10-10-2012

Borrowing Information Across Populations in Estimating Positive and Negative Predictive Values

Ying Huang
Fred Hutchinson Cancer Research Center, Seattle, WA, ying@u.washington.edu

Youyi Fong
Fred Hutchinson Cancer Research Center, Seattle, WA

John Wei
University of Michigan - Ann Arbor

Ziding Feng
Fred Hutchinson Cancer Research Center, Seattle, WA

Published in final edited form as:
J R Stat Soc Ser C Appl Stat. 2011 November 1; 60(5): 633–653. doi:10.1111/j.1467-9876.2011.00761.x.

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.

Summary

A marker's capacity to predict risk of a disease depends on disease prevalence in the target population and its classification accuracy, i.e. its ability to discriminate diseased subjects from non-diseased subjects. The latter is often considered an intrinsic property of the marker; it is independent of disease prevalence and hence more likely to be similar across populations than risk prediction measures. In this paper, we are interested in evaluating the population-specific performance of a risk prediction marker in terms of positive predictive value (PPV) and negative predictive value (NPV) at given thresholds, when samples are available from the target population as well as from another population. A default strategy is to estimate PPV and NPV using samples from the target population only. However, when the marker's classification accuracy as characterized by a specific point on the receiver operating characteristics (ROC) curve is similar across populations, borrowing information across populations allows increased efficiency in estimating PPV and NPV. We develop estimators that optimally combine information across populations. We apply this methodology to a cross-sectional study where we evaluate PCA3 as a risk prediction marker for prostate cancer among subjects with or without previous negative biopsy.

Keywords

Biomarker; Classification; NPV; PPV; Sensitivity; Specificity

1. Introduction

The two most commonly used criteria for biomarker evaluation are classification accuracy and risk prediction capability. Classification accuracy, typically characterized by sensitivity, specificity, and ROC curve (Pepe, 2003), measures the probability that a subject's disease status is correctly identified based on a biomarker. Risk prediction measures, on the other hand, assess how well a marker can inform treatment options based on predicted risk of disease. Among others, two measures often used are positive predictive value (PPV) and negative predictive value (NPV) (Leisenring *et al.*, 2000; Moskowitz and Pepe, 2004, 2006; Steinberg *et al.*, 2008). It is well known that sensitivity, specificity, and ROC curve are intrinsic properties of a test while PPV and NPV depend on both classification accuracy and the external factor, i.e. disease prevalence. However, there has been no method that utilizes

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

this property to gain efficiency in estimating PPVs and NPVs in populations of different disease prevalence when data suggest common intrinsic classification accuracy across populations, as in the application below that motivated this paper.

PCA3, a prostate-specific noncoding mRNA overexpressed in prostate tumors, has been proposed as a risk prediction marker for prostate cancer. In a preliminary cross-sectional study, data were collected from 576 men immediately prior to their prostate biopsy which was scheduled mainly due to elevated PSA (Deras *et al.*, 2006). About half of the subjects had previous negative biopsy and the rest did not. The disease outcome is the prostate cancer status diagnosed by the biopsy. Based on these data, urologists are interested in evaluating PCA3's risk prediction performance in terms of PPV and NPV in the population of subjects who had had previous biopsy and the population of subjects who had not had previous biopsy. In particular the data suggested that PPV at PCA3=60, which is approximately 0.75 in the initial biopsy population could be used as a threshold for recommendation of prostate biopsy, and that NPV at PCA3=20, which is approximately 0.85 in the repeat biopsy population, to recommend against prostate biopsy. These thresholds were recommended by study urologists based on the fact that most prostate cancers are indolent and the fact that prostate cancer prevalence in the initial biopsy population is about 44%, and in the repeat biopsy population the prevalence is much lower around 27%. The difference in prevalence is due to the fact that larger tumors are likely to be detected in the initial biopsy and that most prostate cancer patients were detected from their initial biopsy.

Figure 1(a) shows the density functions of log(PCA3) conditional on disease status within the initial and repeat biopsy populations, while Figure 1(b) shows the empirical ROC curves in the two populations. Interestingly, although the distributions of PCA3 conditional on disease status appear to differ between the two populations (e.g., Wilcoxon rank sum test applied to the non-cancer groups has a p-value 0.043), the two ROC curves appear similar to each other: the test of equal area under the curve has a p-value of 0.66. Existence of scenarios where the ROC curve is similar between different sources is not hard to picture, considering the fact that the ROC curve characterizes the comparison of diseased individuals and non-diseased individuals with respect to their relative ranks rather than actual values. For example, it is common that assays from different clinical centers could have different distributions due to many instrumental and specimen handling factors, leading to some location-scale shifts of the test results across clinical centers yet not changing the classification performance.

One major reason in favor of calculating PPV/NPV separately from each target population is that there are standard formulas for PPV/NPV for a single population as shown in the Method section, but there is no existing method for combining data across populations for estimating PPV/NPV based on the assumption of common classification accuracy, unless one uses stronger assumptions, e.g. a location shift modeled by a population effect indicator in the marker distributions conditional on disease status. The objective of the analysis described in this paper is to develop a statistical method for estimating population-specific PPV and NPV using the ROC curve as a bridge between populations when data strongly suggest the same classification accuracy across populations. This requires the assumption that relative ranks between diseased and non-diseased are the same across populations. Making assumptions based on rank is not uncommon in statistical literature due to the increased robustness compared to making parametric assumptions on marker distribution. Examples include the Friedman test (Friedman, 1937) and Quade test (Quade, 1979) in randomized block design. The procedure proposed in this paper can be thought of as an expansion of these nonparametric methods to PPV/NPV estimation, rather than a simple hypothesis test of equality of rank means. Combining information non-parametrically has a long history. For example, Mantel and Haenszel (1959) combined odds ratios across strata.

In our example, it is desirable to have a method that relies only on the similar rank distribution assumption and does not require explicit modeling e.g. the location-scale shift effects on the marker distribution conditional on disease status.

The settings in which the proposed procedure will be useful assume that any "interaction" effect of biomarker and population in terms of discriminating diseased from non-diseased individuals is negligible. That is, the difference in the marker's discriminatory power between populations is minimal. This assumption should always be checked. When interaction is substantial, results from any of the above methods combining information across populations, include the method proposed in this paper, will be less interpretable and the estimation should be done for each population. The main motivation of this paper is to provide a non-parametric method for combining classification information across population/strata when the combined estimation is desired and justified.

While cross-sectional samples and cohort samples are usually collected in the late phases of biomarker studies, a case-control sampling design is most often used in the early phase of biomarker development (Pepe *et al.*, 2001). In Section 2, we start by considering a case-control design and investigate cross-sectional and cohort designs later in Section 3. We present simulation studies in Section 4 and detailed analyses of the PCA3 example in Section 5. Finally we provide concluding remarks in Section 6.

2. Methods in Case-Control Design

Let D be a binary disease status and let Y be a continuous biomarker of interest. Suppose samples are available from two populations, the target population where PPV and/or NPV are of interest, and another population we call the auxiliary population. In the prostate cancer example, the repeat biopsy population serves as the auxiliary population when we are interested in estimating PPV and/or NPV in the initial biopsy population, and the initial biopsy population would serve as the auxiliary population when we are interested in estimating PPV/NPV in the repeat biopsy population. We use subscript D and $\bar{D}$ to indicate diseased and non-diseased status, and use superscript to indicate the auxiliary population. Let Y, Y_D, and $Y_{\bar{D}}$ be the marker measured for a random subject, a case, and a control respectively from the target population, and let $Y^*, Y^*_D, Y^*_{\bar{D}}$ indicate the corresponding quantities in the auxiliary population. Let $S(y) = P(Y > y)$ denote the survival function for Y; S_D, $S_{\bar{D}}$ denote the survival functions for Y_D and $Y_{\bar{D}}$; $Y_{\bar{D}}; Y^*_D, Y^*_{\bar{D}}$ denote the survival functions for Y^*_D and $Y^*_{\bar{D}}$. Suppose we apply a binary classification rule to the target population such that compared to a given threshold, a subject is classified as diseased if his marker value is greater than the threshold and non-diseased otherwise. Then the ROC curve is the plot of true positive rate versus false positive rate for a series of thresholds, and it can be expressed as $\mathrm{ROC}(t) = S_D\left\{S^{-1}_{\bar{D}}(t)\right\}$. Similarly, let ROC* be the corresponding ROC curve in the auxiliary population. We have $\mathrm{ROC}^*(t) = S^*_D\left\{S^{*-1}_{\bar{D}}(t)\right\}$. Throughout this manuscript we assume larger marker values are associated with higher risks of disease.

Next we explore methods for estimating $\mathrm{PPV}(y) = P(D = 1\ Y > y)$. Results for NPV are omitted since they are easy to derive by exploiting the symmetry between the two: $\mathrm{NPV}(y) = P(D = 0\ Y \le y)$ can be represented as $\mathrm{PPV}(-y)$ when D is replace by $1 - D$ and Y replaced by Y.

