



UW Biostatistics Working Paper Series

2-10-2009

Measures to Summarize and Compare the Predictive Capacity of Markers

Wen Gu

University of Washington - Seattle Campus, wengu@u.washington.edu

Margaret Pepe

University of Washington, mspepe@u.washington.edu

Suggested Citation

Gu, Wen and Pepe, Margaret, "Measures to Summarize and Compare the Predictive Capacity of Markers" (February 2009). *UW Biostatistics Working Paper Series*. Working Paper 342.
<http://biostats.bepress.com/uwbiostat/paper342>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Measures to Summarize and Compare the Predictive Capacity of Markers

Wen Gu[†], Margaret Sullivan Pepe

Department of Biostatistics, University of Washington, Box 357232

1705 Northeast Pacific Street, Seattle, WA 98195

and

Fred Hutchinson Cancer Research Center Public Health Sciences

1100 Fairview Avenue N., M2-B500, Seattle, WA 98109-1024

[†]*Corresponding author's address:* wengu@u.washington.edu

Summary. The predictive capacity of a marker in a population can be described using the population distribution of risk (Huang *et al.*, 2007; Pepe *et al.*, 2008_a; Stern, 2008). Virtually all standard statistical summaries of predictability and discrimination can be derived from it (Gail and Pfeiffer, 2005). The goal of this paper is to develop methods for

making inference about risk prediction markers using summary measures derived from the risk distribution. We describe some new clinically motivated summary measures and give new interpretations to some existing statistical measures. Methods for estimating these summary measures are described along with distribution theory that facilitates construction of confidence intervals from data. We show how markers and, more generally, how risk prediction models, can be compared using clinically relevant measures of predictability. The methods are illustrated by application to markers of lung function and nutritional status for predicting subsequent onset of major pulmonary infection in children suffering from cystic fibrosis. Simulation studies show that methods for inference are valid for use in practice.

Keywords: Classification; Diagnosis; Prediction; Prognosis; Risk models

1 Background

Let D denote a binary outcome variable, such as presence of disease or occurrence of an event within a specified time period and let Y denote a set of predictive markers used to predict a bad outcome, $D = 1$, or a good outcome, $D = 0$. For example, elements of the Framingham risk score (age, gender, total and high-density lipoprotein cholesterol, systolic blood pressure, treatment for hypertension and smoking) are used to predict occurrence of a

cardiovascular event within 10 years (<http://hp2010.nhlbihin.net/atpiii/calculator.asp>). We write the risk associated with marker value $Y = y$ as $risk(y) = P[D = 1|Y = y]$.

Huang *et al.* (2007) proposed the predictiveness curve to describe the predictive capacity of Y . It displays the population distribution of risk via the risk quantiles, $R(\nu)$ versus ν , where

$$P[risk(Y) \leq R(\nu)] = \nu.$$

The inverse of the predictiveness curve is simply the cumulative distribution function (cdf) of $risk(Y)$

$$R^{-1}(p) = P[risk(Y) \leq p] = F_{risk}(p)$$

and correspondingly

$$R(\nu) = F_{risk}^{-1}(\nu).$$

Gail and Pfeiffer(2005) noted that standard statistical measures used to quantify the predictive capacity of a risk prediction model can be calculated from the risk distribution function, $F_{risk}(p)$. These include measures derived from the receiver operating characteristic (ROC) curve and the Lorenz curve, predictive values, misclassification rates, and measures of explained variation. Bura and Gastwirth (2001) used the risk quantiles, $R(\nu)$, to assess predictors in binary regression models. They proposed a summary index which they called the total gain.

Summary indices are often used to compare prediction models. The area under the ROC

curve is widely used in practice for this purpose. However there is controversy about its use, particularly in the cardiovascular research community (Cook, 2007; Pencina *et al.*, 2008). This has motivated another approach to evaluating risk prediction markers that relies on defining categories of risk that are clinically meaningful. Several summary indices based on this notion have been proposed. The reclassification percent and the net reclassification index (NRI) are such summary measures derived from reclassification tables and they have recently gained popularity in the applied literature (Ridker *et al.*, 2008; D'Agostino *et al.*, 2008).

