\left[\frac{\rho(1-\rho)S_{\bar{D}}(y)}{P(Y>y)^2}\right]^2 \mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}\right]
$$

$$
+ 2n_{\bar{D}}\left\{\frac{\rho(1-\rho)S_{\bar{D}}(y)}{P(Y>y)^2}\right\}\left[\frac{-\rho(1-\rho)\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}}{P(Y>y)^2}\right]\mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}, \widehat{S}_{\bar{D}}(y)\right],
$$

where $P(Y>y) = \rho S_D(y) + (1-\rho)S_{\bar{D}}(y)$. We now need to find $\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\}$, $\mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}, \widehat{S}_{\bar{D}}(y)\right]$, and $\mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}\right]$.

First, $\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\} = \frac{1}{n_{\bar{D}}}S_{\bar{D}}(y)\{1 - S_{\bar{D}}(y)\}$.

Second, to find $\mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}, \widehat{S}_{\bar{D}}(y)\right]$, we note that

$$
\mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}, \widehat{S}_{\bar{D}}(y)\right] = \mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\} - \mathrm{ROC}\{S_{\bar{D}}(y)\}, \widetilde{S}_{\bar{D}}(y) - S_{\bar{D}}(y)\right]
$$

$$
= \mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\} - \mathrm{ROC}\{S_{\bar{D}}(y)\}, \widetilde{S}_{\bar{D}}(y) - S_{\bar{D}}(y)\right]
$$

$$
+ \mathrm{cov}\left[\mathrm{ROC}\{\widetilde{S}_{\bar{D}}(y)\} - \mathrm{ROC}\{S_{\bar{D}}(y)\}, \widetilde{S}_{\bar{D}}(y) - S_{\bar{D}}(y)\right]
$$

$$
\simeq^\star \mathrm{cov}\left[\widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}, \widetilde{S}_{\bar{D}}(y)\right] + \mathrm{cov}\left[\mathrm{ROC}\{\widetilde{S}_{\bar{D}}(y)\}, \widetilde{S}_{\bar{D}}(y)\right]
$$

$$
\simeq^\dagger -w\frac{f_D(y)}{f_{\bar{D}}(y)}\mathrm{var}\left\{\widehat{S}_{\bar{D}}(y)\right\} + \mathrm{ROC}'\{S_{\bar{D}}(y)\}\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\}
$$

$$
= -w\frac{f_D(y)}{f_{\bar{D}}(y)}\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\} + \frac{f_D(y)}{f_{\bar{D}}(y)}\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\} = (1-w)\frac{f_D(y)}{f_{\bar{D}}(y)}\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\},
$$

where $(^\star)$ holds by equicontinuity of $\sqrt{n_{\bar{D}}}\left(\widehat{\mathrm{ROC}}_w - \mathrm{ROC}\right)$ (van der Vaart and Wellner , 1996), and $(^\dagger)$ holds since

$$\mathrm{cov}\left[\widetilde{S}_{\bar{D}}(y), \widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right] \simeq \mathrm{cov}\left(\widetilde{S}_{\bar{D}}(y), w\frac{1}{n_D}\sum_{i=1}^{n_D} I\left[Y_{Di} > \widetilde{S}_{\bar{D}}^{-1}\{S_{\bar{D}}(y)\}\right]\right)$$

$$\simeq \quad w\mathrm{cov}\left[\widetilde{S}_{\bar{D}}(y), S_D\widetilde{S}_{\bar{D}}^{-1}\{S_{\bar{D}}(y)\}\right] \simeq -w\frac{f_D(y)}{f_{\bar{D}}(y)}\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\}.$$

Lastly

$$\mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{\widetilde{S}_{\bar{D}}(y)\}\right]$$

$$= \quad \mathrm{var}\left[\mathrm{ROC}\{\widetilde{S}_{\bar{D}}(y)\}\right] + \mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right] + 2\mathrm{cov}\left[\mathrm{ROC}\{\widetilde{S}_{\bar{D}}(y)\}, \widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right]$$

$$\simeq \quad \mathrm{ROC}'\{S_{\bar{D}}(y)\}^2\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\} + \mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right] + 2\mathrm{ROC}'\{S_{\bar{D}}(y)\}\mathrm{cov}\left[\widetilde{S}_{\bar{D}}(y), \widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right]$$