Let ρ indicate disease prevalence in the target population, which we assume initially to be known. By an application of Bayes' theorem, PPV can be written as a function of ρ, $S_{\bar{D}}$, and S_D:

$$\mathrm{PPV}(y) = \frac{\rho S_D(y)}{\rho S_D(y) + (1-\rho) S_{\bar{D}}(y)}. \tag{1}$$

Writing y as $S_{\bar{D}}^{-1} S_{\bar{D}}(y)$ and using the definition of the ROC curve, PPV can be represented as a function of ρ, $S_{\bar{D}}$, and $\mathrm{ROC}\{S_{\bar{D}}(y)\}$:

$$\mathrm{PPV}(y) = \frac{\rho S_D\left\{S_{\bar{D}}^{-1} S_{\bar{D}}(y)\right\}}{\rho S_D\left\{S_{\bar{D}}^{-1} S_{\bar{D}}(y)\right\} + (1-\rho) S_{\bar{D}}(y)} = \frac{\rho \mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}}{\rho \mathrm{ROC}\left\{S_{\bar{D}}(y)\right\} + (1-\rho) S_{\bar{D}}(y)}. \tag{2}$$

Suppose we sample n_D cases $\{Y_{D1}, \ldots, Y_{Dn_D}\}$ and $n_{\bar{D}}$ controls $\{Y_{\bar{D}1}, \ldots, Y_{\bar{D}n_{\bar{D}}}\}$ from the target population and n_D^* cases $\left\{Y_{D1}^*, \ldots, Y_{Dn_D^*}^*\right\}$ and $n_{\bar{D}}^*$ controls $\left\{Y_{\bar{D}1}^*, \ldots, Y_{\bar{D}n_{\bar{D}}^*}^*\right\}$ from the auxiliary population. The default strategy for estimating PPV(y) is to estimate $S_{\bar{D}}(y)$ and $S_D(y)$ empirically with $\tilde{S}_{\bar{D}}(y) = \Sigma_{i=1}^{n_{\bar{D}}} \left(Y_{\bar{D}i} > y\right)/n_{\bar{D}}$ and $\tilde{S}_D(y) = \Sigma_{i=1}^{n_D} (Y_{Di} > y)/n_D$ and enter them into (1). Denote this estimator $\widetilde{\mathrm{PPV}}(y)$. This estimator is asymptotically equivalent to estimating $S_{\bar{D}}(y)$ with $\tilde{S}_{\bar{D}}(y)$ and estimating $\mathrm{ROC}\{S_{\bar{D}}(y)\}$ empirically with $\widetilde{\mathrm{ROC}}\left\{\tilde{S}_{\bar{D}}(y)\right\} = \Sigma_{i=1}^{n_D}\left[Y_{Di} > \tilde{S}_{\bar{D}}^{-1}\left\{\tilde{S}_{\bar{D}}(y)\right\}\right]/n_D$, and entering them into (2), since $\widetilde{\mathrm{ROC}}\left\{\tilde{S}_{\bar{D}}(y)\right\} \simeq \Sigma_{i=1}^{n_D}\{Y_{Di} > y\}/n_D = \tilde{S}_D(y)$, where the approximation is exact when y is one of the data points in the sample from the target population.

2.1. Proposed Estimator

If, in addition, we have $\mathrm{ROC}(t) = \mathrm{ROC}^*(t)$ for $t = S_{\bar{D}}(y)$, that is, the sensitivity corresponding to the specificity 1 - $S_{\bar{D}}(y)$ is constant across the two populations, we can then estimate ROC(t) at $t = S_{\bar{D}}(y)$ using samples from both populations. Let $\widehat{\mathrm{ROC}}(t)$ and $\widehat{\mathrm{ROC}}^*(t)$ be the empirical ROC from the target and auxiliary population respectively, the common $\mathrm{ROC}(t)$ at $t = S_{\bar{D}}(y)$ can be estimated as a weighted average of the two $\widehat{\mathrm{ROC}}_w(t) = w\widehat{\mathrm{ROC}}(t) + (1-w)\widehat{\mathrm{ROC}}^*(t)$, where $t = \tilde{S}_{\bar{D}}(y)$ and w indicates the weight given to the empirical ROC estimate from the target population.

Entering $\widehat{\mathrm{ROC}}_w\left\{\tilde{S}_{\bar{D}}(y)\right\}$ and $\tilde{S}_{\bar{D}}(y)$ into (2), the weighted estimator for PPV(y) is

$$\widehat{\mathrm{PPV}}_w(y) = \frac{\rho \widehat{\mathrm{ROC}}_w\left\{\tilde{S}_{\bar{D}}(y)\right\}}{\rho \widehat{\mathrm{ROC}}_w\left\{\tilde{S}_{\bar{D}}(y)\right\} + (1-\rho)\tilde{S}_{\bar{D}}(y)},$$

where $w = 1$ corresponds to estimating PPV using samples from the target population only. Under the equal classification accuracy assumption, the asymptotic unbiasedness of the ROC and consequently that of the PPV estimators are invariant to the choice of w.

Let f_D and $f_{\bar{D}}$ be density functions of the marker for diseased and nondiseased individuals respectively in the target population. In Theorem 1 of Appendix A, we show that under the assumption of equal sensitivity at specificity $1 - S_{\bar{D}}(y)$, $\sqrt{n_{\bar{D}}}\left\{\widehat{\mathrm{PPV}}_w(y) - \mathrm{PPV}(y)\right\}$ is asymptotically normally distributed with zero mean and a variance term that is a function of w, ρ, $S_D(y)$, $S_{\bar{D}}(y)$ and the density ratio $f_D(y)/f_{\bar{D}}(y)$. Interestingly, since the asymptotic variance of $\widehat{\mathrm{PPV}}_w(y)$ as shown in (4) is a quadratic and convex function of w, an optimal w that minimizes it can be uniquely determined, as presented in equation (5) of Appendix A. Moreover, observe that the asymptotic variance term (4) can be written as the product of two terms, one free of w and the other free of ρ. Consequently the asymptotic relative efficiency of any two estimators with specific weights is independent of the disease prevalence. In other words the optimal w is the same for all ρ. As shown in Appendix A, the optimal w is always less than 1. It converges to 1 when $n_D^*/n_D \to 0$ or when $n_{\bar{D}}^*/n_{\bar{D}} \to 0$. This is anticipated intuitively since $\widehat{\mathrm{ROC}}^*$ is less precise than $\widehat{\mathrm{ROC}}$ under these scenarios and we want to put more weight on the latter.

2.2. Alternative Estimator

Earlier we proposed estimating the specificity at a given threshold y empirically using data from the target population, and estimating the corresponding sensitivity using data from both populations. Alternatively, we can start from the other direction. That is, we could estimate the sensitivity at y empirically using data from the target population, and estimate the corresponding specificity using data from both populations. We call this estimator

$$\widehat{\mathrm{PPV}}.A_w(y) = \rho\left\{\widetilde{S}_D(y)\right\} / \left[\rho\left\{\widetilde{S}_D(y)\right\} + (1-\rho)\,\widehat{\mathrm{ROC}}_w^{-1}\widetilde{S}_D(y)\right], \text{ where}$$

$$\widehat{\mathrm{ROC}}_w^{-1}\left\{\widetilde{S}_D(y)\right\} = w\Sigma_{i=1}^{n_{\bar{D}}}\left\{Y_{\bar{D}i} > y\right\}/n_{\bar{D}} + (1-w)\,\Sigma_{i=1}^{n_{\bar{D}}^*}\left[Y_{\bar{D}i}^* > \widetilde{S}_D^{*-1}\left\{\widetilde{S}_D(y)\right\}\right]/n_{\bar{D}}^*.$$

Asymptotic theory for this estimator and the optimal w for minimizing asymptotic variance are established in Theorems 3 and 4 of Appendix A. Again, the optimal w is always less than 1 and independent of ρ. Interestingly, through simple algebra, it can be shown that the minimum asymptotic variance achievable by $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}}.A_w$ are *equivalent*. Consequently, as far as variance is concerned, asymptotically it does not matter whether we use sensitivity at the given specificity as the bridge between populations or the other way around. We evaluate finite sample performance of the two estimators through simulation studies.

2.3. Imperfect Disease Prevalence Estimate

So far we have assumed that the disease prevalence is known. Sometimes this is reasonable, for example, if we obtain ρ from a population disease registry such as SEER (http://seer.cancer.gov/), its value essentially can be treated as known due to the large sample size involved. Alternatively a disease prevalence estimate $\hat{\rho}$ might be derived from a pilot cross-sectional study, like in our PCA3 application. Under such circumstances, the asymptotic variance of $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}}.A_w$ computed in Sections 2.1 and 2.2 could be easily modified to incorporate the variability in $\hat{\rho}$ as shown in Theorem 5 of Appendix A. Suppose we estimate sample prevalence from a pilot cohort study and apply it to the estimate of PPV based on the case-control sample, then the asymptotic variance of $\widehat{\mathrm{PPV}}_w$ or

$\widehat{\text{PPV}}.A_w$ will equal to their asymptotic variance given 'true' ρ plus an extra term due to the estimation of ρ. From Theorem 5, it can be easily seen that the optimal weights are invariant to the extra variability introduced, and are the same as those in equations (5) and (7) where the disease prevalence is considered to be known. The efficiency of the optimal estimator relative to the default estimator is expected to decrease as variability in the disease prevalence estimator increases due to a dampening effect.