In this paper, we explicitly relate existing and new summary measures of prediction to the risk distribution, i.e. to the predictiveness curve. We contrast them qualitatively, paying particular attention to their clinical interpretations and relevance. We then derive distribution theory that can be used for making statistical inference. Note that rigorous methods for inference have not been available heretofore for several of the existing summary measures. Rather the measures are used informally in practice. Small sample performance is investigated for the new and existing summary measures with simulation studies.

The methods are illustrated with data from 12,802 children with cystic fibrosis disease. We describe the data and risk modelling methods in detail later in section 7. Briefly, we compare the capacities of lung function and nutritional measures made in 1995 to predict onset of a pulmonary exacerbation event during the following year. Overall, 41% of children had a pulmonary exacerbation in 1996. Figure 1 displays predictiveness curves for two risk models, one based on lung function (FEV_1) and one based on weight. We see from Figure 1

that lung function is more predictive in the sense that more subjects have lung function based risks that are at the high and low ends of the risk scale than is true for weight based risks. Since a good risk marker is one that is helpful to individuals making medical decisions, and because decisions are more easily made when an individual's risk is high or low than if it is in the middle, we conclude informally from the curves that lung function is a superior predictor than weight. We next define formal summary indices that can be used for descriptive and comparative purposes and illustrate them with the cystic fibrosis data.

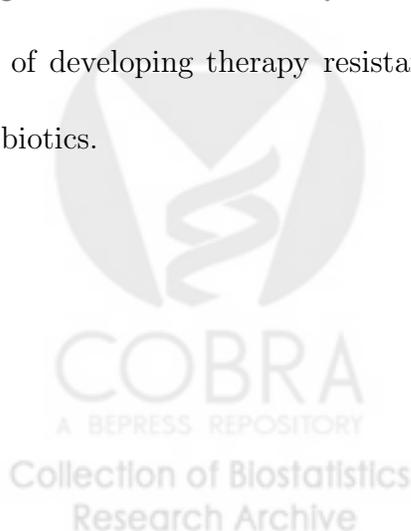
Figure 1

2 Summary Indices Involving Risk Thresholds

In clinical practice, a subject's risk is calculated to assist in medical decision making. If his risk is high, he may be recommended for diagnostic, treatment or preventive interventions. If his risk is low, he may avoid interventions that are unlikely to benefit him. In certain clinical contexts, explicit treatment guidelines exist that are based on individual risk calculations. For example, the Third Adult Treatment Panel recommends that if a subject's 10 year risk of a cardiovascular disease exceeds 20% he should consider low density lipoprotein (LDL)-lowering therapy (Adult Treatment Panel III, 2001). The risk threshold that leads one to opt for an intervention depends on anticipated costs and benefits. These may vary with individuals' perceptions and preferences (Vickers and Elkin, 2006; Hunink *et al.*, 2006). The

choice of threshold may also vary with the availability of health care resources. In this section we discuss summary indices that depend on specifying a risk threshold. To be concrete we suppose that the overall risk in the population is high, $\rho = P[D = 1]$, and that the goal of the risk model is to identify individuals at low risk, $risk(Y) < p_L$, where p_L is the risk threshold that defines low risk in the specific clinical context. Analogous discussion would pertain to a low risk population in which a risk model is sought to identify a subset of individuals at high risk. Extensions to settings where multiple risk categories are of interest occur in practice when multiple treatment options are available, and will be discussed at the end of this section.

For illustration with the cystic fibrosis data, we choose the low risk threshold $p_L = 0.25$ which contrasts with the overall incidence $\rho = 0.41$. Patients with cystic fibrosis now routinely receive inhaled antibiotic treatment to prevent pulmonary exacerbations but this was not the case in the 1990s the time during which our data were collected. If subjects at low risk, $risk(Y) < p_L$, in the absence of treatment could be identified, they could forego treatment and thereby avoid inconveniences, monetary costs and potentially increased risk of developing therapy resistant bacterial strains associated with inhaled prophylactic antibiotics.