$$\simeq \quad (1 - 2w)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2\mathrm{var}\left\{\widetilde{S}_{\bar{D}}(y)\right\} + \mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right],$$

where

$$\mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right] = \mathrm{var}\left(\frac{w}{n_D}\sum_{i=1}^{n_D} I\left[Y_{Di} > \widetilde{S}_{\bar{D}}^{-1}\{S_{\bar{D}}(y)\}\right] + \frac{1-w}{n_D^*}\sum_{i=1}^{n_D^*} I\left[Y_{Di}^* > \widetilde{S}_{\bar{D}}^{*-1}\{S_{\bar{D}}(y)\}\right]\right)$$

$$\simeq \quad \frac{w^2}{n_D}\mathrm{ROC}\{S_{\bar{D}}(y)\}\left[1 - \mathrm{ROC}\{S_{\bar{D}}(y)\}\right] + \frac{w^2}{n_{\bar{D}}}\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 S_{\bar{D}}(y)\{1 - S_{\bar{D}}(y)\}$$

$$+ \quad \frac{(1-w)^2}{n_D^*}\mathrm{ROC}\{S_{\bar{D}}(y)\}\left[1 - \mathrm{ROC}\{S_{\bar{D}}(y)\}\right] + \frac{(1-w)^2}{n_{\bar{D}}^*}\left(\frac{f_D^*\left[S_{\bar{D}}^{*-1}\{S_{\bar{D}}(y)\}\right]}{f_{\bar{D}}^*\left[S_{\bar{D}}^{*-1}\{S_{\bar{D}}(y)\}\right]}\right)^2 S_{\bar{D}}(y)\{1 - S_{\bar{D}}(y)\}$$

$$= \quad \left[\frac{w^2}{n_D} + \frac{(1-w)^2}{n_D^*}\right]S_D(y)\{1 - S_D(y)\} + \left[\frac{w^2}{n_{\bar{D}}} + \frac{(1-w)^2}{n_{\bar{D}}^*}\right]\left(\frac{f_D(y)}{f_{\bar{D}}(y)}\right)^2 S_{\bar{D}}(y)\{1 - S_{\bar{D}}(y)\}.$$

Thus

$$\mathrm{var}\left[\widehat{\mathrm{ROC}}_w\{S_{\bar{D}}(y)\}\right] = \left[(1-w)^2\left(1 + \frac{n_{\bar{D}}}{n_{\bar{D}}^*}\right)\right]\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2\frac{1}{n_{\bar{D}}}S_{\bar{D}}(y)\{1 - S_{\bar{D}}(y)\} + \left[w^2 + (1-w)^2\frac{n_D}{n_D^*}\right]\frac{1}{n_D}S_D(y)\{1 - S_D(y)\}.$$

A2: Proof of Theorem 2

Observe that Σ_w is a quadratic function of w,

$$\Sigma_w \quad = \quad w^2\left[A_{22}\left(1 + \frac{1}{\lambda_1}\right)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 V_{\bar{D}}(y) + A_{22}\frac{1}{\lambda\lambda_2}V_D(y) + A_{22}\frac{1}{\lambda}V_D(y)\right]$$

$$- \quad w\left[A_{12}\frac{f_D(y)}{f_{\bar{D}}(y)}V_{\bar{D}}(y) + 2A_{22}\left(1 + \frac{1}{\lambda_1}\right)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 V_{\bar{D}}(y) + A_{22}\frac{1}{\lambda\lambda_2}V_D(y)\right]$$

$$+ \quad A_{11}V_{\bar{D}}(y) + A_{12}\frac{f_D(y)}{f_{\bar{D}}(y)}V_{\bar{D}}(y) + A_{22}\left(1 + \frac{1}{\lambda_1}\right)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 V_{\bar{D}}(y) + A_{22}\frac{1}{\lambda\lambda_2}V_D,$$

which is convex since the coefficient for w is greater than zero. Thus Σ_w has a minimum and the results follow from simple algebra.

A3: Proof of Theorem 3

Proof of Theorem 3 follows similar arguments as in proof of Theorem 1 and is thus omitted.

A4: Proof of Theorem 4
Proof of Theorem 4 follows similar arguments as in proof of Theorem 2 and is thus omitted.