2.4. Robustness

The estimators proposed in Sections 2.1 and 2.2 gain precision by assuming equality between ROC$\{S_{\bar{D}}(y)\}$ and ROC*$\{S_{\bar{D}}(y)\}$ or between $S^*_{\bar{D}} S^{*-1}_{\bar{D}} \{S_D(y)\}$ and $S_{\bar{D}} S^{-1}_{D} \{S_D(y)\}$, it is important to be aware of the magnitude of the bias in $\widehat{\text{PPV}}_w$ or $\widehat{\text{PPV}}.A_w$ when the corresponding assumptions are violated.

Let $\delta = \text{ROC}^*\{t\} - \text{ROC}(t)$ for $t = S_{\bar{D}}(y)$ and let $\eta = -\left[S^*_{\bar{D}} S^{*-1}_{\bar{D}} \{S_D(y)\} - S_{\bar{D}} S^{-1}_{\bar{D}} \{S_D(y)\}\right]$. As shown in Theorems 6 and 7 of Appendix B, the asymptotic bias of $\widehat{\text{PPV}}_w$ can be represented as a monotone increasing function of $(1 - w)\delta$, and the asymptotic bias of $\widehat{\text{PPV}}.A_w(y)$ is a monotone increasing function of $(1 - w)\eta$.

In practice, researchers might be able to guess a suitable the range for δ or η based on experience. Alternatively, an interval of δ or η consistent with the data can be derived at, say 95% confidence level. Then the asymptotic bias of the proposed estimator can be calculated, and combined with the reduction in variance to determine the "worst-case" impact on the mean squared error. Conversely, given a range of tolerable bias in $\widehat{\text{PPV}}_w(y)$ or $\widehat{\text{PPV}}.A_w(y)$, we can derive the corresponding tolerable range for δ or η.

2.5. Weight Determination and Variance Estimation

We propose two approaches for determining the optimal weight w for computing $\widehat{\text{PPV}}_w$ or $\widehat{\text{PPV}}.A_w$ and subsequently estimating variance of the weighted estimators. The first approach is based on the closed-form formula for w as presented in (5) and (7) in Appendix A for minimizing asymptotic variance of the weighted estimators under equal classification accuracy condition. Equations (6) and (7) involve a density ratio $f_D/f_{\bar{D}}$, which would be difficult to estimate without making any parametric assumption about the marker distribution. We thus propose to assume normality of Y in the target population conditional on disease status and then compute (6) and (7) based on estimated distribution parameters. In practice, if we could transform data such that the normality assumption is not grossly violated, then we expect that the weight estimated assuming normality would be a good approximation to the true entity. Note that since the choice of w will affect only efficiency of the estimator but not its consistency, robustness to deviation from normality is not a big issue for weight determination. Given selected w, one could apply asymptotic formula (4) and (5) based on normality assumption for estimation of variance. However, here deviation from normality could potentially bias the variance estimation and invalidate the inference. Therefore, we recommend instead using bootstrap resampling to estimate variance of the weighted estimator after the optimal w is obtained through the asymptotic formula. The resampling scheme will be chosen to reflect sampling design.

Validity of the above approach for determining w relies on the equal classification accuracy assumption. In practice, a researcher's choice of approaches for weight determination and variance estimation depends on the problem investigated and reflects how strong one's belief is about the equal classification accuracy assumption and how heavily one is concerned

about the possible bias under the violation of this assumption. There are scenarios where the equal classification accuracy is expected to hold where the approach described above is best suited. For example, consider a medical test performed at two different labs. It is not uncommon to assume that the lab difference leads to a location-scale shift in distribution of the test results but does not change the ranks of diseased versus non-disease, and thus a common ROC curve exists. In other scenarios where the equal classification accuracy assumption is built largely upon statistical tests rather than prior knowledge about the underlying biological mechanism, as in our PCA3 application, researchers might want to be conservative in terms of controlling possible bias while improving efficiency.

With an objective of maintaining a balance between bias and variance, here we propose a second bootstrap-based approach for determining w. Specifically, we generate a bootstrap set based on the observed dataset and implement a grid search algorithm to examine a series of candidate w values. In our simulation studies and application, a grid size of 0.01 is used. For each w, we estimate the bootstrap variance of the weighted estimators. At the same time, to account for possible deviation from the equal classification accuracy assumption, we also compute a 'bias' or penalty term as the difference between means of the weighted estimators over the bootstrap distribution and the default estimator based on the original data. A weighted estimator with minimum 'pseudo mean squared error' (PMSE), which is defined as the sum of the squared penalty and bootstrap variance, can then be selected out of all possible w values and between $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}}.A_w$. Note that here we use the same set of bootstrap samples for choosing w and for variance estimation. Doing so ignores the variability due to estimation of w. Conceptually, a more complicated bootstrap procedure could be implemented to account for the variability in estimating w. However, it appears that given practical sample size, ignoring the contribution to variability due to estimating w has minimal impact on the inference, as shown by the satisfactory coverage of the weighted estimators in simulation studies. We thus adopt this simpler bootstrap procedure instead of going for more complicated procedure.

3. Estimation in Cross-sectional or Cohort Design

The estimators we developed in Section 2 for case-control design is directly applicable to prospective sampling design. Consider the setting where n individuals in the target populations are randomly sampled, among which n_D subjects are diseased. Then disease prevalence in the target population can be estimated by $\hat{\rho} = n_D/n$, while estimators $\hat{S}_D(y)$ and $\hat{S}_{\bar{D}}(y)$ are computed in the same way as in Section 2. As demonstrated in Appendix C, here $\hat{\rho}$ is uncorrelated with $\hat{S}_D(y)$ or $\hat{S}_{\bar{D}}(y)$, considering the fact that $\hat{S}_D(y)$ and $\hat{S}_{\bar{D}}(y)$ are estimated from the conditional distributions of marker given disease status, while $\hat{\rho}$ is a function only of disease status data. Consequently, the asymptotic properties of $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}}.A_w(y)$ are the same as those presented in Theorem 5.

4. Simulation

We conduct simulation studies to investigate the performance of the weighted estimators developed in earlier sections, using a case-control design. Assuming

$$\begin{aligned} Y_{\bar{D}} &\sim N(0,1), \quad Y_D \sim N(1,1), \\ Y^*_{\bar{D}} &\sim N(0.5,1) \quad Y^*_D \sim N(1.5,1). \end{aligned} \tag{3}$$

Our goal is to estimate PPV(y) in the target population. In the simulation, equal number of samples are obtained from the target population and from the auxiliary population, and within each population equal number of cases and controls are sampled. We study the

setting where $\rho = 0.4$, close to that of the initial biopsy population in the PCA3 example. Results are presented for y being the 90^{th} percentile within controls, w varying from 0.1 to 0.9, and total sample size of either 500 or 1000. Results are based on 1000 Monte-Carlo simulations with bootstrap sample size 250.

First we assume ρ is known. Table 1 shows that both $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}}.A_w(y)$ have minimal biases. Asymptotic variances under a series of w are fairly close to the corresponding finite sample variances. Large efficiency gain can be achieved by borrowing information across populations compared to the default strategy. Wald confidence intervals (CI) based on bootstrap variance estimate have coverage close to nominal level assuming logit of the estimators are normally distributed.

Also presented are the results when we assume disease prevalence in the target population is estimated from a pilot cohort study with sample size 250 or 500 respectively for a follow-up case-control study of sample size 500 or 1000 (Table 2). Again the proposed estimators have good performances. The efficiency of the proposed estimators relative to the default estimator is smaller with imperfect disease prevalence estimate compared to that given perfect disease prevalence.

Next we examine the performance of the weighted estimators when weight is selected by assuming a normal marker distribution conditional on disease status or through the bias-penalized bootstrap procedure. With marker distributions following (3), we study the efficiency of $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}}.A_w(y)$ relative to $\widehat{\mathrm{PPV}}(y)$ as well as their coverage property. Table 3 presents efficiency of the weighted estimator relative to the default estimator for varying disease prevalence in the target population, $\rho = \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and varying threshold y corresponding to $v = 1 - S_{\bar{D}}(y) = \{0.1, 0.3, 0.5, 0.7, 0.9\}$, $S_D(y) = \{0.989, 0.936, 0.841, 0.682, 0.389\}$, for $n_D = n_{\bar{D}} = n_D^* = n_{\bar{D}}^* = 250$. It appears that weight selected under normality assumption achieves the optimal efficiency in general. The efficiency gain is similar between $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}}.A_w$. The weight selected by the bias-penalized bootstrap procedure achieves smaller but still sizable efficiency compared to the model-based procedure assuming equal classification accuracy. This is not surprising considering that the penalty terms adopted by the bootstrap procedure essentially 'shrink' the weighted estimator towards the default estimator. Table 4 shows coverage of 95% Wald CI based on bootstrap estimated variance for the weighted estimators, assuming normality of the logit-transformed estimator. Both procedures of weight selection have satisfactory coverage.