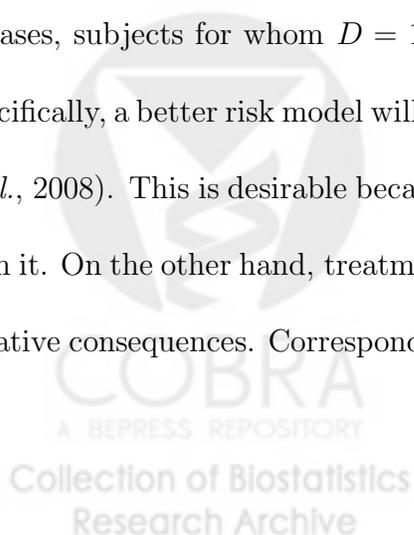


2.1 Population proportion at low risk

A simple compelling summary measure is the proportion of the population deemed to be at low risk according to the risk model. This is $R^{-1}(p_L)$, the inverse function of the predictive-ness curve, as noted earlier. A good risk prediction marker should identify more people at low risk. That is, a better model will have larger values for $R^{-1}(p_L)$. In the Cystic Fibrosis example, we see from Figure 1 that 32% of subjects in the population are in the low risk stratum based on lung function measures while 11%, are in the low risk stratum according to weight. A completely uninformative marker would put none in the low risk stratum since it assigns $risk(Y) = \rho$ to all subjects.

2.2 Cases and controls classified as low risk

Another important perspective from which to evaluate risk prediction markers is classification accuracy (Pepe *et al.*, 2008_a, Janes *et al.*, 2008). This is characterized by the risk distribution in cases, subjects for whom $D = 1$, and in controls, subjects with a good outcome $D = 0$. Specifically, a better risk model will classify fewer cases and more controls as low risk (Pencina *et al.*, 2008). This is desirable because cases should not forego treatment as they may benefit from it. On the other hand, treatment should be avoided for controls since they only suffer its negative consequences. Corresponding summary measures are termed true and false positive



rates,

$$\text{TPR}(p_L) = P(\text{risk}(Y) \geq p_L | D = 1); \quad \text{FPR}(p_L) = P(\text{risk}(Y) \geq p_L | D = 0).$$

Higher $\text{TPR}(p_L)$ and lower $\text{FPR}(p_L)$ are desirable.

Figure 2 shows cumulative distributions of $\text{risk}(Y)$ in cases and controls separately. From this, $\text{TPR}(p)$ and $\text{FPR}(p)$ can be gleaned for any value of p . We see that the proportion of controls in the low risk stratum is much larger when using lung function as the risk prediction marker than for weight, $1-\text{FPR}(p_L) = 46\%$ for lung function as opposed to 15% for weight. However the proportion of cases whose risks exceed p_L is also lower for the lung function model ($\text{TPR}(p_L) = 87\%$) than for the weight model ($\text{TPR}(p_L) = 93\%$)

Figure 2

Observe that $\text{TPR}(p_L)$ and $\text{FPR}(p_L)$ are indexed by the threshold p_L . This contrasts with the display of TPR and FPR that constitutes the ROC curve. ROC curves (Figure 3) suppress the risk thresholding values by showing TPR just as a function of FPR, not TPR and FPR as functions of risk threshold (Figure 2). When specific risk thresholds define clinically meaningful risk categories, the TPR and FPR associated with those risk category definitions are of intrinsic interest, more so than the TPR achieved at a fixed FPR value.

Figure 3

2.3 Event rates in risk strata

Another pair of summary measures is the event rates in the two risk strata. These can be thought of as predictive values, $PPV(p_L)$ and $1-NPV(p_L)$, defined as

$$PPV(p_L) = P(D = 1 | risk(Y) > p_L), \quad \text{and} \quad 1-NPV(p_L) = P(D = 1 | risk(Y) < p_L). \quad (1)$$

$PPV(p_L)$ is the event rate in the high risk stratum and $1-NPV(p_L)$ is the event rate in the low risk stratum. For a good marker, the event rate $PPV(p_L)$ will be high and the event rate $1-NPV(p_L)$ will be low.