A5: Proof of Theorem 5
Theorem 5 is a result of Delta method and the independence between the pilot cohort study (and $\hat{\rho}$) with the case-control sample.

B: Proof of Asymptotic Bias of $\widehat{\mathrm{PPV}}_w$

B1: Proof of Theorem 6
Let $\delta = \mathrm{ROC}^*\{t\} - \mathrm{ROC}(t)$ for $t = S_{\bar{D}}(y)$. Since $\widetilde{S_{\bar{D}}}(y) \rightarrow^p S_{\bar{D}}(y)$, we have

$$
\begin{aligned}
\widehat{\mathrm{ROC}}_w\left\{\widetilde{S_{\bar{D}}}(y)\right\} &= w\widetilde{S_D}(y) + (1-w)\widehat{\mathrm{ROC}}^*\left\{\widetilde{S_{\bar{D}}}(y)\right\} \\
&\rightarrow^p w\mathrm{ROC}\{S_{\bar{D}}(y)\} + (1-w)\mathrm{ROC}^*\{S_{\bar{D}}(y)\} \\
&= \mathrm{ROC}\{S_{\bar{D}}(y)\} + (1-w)\delta \equiv \mathrm{ROC}^\dagger\{S_{\bar{D}}(y)\}.
\end{aligned}
$$

As sample size goes to infinity,

$$
\widehat{\mathrm{PPV}}_w(y) \rightarrow \mathrm{PPV}_w^\dagger(y) = \frac{\mathrm{ROC}^\dagger\left\{S_{\bar{D}}(y)\right\}\rho}{\mathrm{ROC}^\dagger\left\{S_{\bar{D}}(y)\right\}\rho + S_{\bar{D}}(y) \times (1-\rho)}.
$$

The asymptotic bias of $\mathrm{PPV}_w^\dagger(y)$, denoted by γ, is

$$
\begin{aligned}
\gamma &= \mathrm{PPV}^\dagger(y) - \mathrm{PPV}(y) \\
&= \frac{\mathrm{ROC}^\dagger\left\{S_{\bar{D}}(y)\right\}\rho}{\mathrm{ROC}^\dagger\left\{S_{\bar{D}}(y)\right\}\rho + S_{\bar{D}}(y) \times (1-\rho)} - \frac{\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}\rho}{\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}\rho + S_{\bar{D}}(y) \times (1-\rho)} \\
&= C^+ \frac{\mathrm{ROC}^\dagger\{S_{\bar{D}}(y)\} - \mathrm{ROC}\{S_{\bar{D}}(y)\}}{\mathrm{ROC}^\dagger\left\{S_{\bar{D}}(y)\right\}\rho + S_{\bar{D}}(y) \times (1-\rho)} = C^+ \frac{(1-w)\delta}{\rho(1-w)\delta + \rho\mathrm{ROC}\{S_{\bar{D}}(y)\} + S_{\bar{D}}(y)(1-\rho)},
\end{aligned}
$$

where

$$
C^+ = \frac{\rho(1-\rho)S_{\bar{D}}(y)}{\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}\rho + S_{\bar{D}}(y) \times (1-\rho)}.
$$

It is easy to see that γ is monotonically increasing as $(1-w)\delta$ increases, thus for $\delta_0 \leq \delta \leq \delta_1$, γ falls into

$$
\left[C^+ \frac{(1-w)\delta_0}{\rho(1-w)\delta_0 + \rho\mathrm{ROC}\{S_{\bar{D}}(y)\} + S_{\bar{D}}(y)(1-\rho)}, C^+ \frac{(1-w)\delta_1}{\rho(1-w)\delta_1 + \rho\mathrm{ROC}\{S_{\bar{D}}(y)\} + S_{\bar{D}}(y)(1-\rho)}\right].
$$

B2: Proof of Theorem 7
Proof of Theorem 7 follows similar arguments as in proof of Theorem 6 and is thus omitted.

C: Illustration Using a Case-Control Example

To illustrate application of our methodology to a case-control data, we use a constructed case-control sample from the pilot PCA3 data, which includes all cases and an equal number of controls for each population. Specifically, 118 cases and 118 controls from the initial biopsy population are included, as well as 72 cases and 72 controls from the repeat biopsy population. Define $\widehat{\mathrm{NPV}}_w$ to be the weighted estimator for NPV using specificity at a specific sensitivity as the bridge between populations and let $\widehat{\mathrm{NPV}.A}_w$ be the alternative estimator where sensitivity at a specific specificity is used as the bridge.