We also investigate robustness of the weighted estimators to violation of the common classification accuracy assumption. We simulate data from two populations with difference in ROC curves:

$$Y_{\bar{D}} \sim N(0,1), \quad Y_D \sim N(1,1),$$
$$Y_{\bar{D}}^* \sim N(0.5,1) \quad Y_D^* \sim N(1.8,1),$$

$n_D = n_{\bar{D}} = n_D^* = n_{\bar{D}}^* = 250$. Again, varying ρ, {0.1, 0.3, 0.5, 0.7, 0.9}, and thresholds y corresponding to $v = 1 - S_{\bar{D}}(y) = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and $S_D(y) = \{0.989, 0.936, 0.841, 0.682, 0.389\}$ in the target population are considered. In the auxiliary population, corresponding to the same set of specificity as in the target population, values of S_D^* are {0.995, 0.966, 0.903, 0.781, 0.507} respectively, while corresponding to the same set of sensitivity as in the target population, values of $1 - S_{\bar{D}}^*$ are {0.163, 0.411, 0.618, 0.795,

0.943} respectively. Results of relative bias for PPV estimators with weights selected by normal model or by bootstrap are presented in Table 5 as a function of v and ρ, where ρ is assumed to be known. Note that overall by including the extra penalty term, estimators with weights selected by penalized bootstrap have much smaller bias compared to the estimators with weights selected assuming normality under equal classification accuracy assumption. When weights are determined parametrically, the magnitude of bias for $\widehat{\mathrm{PPV}.A_w}(y)$ relative to $\widehat{\mathrm{PPV}}_w(y)$ tends to be larger when y is at the lower end of its distribution and smaller when y is at the upper end of its distribution. Intuitively this makes sense considering that bias in $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}.A_w}$ relates to the difference between sensitivity at given specificity and the difference between specificity at given sensitivity respectively. For two unequal ROC curves, the horizontal difference tends to be smaller than the vertical difference at the lower end of the curve, i.e. where the ROC curve is steeper, which corresponds to large y; whereas the order of the horizontal and vertical distance reverses at the upper end of the ROC curve where the ROC curve is flatter and y is small. When the bias-penalized bootstrap procedure is used for weight selection, the bias is similar between $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}.A_w}$.

5. Application to PCA3 Study

In the PCA3 study (Deras *et al.*, 2006), information was collected for 267 subjects from the initial biopsy population and another 269 different subjects from the repeat biopsy population. As mentioned in the Introduction, researchers are interested in evaluating PCA3's ability to identify high risk subjects in the initial biopsy population and its ability to identify low risk subjects in the repeat biopsy population. PPV(60) and NPV(20) were chosen as the measures to evaluate.

Define $\widehat{\mathrm{NPV}}_w$ to be the weighted estimator for NPV using specificity at a particular sensitivity as the bridge between populations and let $\widehat{\mathrm{NPV}.A_w}$ be the alternative estimator where sensitivity at a particular 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 to (i) sensitivity corresponding to 1-specificity = $S_D(60)$, (ii) specificity corresponding to sensitivity = $S_D(60)$, (iii) specificity corresponding to sensitivity = $S_D(20)$, and (iv) sensitivity corresponding to 1-specificity = $S_D(20)$. With respect to these four measures, point estimates in the initial and repeat biopsy populations are (i) {0.314, 0.236}, (ii) {0.081, 0.132}, (iii) {0.730, 0.764}, and (iv) {0.503, 0.487} respectively. None of the test results are significant. P-values are 0.433, 0.315, 0.665, 0.864 respectively.

While the equal classification accuracy assumption appears plausible from the data, without a better understanding of the potential biological mechanism behind it, we decide to be conservative and apply the bias-penalized bootstrap method of weight selection for robustness against possible difference in classification accuracy between the two populations. 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 2,000 bootstrap samples, where individuals are sampled separately from each population. Ratio of PMSE for default estimator versus weighted estimators is plotted as function of w (Figure 2). The optimal weights that minimize PMSE for estimating PPV and NPV are identified. Observe that $\widehat{\mathrm{PPV}.A_w}(60)$ is slightly more efficient compared to $\widehat{\mathrm{PPV}}_w(60)$ at optimal weights. $\widehat{\mathrm{NPV}}_w(20)$ and $\widehat{\mathrm{NPV}.A_w}(20)$ have similar optimal efficiency, with the latter slightly better.

Results comparing $\widehat{\text{PPV}}.A_w(60)$ and $\widehat{\text{NPV}}.A_w(20)$ at their optimal weights and corresponding default estimators are presented in Table 6. For both PPV(60) and NPV(20), the weighted estimate and the default estimate are fairly similar to each other. In terms of variance, efficiency gain based on the weighted estimator is around 38% for PPV(60) and 93% for NPV(20). This is not surprising, considering that in the initial biopsy population the numbers of cases and controls are more balanced and there is more variability due to the disease prevalence estimate (since ρ is closer to 0.5). The PMSE of the default estimator exceeds that of the weighted estimator by around 20% for PPV(60) and 78% for NPV(20).

Next we study robustness of $\widehat{\text{PPV}}.A_w(60)$ and $\widehat{\text{NPV}}.A_w(20)$ at their optimal weights to violation from the equal classification accuracy assumption. Figure 3 shows how large the difference in 1-specificity corresponding to sensitivity = $S_D(60)$ needs to be between the two populations to cause 5% (relative bias) over- or under-estimation in PPV(60). 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 0.13 or larger by 0.14 in the repeat biopsy population compared to the initial biopsy population. These correspond to zero and 91.6 percentiles in the distribution of the 1-specificity differences constructed by bootstrap resampling. Consequently, it is unlikely that the optimally-weighted estimator can lead to 5% over-estimation in PPV(60), although there is some chance that PPV(60) might be under-estimated. 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 by 0.16 or smaller by 0.18 in the initial biopsy population than the repeat biopsy population. These correspond to 99.0 and 1.7 percentiles in the bootstrap distribution of the sensitivity difference. Therefore, it is highly unlikely that the optimally-weighted NPV(20) estimator can lead to 5% over- or under-estimation. The weighted estimators seem to be fairly robust in this example.

To get a more conservative view of the bias-variance trade-off in our example. We entertained the “worst-case” bias defined as the boundary of the 95% CI for difference in classification accuracy between the two populations. We look at upward or downward bias in the weighted PPV/NPV estimators separately. Suppose the true predictive values are over-estimated by weighting. Weighting leads to 25.7% and 15.5% decreases in PMSE for estimating PPV(60) and NPV(20) respectively. If the true predictive values are under-estimated, weighting leads to a 4.0% drop in PMSE for estimating NPV(20), and a 21.3% increase in PMSE for estimating PPV(60). These results further press our point that the weighted estimator is desirable in the PCA3 example especially for estimating NPV(20) in terms of reducing mean squared error.

We also try the model-based procedure for weight selection assuming normality of log(PCA3) conditional on disease status. Smaller optimal weights are selected compared to bias-penalized bootstrap weight selection ($w = 0.60$ for $\widehat{\text{PPV}}_w$ and $w = 0.49$ for $\widehat{\text{NPV}}.A_w$). Corresponding PPV(60) and NPV(20) estimates are 0.73 (95% CI: 0.62-0.82) and 0.85 (95% CI: 0.74-0.91) respectively, with 70% and 93% efficiency gain compared to default estimator based on bootstrap variance. While the model-based procedure appears to be more efficient compared to the bias-penalized procedure for estimating PPV(60), the corresponding estimators are further away from the default estimators as expected.

Finally, to illustrate application of our methodology to a case-control design, we generated a case-control sample from the PCA3 data. Results are shown in the online Supplementary Material. Again, substantial efficiency gain could potentially be achieved through weighting.

6. Concluding Remarks

In this paper we proposed more efficient estimators for population-specific PPV and NPV, when samples are available from both the target population and an auxiliary population which share similar classification accuracy as measured by particular points on the ROC curve. Note that even if accuracy of the marker might depend on other variables, which are distributed differently across populations, our method will still work as long as the marginal classification accuracy is similar between the two populations. Our proposed estimators assign weights to samples from each population. We propose two methods for weight selection to maximize estimation efficiency. The one based on asymptotic variance formula and normality assumption is easy to implement and more efficient when the assumptions hold exactly. The bias-penalized bootstrap method for weight selection provides a more robust alternative against possible violation of the common classification accuracy assumption, although it does lose quite a bit of efficiency relative to the correctly specified model based procedure.

In theory, the common classification accuracy assumption holds in the following scenario. Suppose cases and controls in the auxiliary population after some monotone transformation g, follow the same distributions as cases and controls in the target population, then $S^*_{\bar{D}}\left(Y^*_D\right)=P\left(Y^*_{\bar{D}} \geq Y^*_D\right)=P\left\{g\left(Y^*_{\bar{D}}\right) \geq g\left(Y^*_D\right)\right\}=P\left(Y_{\bar{D}} \geq Y_D\right)=S_{\bar{D}}\left(Y_D\right)$, which implies the equivalence between the ROC curves. This holds because $P\left\{S_{\bar{D}}\left(Y_D\right) \leq t\right\}=P\left\{Y_D \geq S^{-1}_{\bar{D}}\left(t\right)\right\}=\mathrm{ROC}\left(t\right)$, i.e., ROC is the cdf of $S_{\bar{D}}(Y_D)$, the 'placement' of Y_D among the control distribution (Pepe and Cai, 2004). Here the population indicator is a confounder in evaluating classification accuracy of the marker; the threshold of marker value to achieve a given specificity is different across populations but the sensitivity corresponding to a given specificity remains the same (Janes and Pepe, 2008a,b). Our methods provide a way to adjust for the confounding effect of population with a goal of estimating population-specific predictive values. In practice, whether classification accuracy of a biomarker is similar across populations can be explored using the data. And we can further conduct tests for equal classification accuracy as we did in the PCA3 example. This is analogous to a test of the interaction between marker and covariate in a standard regression setting to rule out the possibility that the covariate (in our setting the population indicator) would affect the marker's discriminatory performance. We should also work closely with scientists to decide whether a reasonable true difference in ROC curves would lead to intolerable bias.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful for support provided by NCI grants CA86368. And we thank the editor, AE, and referees for their helpful comments and suggestions.