To evaluate validity of assumptions for $\widehat{\mathrm{PPV}}_w(60)$, $\widehat{\mathrm{PPV}.A}_w(60)$, $\widehat{\mathrm{NPV}}_w(20)$ and $\widehat{\mathrm{NPV}.A}_w(20)$ respectively, tests are conducted using bootstrap variance estimates for equivalence between the two populations with respect

Table 1. Comparison of the two strategies for estimating PPV and NPV. Here "Efficiency" is the ratio of PMSE of the default estimator ($\widetilde{\mathrm{PPV}}$ or $\widetilde{\mathrm{NPV}}$) vs PMSE of the weighted estimator.

	$\widetilde{\mathrm{PPV}}(60)$	$\widehat{\mathrm{PPV}.A_w}(60)$	$\widetilde{\mathrm{NPV}}(20)$	$\widehat{\mathrm{NPV}.A_w}(20)$
		Fixed ρ		
Weight	1	0.77	1	0.60
Est (95% CI)	0.76 (0.63, 0.86)	0.76 (0.65, 0.84)	0.86 (0.79, 0.91)	0.84 (0.80, 0.88)
Bias	-	-0.009	-0.001	-0.01
Variance	0.0034	0.0024	0.0009	0.0004
Efficiency*	1.00	1.41	1.00	1.80
		Estimated ρ		
Weight	1	0.88	1	0.58
Est (95% CI)	0.76 (0.62, 0.86)	0.76 (0.64, 0.85)	0.86 (0.78, 0.91)	0.84 (0.78, 0.89)
Bias	-0.026	-0.033	-0.0016	-0.002
Variance	0.0038	0.0031	0.001	0.0007
Efficiency	1.00	1.07	1.00	1.63

to (i) sensitivity corresponding to 1-specificity $= S_{\bar{D}}(60)$, (ii) specificity corresponding to sensitivity $= S_D(60)$, (iii) specificity corresponding to sensitivity $= S_D(20)$, and (iv) sensitivity corresponding to 1-specificity $= S_{\bar{D}}(20)$. With respect to above four measures, point estimates in the initial and repeat biopsy populations are (i) $\{0.347, 0.315\}$, (ii) $\{0.085, 0.097\}$, (iii) $\{0.525, 0.486\}$, and (iv) $\{0.706, 0.764\}$ respectively. None of the test results are significant. P-values are 0.759, 0.842, 0.719, 0.539 respectively.

We investigate performance of the four estimators over a series of w varying from 0 to 1. Variance and bias of the weighted estimators are computed based on 1000 bootstrap samples, where cases and controls are sampled separately from each population. First, we employ sample disease prevalence from the pilot study as if they were known. Relative efficiency of weighted estimators versus default estimators in terms of PMSE is plotted as function of w (Figure 1). The optimal weights that minimize PMSE for estimating PPV and NPV are identified. Observe that $\widehat{\mathrm{PPV}}_w(60)$ and $\widehat{\mathrm{PPV}.A_w}(60)$ have similar optimal efficiency, with the latter slightly better. $\widehat{\mathrm{NPV}.A_w}(20)$ is slightly more efficient than $\widehat{\mathrm{NPV}}_w(20)$ at optimal weights. When variability in prevalence estimate is taken into account, the pattern of varying efficiency as weight changes are similar to that when ρ is treated as fixed, but the maximal efficiency relative to the default estimator decreases (Figure 1). Again, $\widehat{\mathrm{PPV}.A_w}(60)$ and $\widehat{\mathrm{NPV}.A_w}(20)$ achieve similar or slightly larger optimal efficiency compared to $\widehat{\mathrm{PPV}}_w(60)$ and $\widehat{\mathrm{NPV}}_w(20)$.

Results comparing $\widehat{\mathrm{PPV}.A_w}(60)$ and $\widehat{\mathrm{NPV}.A_w}(20)$ at their optimal weights and corresponding default estimators are presented in Table 1. For both PPV(60) and NPV(20), the weighted estimate and the default estimate are pretty similar to each other. When disease prevalence is treated as fixed, efficiency gain based on the weighted estimator is around 35% for PPV(60) and 86% for NPV(20), which is not surprising given that there are more subjects in the initial biopsy population. When variability in disease prevalence is taken into account, efficiency gains using the weighted estimators are around 26% for PPV(60) and 43% for NPV(20).

Next we study robustness of $\widehat{\mathrm{PPV}.A_w}(60)$ and $\widehat{\mathrm{NPV}.A_w}(20)$ at their optimal weights to violation from the assumption, taking variability in $\hat{\rho}$ into consideration. In Figure 2 we show that in order to cause 5% (*relative bias*) over- or under-estimation in PPV(60), how big the difference in 1-specificity corresponding to sensitivity $= S_D(60)$ needs to be between the two populations. Also displayed is the required difference in sensitivity corresponding to 1-specificity $= S_{\bar{D}}(20)$, in order to cause 5% over- or under-estimation in NPV(20). Note that for PPV(60) to be over- or under-estimated by 5% using the optimally weighted estimator, 1-specificity corresponding to sensitivity $= S_D(60)$ needs to be smaller by 0.143 or larger by 0.158 in the repeat biopsy population than the initial biopsy population. These correspond to 0 and 95.9 percentiles in the distribution of the 1-specificity differences constructed by bootstrap resampling. On the other hand, for NPV(20) to be over- or under-estimated by 5% by the optimally-weighted estimator, sensitivity corresponding to 1-specificity $= S_{\bar{D}}(20)$ needs to be larger
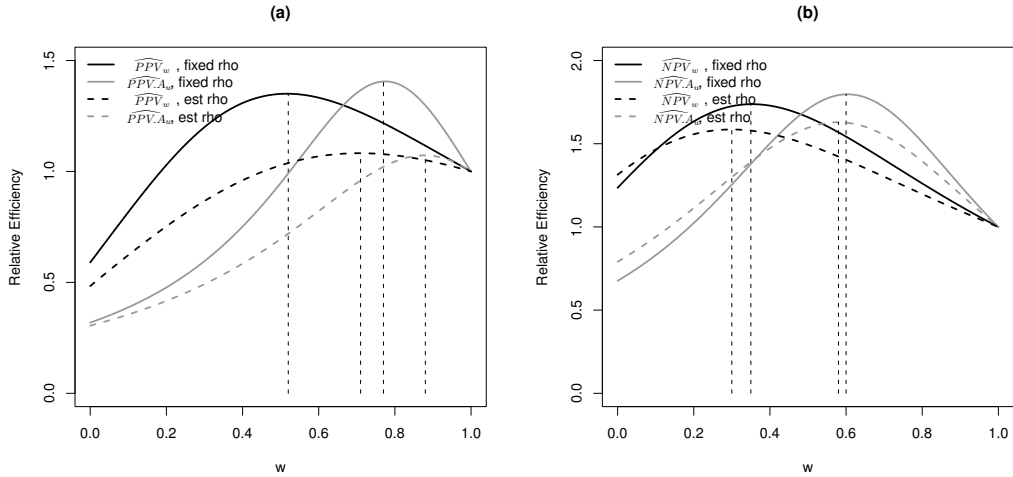
Fig. 1. Relative efficiency of the proposed estimator vs default estimator of (a) PPV(60) and (b) NPV(20) as function of weight.

by 0.186 or smaller by 0.206 in the initial biopsy population than the repeat biopsy population. These correspond to 99.4 and 4.4 percentiles in the bootstrap distribution of the sensitivity difference. Therefore, it is highly unlikely that the optimally-weighted PPV(60) or $NPV(20)$ estimator can lead to 5% over-estimation, although there is a small chance that these estimators could be under-estimated by 5%.

References

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.

Fig. 2. Difference in classification accuracy between the two populations to achieve 5% over- or under-estimation (relative bias) in PPV(60) and NPV(20). The black arrowheads are sensitivities in the initial population corresponding to 1-specificity $= S_{\bar{D}}(20)$, in order to cause 5% over- or under-estimation in NPV(20) of the repeat biopsy population; the grey arrowheads are 1-specificities in the repeat biopsy population corresponding to sensitivity $= S_D(60)$, in order to cause 5% over- or under-estimation in PPV(60) of the initial biopsy population.