Appendix

Proofs of all results that are not given explicitly in the text are available in the supplementary material.

Appendix A: Asymptotic variance of the weighted PPV estimators

Here we present asymptotic theory for the proposed estimator defined in Sections 2.1 and 2.2. We assume the following conditions hold:

i. the distribution functions of Y_D, $Y_{\bar{D}}$, Y_D^*, and $Y_{\bar{D}}^*$ are differentiable with density functions f_D, $f_{\bar{D}}$, f_D^*, and $f_{\bar{D}}^*$ respectively,

ii. as $n_{\bar{D}} \to \infty, n_D/n_{\bar{D}} \to \lambda, n_{\bar{D}}^*/n_{\bar{D}} \to \lambda_1$, and $n_D^*/n_D \to \lambda_2$. This implies $\left(n_D^*+n_{\bar{D}}^*\right)/\left(n_D+n_{\bar{D}}\right) \to (\lambda_1+\lambda\lambda_2)/(1+\lambda)$, $n_D/\left(n_D+n_{\bar{D}}\right) \to \lambda/(1+\lambda)$, and $n_D^*/\left(n_D^*+n_{\bar{D}}^*\right) \to \lambda\lambda_2/(\lambda\lambda_2+\lambda_1)$, i.e. the ratio of the sample sizes from the two populations converges to a constant, and the proportion of diseased in each population converges to a population-specific constant.

Consistency of $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}}.A_w(y)$ follow from the Continuous Mapping Theorem.

Theorem 1

$\sqrt{n_{\bar{D}}}\left\{\widehat{\mathrm{PPV}}_w(y) - \mathrm{PPV}(y)\right\}$ is asymptotically normally distributed with mean zero and variance

$$\begin{aligned}\Sigma_w &= A_{11}V_{\bar{D}}(y) + A_{12}\frac{f_D(y)}{f_{\bar{D}}(y)}(1-w)V_{\bar{D}}(y) \\ &\quad + A_{22}\left[\left\{(1-w)^2\left(1+\frac{1}{\lambda_1}\right)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 V_{\bar{D}}(y)\right\} + \frac{1}{\lambda}\left\{w^2+(1-w)^2\frac{1}{\lambda_2}\right\}V_D(y)\right],\end{aligned} \tag{4}$$

where $V_{\bar{D}}(y) = S_{\bar{D}}(y)\{1 - S_{\bar{D}}(y)\}$, $V_D(y) = S_D(y)\{1 - S_D(y)\}$,

$$A_{11} = \left[\frac{\rho(1-\rho)}{\left\{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)\right\}^2}\right]^2 S_D(y)^2,$$

$$A_{12} = -2\left[\frac{\rho(1-\rho)}{\left\{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)\right\}^2}\right]^2 S_D(y)\,S_{\bar{D}}(y),$$

$$A_{22} = \left[\frac{\rho(1-\rho)}{\left\{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)\right\}^2}\right]^2 S_{\bar{D}}(y)^2.$$

Note that when $w = 1$, Σ_w reduces to $A_{11}V_{\bar{D}}(y) + A_{22}V_D(y)/\lambda$, the asymptotic variance of the default estimator $\widetilde{\mathrm{PPV}}(y)$.

Observe that Σ_w is a quadratic function of w, which is convex since $A_{22} > 0$. In addition, Σ_w can be written as the product of $[\rho(1-\rho)/\{\rho S_D(y) + (1-\rho)S_{\bar{D}}(y)\}^2]^2$ and another term that is free of ρ.

Theorem 2

Asymptotic variance of $\widehat{\mathrm{PPV}}_w(y)$ is minimized when

$$w=\frac{A_{12}\frac{f_D(y)}{f_{\bar{D}}(y)}V_{\bar{D}}(y)+2A_{22}\left(1+\frac{1}{\lambda_1}\right)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 V_{\bar{D}}(y)+2A_{22}\frac{1}{\lambda_2}\frac{1}{\lambda}V_D(y)}{2A_{22}\left(1+\frac{1}{\lambda_1}\right)\left\{\frac{f_D(y)}{f_{\bar{D}}(y)}\right\}^2 V_{\bar{D}}(y)+2A_{22}\frac{1}{\lambda_2}\frac{1}{\lambda}V_D(y)+2A_{22}\frac{1}{\lambda}V_D(y)}. \quad (5)$$

Since $A_{12} < 0$, the optimal w is always less than 1.

Theorem 3

$\sqrt{n_{\bar{D}}}\left\{\widehat{\mathrm{PPV}.A_w}(y)-\mathrm{PPV}(y)\right\}$ is asymptotically normally distributed with mean zero and variance

$$\begin{aligned}\Sigma.A_w \quad =&A_{11}\left[\left\{(1-w)^2\left(1+\frac{1}{\lambda_2}\right)\left\{\frac{f_{\bar{D}}(y)}{f_D(y)}\right\}^2\frac{1}{\lambda}V_D(y)\right\}+\left\{w^2+(1-w)^2\frac{1}{\lambda_1}\right\}V_{\bar{D}}(y)\right],\\ &+A_{12}\frac{f_{\bar{D}}(y)}{f_D(y)}(1-w)\frac{1}{\lambda}V_D(y)+A_{22}\frac{1}{\lambda}V_D(y),\end{aligned} \quad (6)$$

Theorem 4

Asymptotic variance of $\widehat{\mathrm{PPV}.A_w}(y)$ is minimized when

$$w=\frac{A_{12}\frac{f_{\bar{D}}(y)}{f_D(y)}\frac{1}{\lambda}V_D(y)+2A_{11}\left(1+\frac{1}{\lambda_2}\right)\left\{\frac{f_{\bar{D}}(y)}{f_D(y)}\right\}^2\frac{1}{\lambda}V_D(y)+2A_{11}\frac{1}{\lambda_1}V_{\bar{D}}(y)}{2A_{11}\left(1+\frac{1}{\lambda_2}\right)\left\{\frac{f_{\bar{D}}(y)}{f_D(y)}\right\}^2\frac{1}{\lambda}V_D(y)+2A_{11}\frac{1}{\lambda_1}V_{\bar{D}}(y)+2A_{11}V_{\bar{D}}(y)}. \quad (7)$$

The optimal w is always less than 1.

Theorem 5

Suppose we use sample prevalence $\hat{\rho}$ derived from a pilot cohort study with sample size n_c, such that var $(\hat{\rho}) = \sigma^2/n_c$, and suppose $n_c/n_{\bar{D}} \to \xi$ as $n_{\bar{D}} \to \infty$. Then compared to known ρ, the asymptotic variance of $\sqrt{n_{\bar{D}}}\left\{\widehat{\mathrm{PPV}}_w(y)-\mathrm{PPV}(y)\right\}$ as $n_{\bar{D}} \to \infty$ increases by a term

$$\frac{\sigma^2}{\xi}\frac{S_D(y)^2 S_{\bar{D}}(y)^2}{\left\{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)\right\}^4}.$$

Same for the asymptotic variance of $\widehat{\mathrm{PPV}.A_w}(y)$.

Appendix B: Asymptotic Bias of the Weighted PPV estimators

Theorems 6 and 7 present the asymptotic bias of $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}.A_w}$ as a function of the difference in sensitivity between the two populations with specificity fixed at $1 - S_{\bar{D}}(y)$ and the difference in specificity between the two populations with sensitivity fixed at $S_D(y)$. The derivation is presented in the supplementary material.

Theorem 6

Let $\delta = \mathrm{ROC}^*\{t\} - \mathrm{ROC}(t)$ for $t = S_{\bar{D}}(y)$. The asymptotic bias of $\widehat{\mathrm{PPV}}_w(y)$ is monotonically increasing in $(1-w)\delta$, and equals

$$\frac{\rho(1-\rho)S_{\bar{D}}(y)}{\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}\rho+S_{\bar{D}}(y)\times(1-\rho)}\times\frac{(1-w)\delta}{\rho(1-w)\delta+\rho\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}+S_{\bar{D}}(y)(1-\rho)}. \tag{8}$$

On the other hand, to cause an asymptotic bias r (such that $|r|$ is smaller than or equal to the maximum possible asymptotic bias that can be achieved) in terms of PPV, according to (8), we have

$$\delta=\frac{r}{1-w}\frac{\rho\mathrm{ROC}\left\{S_{\bar{D}}(y)\right\}+(1-\rho)S_{\bar{D}}(y)}{C^+-\rho r}, \tag{9}$$

where

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

Theorem 7

Let $\eta=-\left[S^*_{\bar{D}}S^{*-1}_{D}\{S_D(y)\}-S_{\bar{D}}S^{-1}_{D}\{S_D(y)\}\right]$, the asymptotic bias of $\widehat{\mathrm{PPV}}.A_w(y)$ equals

$$\frac{\rho(1-\rho)S_D(y)}{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)}\times\frac{(1-w)\eta}{-(1-\rho)(1-w)\eta+\rho S_D(y)+S_{\bar{D}}(y)(1-\rho)}. \tag{10}$$

On the other hand, to cause an asymptotic bias r (such that $|r|$ is smaller than or equal to the maximum possible asymptotic bias that can be achieved) in terms of PPV, according to (10), we have

$$\eta=\frac{r}{1-w}\frac{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)}{C^-+(1-\rho)r}, \tag{11}$$

where

$$C^-=\frac{\rho(1-\rho)S_D(y)}{\rho S_D(y)+(1-\rho)S_{\bar{D}}(y)}.$$

Appendix C: Proof for cross-sectional or cohort study

Suppose we randomly sample n observations Y, D from the target population. Calculating $\widehat{\rho}=\Sigma_{i=1}^{n}D_i/n$, and

$$\widehat{S}_{D}(y)=\frac{\Sigma_{i=1}^{n} I(Y_i>y)\,D_i}{\sum_{i=1}^{n} D_i},\ \widehat{S}_{\bar{D}}(y)=\frac{\Sigma_{i=1}^{n} I(Y_i>y)\,(1-D_i)}{\Sigma_{i=1}^{n} 1-D_i}.$$

Let $\mathbf{D} = (D_1, D_2, \ldots, D_n)$, then

$$\begin{aligned} & \operatorname{cov}\left\{\widehat{S}_{D}(y),\widehat{\rho}\right\} \\ = \ & \operatorname{cov}\left[E\left\{\frac{\Sigma_{i=1}^{n} I(Y_i>y)D_i}{\Sigma_{i=1}^{n} D_i}\middle|\mathbf{D}\right\}, E\left(\frac{1}{n}\sum_{i=1}^{n} D_i\middle|\mathbf{D}\right)\right]+E\left[\operatorname{cov}\left\{\frac{\Sigma_{i=1}^{n} I(Y_i>y)D_i}{\Sigma_{i=1}^{n} D_i}, \frac{1}{n}\sum_{i=1}^{n} D_i\middle|\mathbf{D}\right\}\right] \\ =^{*} \ & \operatorname{cov}\left(S_{D}(y), \sum_{i=1}^{n} D_i\right)+E(0) \\ = \ & 0+0=0, \end{aligned}$$

where holds since

$$\begin{aligned} & E\left\{\frac{\Sigma_{i=1}^{n} I(Y_i>y)D_i}{\Sigma_{i=1}^{n} D_i}\middle|\mathbf{D}\right\}=\frac{1}{\Sigma_{i=1}^{n} D_i}E\{I(Y_i>y)\,|D_i\} \\ = \ & \frac{\Sigma_{i=1} D_i}{\Sigma_{i=1}^{n}\left[D_i\left\{S_{D}(y)D_i+S_{\bar{D}}(y)(1-D_i)\right\}\right]} \\ = \ & \frac{\Sigma_{i=1} D_i}{\Sigma_i S_{D}(y)D_i}=S_{D}(y). \end{aligned}$$

References

Deras IL, Aubin SMJ, Blase A, Day JR, Koo S, Partin AW, Ellis WJ, Marks LS, Fradet Y, Rittenhouse H, Groskopf J. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2006; 179:1587–1592. [PubMed: 18295257]

Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. JASA. 1937; 32(200):675–701.

Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. American Journal of Epidemiology. 2008 In Press.

Janes H, Pepe MS. Adjusting for covariate effects on classification accuracy using the covariate-adjusted ROC curve. Biometrika. In Press.

Leisenring W, Alonzo TA, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. Biometrics. 2000; 56:345–351. [PubMed: 10877288]

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. JNCI. 1959; 22(4):719–748. [PubMed: 13655060]

Moskowitz CS, Pepe MS. Quantifying and comparing the predictive accuracy of continuous prognostic factors for binary outcomes. Biostatistics. 2004; 5:113–127. [PubMed: 14744831]

Moskowitz CS, Pepe MS. Comparing the predictive values of diagnostic tests: sample size and analysis for paired study designs. Clinical Trials. 2006; 3(3):272–279. [PubMed: 16895044]

Pepe, MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press; Oxford, United Kingdom: 2003.

Pepe MS, Cai T. The analysis of placement values for evaluating discriminatory measures. Biometrics. 2004; 60:528–535. [PubMed: 15180681]

Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. Journal of the National Cancer Institute. 2001; 93(14):1054–1061. [PubMed: 11459866]

Quade D. Using weighted rankings in the analysis of complete blocks with additive block effects. JASA. 74(367):680–683.

Steinberg DM, Fine J, Chappell R. Sample size for positive and negative predictive value in diagnostic research. Biostatistics. 2008 In Press.

van der Vaart, AW.; Wellner, JA. Weak convergence and empirical processes. Springer-Verlag; New York: 1996.

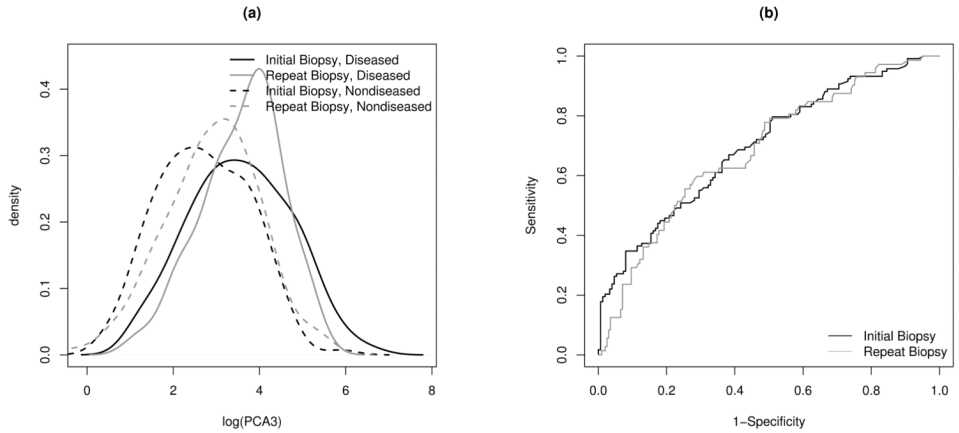

Fig. 1.

(a) Distribution of log(PCA3) conditional on disease status within the initial and repeat biopsy populations based on the pilot cohort study, (b) Empirical ROC curves for PCA3 within the initial and the repeat biopsy populations based on the pilot cohort study.

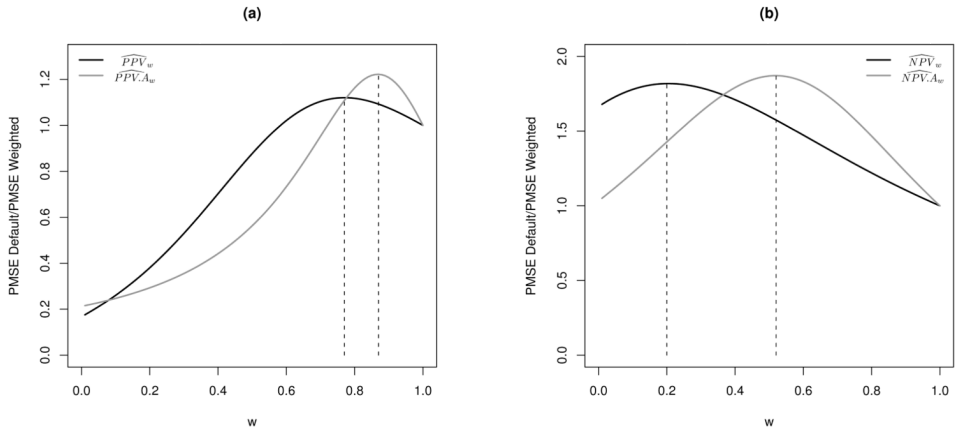

Fig. 2.
Ratio of PMSE for the default estimator versus the weighted estimator of (a) PPV(60) and (b) NPV(20) as functions of weight.

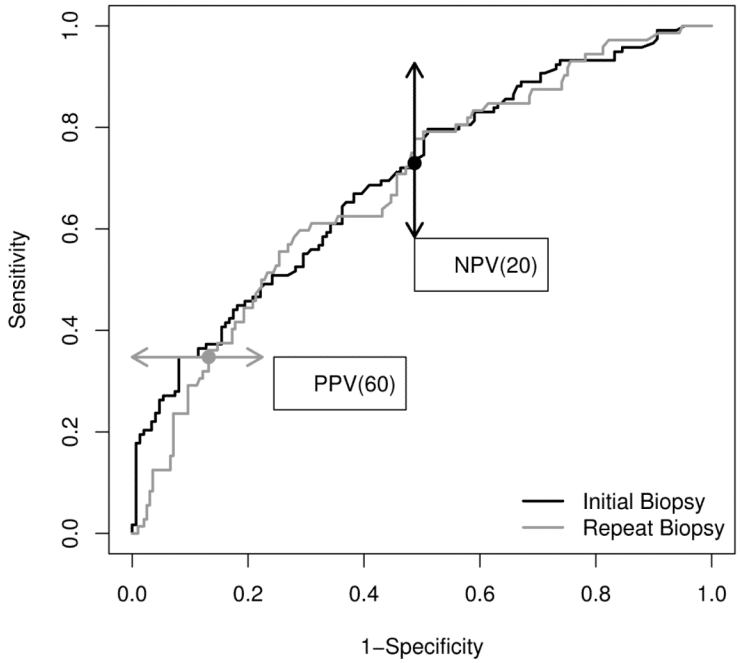

Fig. 3.

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.

Table 1

Performance of $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}}.A_w$ for fixed ρ = 0.4. Here PPV = 0.722, asymptotically optimal w is 0.249 for $\widehat{\mathrm{PPV}}_w(y)$ and 0.578 for $\widehat{\mathrm{PPV}}.A_w(y)$. Efficiency of the weighted estimator relative to $\widetilde{\mathrm{PPV}}$ is defined to be the ratio of variance for $\widetilde{\mathrm{PPV}}$ to variance for the weighted estimator. Wald CI based on bootstrap variance estimate is constructed assuming normality of the logit-transformed estimator. Here $n_D = n_{\bar{D}} = n_D^* = n_{\bar{D}}^* = n/2$.

	w=0.1	w=0.2	w=0.3	w=0.4	w=0.5	w=0.6	w=0.7	w=0.8	w=0.9
Bias									
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	−0.002	−0.0005	0.0003	0.0009	0.001	0.001	0.001	0.001	0.0006
n=500	−0.0003	0.0002	0.0006	0.0009	0.001	0.001	0.001	0.001	0.0009
$\widehat{\mathrm{PPV}}.A_w(y)$									
n=250	0.002	0.0004	−0.0005	−0.001	−0.002	−0.002	−0.002	−0.001	−0.0008
n=500	0.001	0.0008	0.0004	0.0001	−0.0002	−0.0003	−0.0003	−0.0001	0.0002
Variance : $\frac{\text{Asymptotic - Observed}}{\text{Observed}} \times 100\ \%$									
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	−5.28	−3.83	−3.83	−3.31	−3.50	−3.58	−3.25	−2.62	−2.34
n=500	−2.51	−2.16	−1.74	−1.27	−1.20	−1.47	−1.00	−0.60	−0.79
$\widehat{\mathrm{PPV}}.A_w(y)$									
n=250	−0.097	−0.72	−0.83	−0.48	0.18	0.82	1.23	1.41	1.56
n=500	1.58	1.06	0.73	0.57	0.52	0.49	0.44	0.40	0.44
Efficiency relative to $\widetilde{\mathrm{PPV}}$									
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	1.77	1.85	1.85	1.79	1.69	1.56	1.41	1.27	1.13
n=500	1.79	1.86	1.87	1.81	1.71	1.57	1.43	1.28	1.13
$\widehat{\mathrm{PPV}}.A_w(y)$									
n=250	0.90	1.13	1.40	1.67	1.87	1.91	1.77	1.52	1.25
n=500	0.87	1.10	1.36	1.63	1.84	1.89	1.76	1.52	1.25
Coverage of 95% CI									

J R Stat Soc Ser C Appl Stat. Author manuscript; available in PMC 2011 November 1.

	w=0.1	w=0.2	w=0.3	w=0.4	w=0.5	w=0.6	w=0.7	w=0.8	w=0.9
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	96.9	96.5	96.4	95.9	95.9	95.7	95.6	95.4	95.5
n=500	96.1	95.8	95.6	95.3	95.3	95.4	95.1	95.0	95.0
$\widehat{\mathrm{PPV}}_w A_w(y)$									
n=250	95.4	95.3	95.4	95.5	95.6	95.6	95.6	95.2	95.2
n=500	95.1	95.0	95.3	95.6	95.6	95.3	94.9	94.6	95.0

Table 2

Performance of $\widehat{\mathrm{PPV}}_w$ and $\widehat{\mathrm{PPV}}.A_w$ with $\hat{\rho}$ estimated for ρ = 0.4. Here PPV = 0.722, asymptotically optimal w is 0.249 for $\widehat{\mathrm{PPV}}_w(y)$ and 0.578 for $\widehat{\mathrm{PPV}}.A_w(y)$. Wald CI based on bootstrap variance estimate is constructed assuming normality of the logit-transformed estimator. Here $n_D = n_{\bar{D}} = n_D^* = n_{\bar{D}}^* = n/2$.

	w=0.1	w=0.2	w=0.3	w=0.4	w=0.5	w=0.6	w=0.7	w=0.8	w=0.9
Bias									
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	−0.0004	0.0008	0.002	0.002	0.003	0.003	0.003	0.003	0.002
n=500	−0.001	−0.0006	−4e-5	0.0004	0.0006	0.0008	0.001	0.001	0.001
$\widehat{\mathrm{PPV}.A}_w(y)$									
n=250	−0.002	0.0008	3e-5	−0.0005	−0.0007	−0.0007	−0.0004	6e-5	0.0003
n=500	−0.002	−0.002	−0.002	−0.002	−0.001	−0.001	−0.0009	−0.0003	0.0003
Variance : $\frac{\text{Asymptotic - Observed}}{\text{Observed}} \times 100\ \%$									
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	4.39	3.63	3.48	3.48	3.37	3.03	2.48	1.80	1.12
n=500	−2.97	−3.29	−3.34	−3.31	−3.33	−3.46	−3.71	−4.05	−4.42
$\widehat{\mathrm{PPV}.A}_w(y)$									
n=250	1.61	1.81	2.33	2.93	3.28	3.14	2.48	1.60	0.87
n=500	−3.18	−3.01	−2.76	−2.57	−2.59	−2.92	−3.49	−4.11	−4.57
Efficiency relative to $\widetilde{\mathrm{PPV}}$									
$\widehat{\mathrm{PPV}}_w(y)$									
n=250	1.55	1.59	1.60	1.56	1.49	1.41	1.31	1.21	1.10
n=500	1.58	1.62	1.62	1.58	1.51	1.42	1.32	1.22	1.10
$\widehat{\mathrm{PPV}.A}_w(y)$									
n=250	0.89	1.07	1.26	1.44	1.57	1.60	1.53	1.38	1.19
n=500	0.88	1.06	1.26	1.44	1.58	1.61	1.54	1.39	1.19
Coverage of 95% CI									
$\widehat{\mathrm{PPV}}_w(y)$									

J R Stat Soc Ser C Appl Stat. Author manuscript; available in PMC 2011 November 1.

	w=0.1	w=0.2	w=0.3	w=0.4	w=0.5	w=0.6	w=0.7	w=0.8	w=0.9
n=250	96.5	96.3	96.1	96.0	95.8	95.9	95.8	95.8	95.7
n=500	96.3	96.0	95.9	95.8	95.9	96.0	95.9	96.0	96.0
$\widehat{\mathrm{PPV}}.A_w(y)$									
n=250	95.7	95.3	95.1	95.1	95.0	95.4	95.3	95.2	95.2
n=500	96.1	96.1	96.4	96.0	96.0	96.1	96.1	95.8	95.8

Table 3

Relative efficiency of $\widehat{\mathrm{PPV}}_w(y)$ or $\widehat{\mathrm{PPV}}.A_w(y)$ versus $\widehat{PPV}(y)$ for varying ρ and specificity $v = F_{\bar{D}}(y)$, assuming ρ is fixed (i.e. $\mathrm{var}\left\{\widehat{\mathrm{PPV}}(y)\right\}/\mathrm{var}\left\{\widehat{\mathrm{PPV}}_w(y)\right\}$ or $\mathrm{var}\left\{\widehat{\mathrm{PPV}}(y)\right\}/\mathrm{var}\left\{\widehat{\mathrm{PPV}}.A_w(y)\right\}$. The weight w is selected using the asymptotic formula assuming normal model or based on the bootstrap procedure to minimize PMSE. Note that asymptotically, the efficiency of the weighted estimators with optimal weight relative to $\widetilde{\mathrm{PPV}}$ is 1.31, 1.52, 1.66, 1.78, 1.89 respectively for v = 0.1, 0.3, 0.5, 0.7, 0.9.

			v = 0.1	**v = 0.3**	**v = 0.5**	**v = 0.7**	**v = 0.9**
Disease Prevalence	Weight Selection		Value of PPV(y)				
ρ = 0.1			0.109	0.129	0.158	0.202	0.302
ρ = 0.3			0.320	0.364	0.419	0.494	0.625
ρ = 0.5			0.523	0.572	0.627	0.695	0.796
ρ = 0.7			0.719	0.757	0.797	0.842	0.901
ρ = 0.9			0.908	0.923	0.938	0.953	0.972
			Efficiency relative to $\widetilde{\mathrm{PPV}}(y)$				
ρ = 0.1	Normal	$\widehat{\mathrm{PPV}}_w(y)$	1.33	1.45	1.72	1.75	1.99
		$\widehat{\mathrm{PPV}}.A_w(y)$	1.39	1.48	1.73	1.74	2.02
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	1.15	1.20	1.30	1.32	1.36
		$\widehat{\mathrm{PPV}}.A_w(y)$	1.18	1.22	1.32	1.32	1.35
ρ = 0.3	Normal	$\widehat{\mathrm{PPV}}_w(y)$	1.29	1.59	1.71	1.82	1.86
		$\widehat{\mathrm{PPV}}.A_w(y)$	1.28	1.60	1.69	1.79	1.86
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	1.13	1.25	1.28	1.27	1.29
		$\widehat{\mathrm{PPV}}.A_w(y)$	1.14	1.26	1.27	1.27	1.30
ρ = 0.5	Normal	$\widehat{\mathrm{PPV}}_w(y)$	1.28	1.55	1.58	1.84	1.89
		$\widehat{\mathrm{PPV}}.A_w(y)$	1.33	1.58	1.55	1.83	1.92
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	1.15	1.21	1.23	1.28	1.32
		$\widehat{\mathrm{PPV}}.A_w(y)$	1.18	1.22	1.23	1.29	1.33
ρ = 0.7	Normal	$\widehat{\mathrm{PPV}}_w(y)$	1.31	1.60	1.69	1.71	1.96

J R Stat Soc Ser C Appl Stat. Author manuscript; available in PMC 2011 November 1.

			$\nu = 0.1$	$\nu = 0.3$	$\nu = 0.5$	$\nu = 0.7$	$\nu = 0.9$
		$\widehat{\text{PPV}.A_w(y)}$	1.33	1.57	1.66	1.79	2.03
	Bootstrap	$\widehat{\text{PPV}_w(y)}$	1.15	1.23	1.29	1.29	1.31
		$\widehat{\text{PPV}.A_w(y)}$	1.15	1.23	1.29	1.30	1.35
$\rho = 0.9$	Normal	$\widehat{\text{PPV}_w(y)}$	1.31	1.50	1.59	1.79	1.93
		$\widehat{\text{PPV}.A_w(y)}$	1.35	1.48	1.56	1.82	1.91
	Bootstrap	$\widehat{\text{PPV}_w(y)}$	1.14	1.21	1.26	1.31	1.31
		$\widehat{\text{PPV}.A_w(y)}$	1.18	1.21	1.25	1.31	1.33

Table 4

Coverage of 95% logit-transformed Wald CI using bootstrap variance estimate of $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}.A}_w(y)$ for varying ρ and specificity $v = F_{\bar{D}}(y)$, assuming ρ is fixed. The weight w is selected using the asymptotic formula assuming normal model or based on the bootstrap procedure to minimize PMSE. Note that asymptotically, the efficiency of the weighted estimators with optimal weight relative to $\widetilde{\mathrm{PPV}}$ is 1.31, 1.52, 1.66, 1.78, 1.89 respectively for v = 0.1, 0.3, 0.5, 0.7, 0.9.

	Weight Selection		***v* = 0.1**	***v* = 0.3**	***v* = 0.5**	***v* = 0.7**	***v* = 0.9**
$\rho = 0.1$	Normal	$\widehat{\mathrm{PPV}}_w(y)$	94.20	94.90	96.00	94.50	93.80
		$\widehat{\mathrm{PPV}.A}_w(y)$	94.10	96.00	96.40	94.10	93.30
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	92.60	92.60	94.60	93.50	94.20
		$\widehat{\mathrm{PPV}.A}_w(y)$	92.00	92.70	94.70	93.20	92.90
$\rho = 0.3$	Normal	$\widehat{\mathrm{PPV}}_w(y)$	92.23	95.50	94.60	93.80	94.70
		$\widehat{\mathrm{PPV}.A}_w(y)$	92.52	95.00	94.30	94.40	94.70
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	91.70	94.70	92.90	92.30	93.10
		$\widehat{\mathrm{PPV}.A}_w(y)$	91.00	94.30	92.50	91.70	92.70
$\rho = 0.5$	Normal	$\widehat{\mathrm{PPV}}_w(y)$	95.00	95.10	94.00	95.20	95.90
		$\widehat{\mathrm{PPV}.A}_w(y)$	95.50	95.90	94.10	94.60	95.90
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	94.40	93.00	92.90	93.70	95.20
		$\widehat{\mathrm{PPV}.A}_w(y)$	94.60	92.60	92.90	93.60	94.60
$\rho = 0.7$	Normal	$\widehat{\mathrm{PPV}}_w(y)$	93.79	95.10	94.70	95.70	94.90
		$\widehat{\mathrm{PPV}.A}_w(y)$	94.35	96.30	94.10	96.20	95.60
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	93.10	94.20	92.90	94.60	93.80
		$\widehat{\mathrm{PPV}.A}_w(y)$	92.80	94.30	92.70	94.50	93.80
$\rho = 0.9$	Normal	$\widehat{\mathrm{PPV}}_w(y)$	93.8	94.70	95.80	94.60	94.30
		$\widehat{\mathrm{PPV}.A}_w(y)$	94.3	95.20	95.90	95.60	93.90
	Bootstrap	$\widehat{\mathrm{PPV}}_w(y)$	92.80	94.00	94.80	93.90	93.30

J R Stat Soc Ser C Appl Stat. Author manuscript; available in PMC 2011 November 1.

Weight Selection	$\widehat{\mathrm{PPV}.A_w(y)}$	$\nu = 0.1$	$\nu = 0.3$	$\nu = 0.5$	$\nu = 0.7$	$\nu = 0.9$
		92.00	93.80	95.30	94.30	92.50

Table 5

Relative bias of $\widehat{\mathrm{PPV}}_w(y)$ and $\widehat{\mathrm{PPV}.A}_w(y)$ for varying ρ and $v = F_{\bar{D}}(y)$. The weight w is selected using the asymptotic formula assuming normal model or based on the bootstrap procedure to minimize PMSE.

			$v = 0.1$	$v = 0.3$	$v = 0.5$	$v = 0.7$	$v = 0.9$
$\rho = 0.1$		PPV(y)	0.109	0.129	0.157	0.202	0.302
	Normal	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.44	2.56	5.25	9.13	15.78
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.94	3.71	6.67	10.22	15.52
	Bootstrap	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.09	0.78	1.21	1.98	4.53
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.06	0.73	1.10	1.72	4.00
$\rho = 0.3$		PPV(y)	0.320	0.364	0.419	0.494	0.625
	Normal	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.40	1.76	3.53	5.61	7.36
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.79	2.65	4.51	6.27	7.07
	Bootstrap	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.14	0.47	1.05	1.35	1.92
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.11	0.42	0.97	1.19	1.64
$\rho = 0.5$		PPV(y)	0.523	0.572	0.627	0.695	0.796
	Normal	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.25	1.19	2.20	3.27	3.94
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.50	1.75	2.75	3.64	3.82
	Bootstrap	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.08	0.35	0.61	0.94	1.17
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.05	0.33	0.56	0.87	1.06
$\rho = 0.7$		PPV(y)	0.719	0.757	0.797	0.842	0.901
	Normal	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.18	0.70	1.23	1.59	1.83
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.34	1.00	1.55	1.82	1.77
	Bootstrap	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.08	0.25	0.39	0.36	0.54
		% Bias of $\widehat{\mathrm{PPV}.A}_w(y)$	0.07	0.23	0.37	0.33	0.50
$\rho = 0.9$		PPV(y)	0.908	0.923	0.938	0.953	0.972
	Normal	% Bias of $\widehat{\mathrm{PPV}}_w(y)$	0.06	0.22	0.36	0.49	0.51

J R Stat Soc Ser C Appl Stat. Author manuscript; available in PMC 2011 November 1.

		$\nu = 0.1$	$\nu = 0.3$	$\nu = 0.5$	$\nu = 0.7$	$\nu = 0.9$
Bootstrap	% Bias of $\widehat{\text{PPV}}_w(y)$	0.11	0.31	0.46	0.56	0.50
	% Bias of $\widehat{\text{PPV}}_w(y)$	0.03	0.08	0.10	0.15	0.16
	% Bias of $\widehat{\widehat{\text{PPV}}_w(y)}$	0.03	0.08	0.10	0.14	0.15

Table 6

Comparison of the two strategies for estimating PPV and NPV. Here Bias* is the difference between the weighted estimate and the default estimate; Efficiency† is the ratio of variance of the default estimator ($\widetilde{\mathrm{PPV}}$ or $\widetilde{\mathrm{NPV}}$) vs variance of the weighted estimator; Efficiency* is the ratio of PMSE of the default estimator ($\widetilde{\mathrm{PPV}}$ or $\widetilde{\mathrm{NPV}}$) to PMSE of the weighted estimator.

	$\widetilde{\mathrm{PPV}}(60)$	$\widetilde{\mathrm{PPV}}.A_w(60)$	$\widetilde{\mathrm{NPV}}(20)$	$\widetilde{\mathrm{NPV}}.A_w(20)$
Weight	1	0.87	1	0.52
Est (95% CI)	0.77 (0.62, 0.88)	0.76 (0.63, 0.85)	0.86 (0.69, 0.94)	0.85 (0.74, 0.92)
Bias*	0.0012	−0.020	0.004	−0.008
Variance	0.0044	0.0032	0.0037	0.0019
PMSE	0.0044	0.0036	0.0038	0.0020
Efficiency†	1.00	1.38	1.00	1.93
Efficiency*	1.00	1.22	1.00	1.87