By applying Bayes theorem to (1), PPV and NPV can be written in terms of TPR and FPR:

$$PPV(p) = \frac{\rho}{1 - \rho} \frac{TPR(p)}{FPR(p)}; \quad NPV(p) = \frac{1 - \rho}{\rho} \frac{1 - FPR(p)}{1 - TPR(p)}, \quad (2)$$

These expressions facilitate estimation of $PPV(p)$ and $NPV(p)$, which we discuss in section 4.

Event rates are also functions of the predictiveness curve. Specifically they average the curve over the ranges $(\nu_L, 1)$ and $(0, \nu_L)$ where $\nu_L = R^{-1}(p_L)$.

$$PPV(p_L) = \int_{\nu_L}^1 R(u) du / (1 - \nu_L); \quad \text{and} \quad 1-NPV(p_L) = \int_0^{\nu_L} R(u) du / \nu_L.$$

For the cystic fibrosis example, estimates of the event rates, $1-NPV(p_L)$ and $PPV(p_L)$,

are 17% and 53% for the risk strata defined by lung function. In contrast the event rates are much closer to each other, 24% and 43%, in the two risk strata defined by weight. Again lung function appears to be the better predictor of low risk. Not only is $R^{-1}(p_L)$, the size of the low risk stratum, bigger when using lung function but $1-\text{NPV}(p_L)$, the event rate in the low risk stratum, is also smaller.

2.4 ν^{th} Risk percentile

In the applied literature, variables are often categorized using quantiles. In this vein, categories of risk are sometimes defined using risk quantiles for which we have used the notation $R(\nu)$. For example, Ridker *et al.* (2000) used quartiles of risk and noted that high sensitivity c-reactive protein (hs-CRP) was more predictive of cardiovascular risk than standard lipid screening because the level of hs-CRP in the highest versus lowest quartile was associated with a much higher relative risk for future coronary events than was the case for standard lipid measurements.

Another context in which $R(\nu)$ is well motivated is when availability of medical resources is limited. Suppose resources are available to provide an intervention to a fraction $1 - \nu$ of the population, those $1 - \nu$ at highest risk. Since $R(\nu)$ is the corresponding risk quantile, subjects given the intervention have risks $\geq R(\nu)$. A marker or risk model for which $R(\nu)$ is larger is preferable because it ensures that those receiving intervention are at greater risk of a bad outcome in the absence of the intervention.

In the cystic fibrosis example, suppose the 10% of the population deemed to be at highest risk will be treated. If lung function is used to calculate risk, subjects with risks at or above 0.76 receive treatment. On the other hand if weight is used to calculate risk, subjects whose risks are as low as 0.52 will be offered treatment.

2.5 Risk threshold yielding specified TPR or FPR

In a diagnostic setting, it may be important to flag most people with disease as high risk so that people with disease get necessary treatment. In other words, we may require that the TPR exceed a certain minimum value, $TPR=t$. The corresponding risk threshold is an important entity to report. We denote it by $R(\nu_T(t))$. The decision rule that yields $TPR=t$, requires people whose risks are as low as $R(\nu_T(t))$ to undergo treatment. If the treatment is cumbersome or risky the decision rule may be unacceptable or unethical if the threshold $R(\nu_T(t))$ is low.

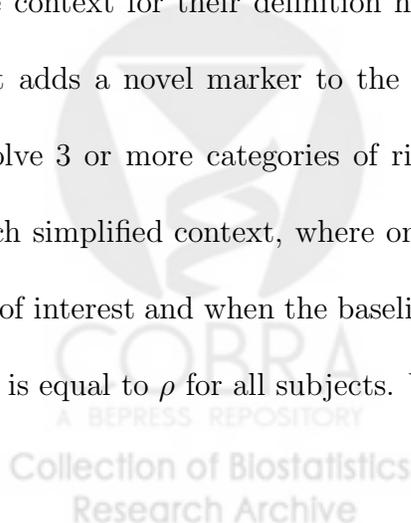
In screening healthy populations for a rare disease such as ovarian cancer, the false positive rate must be very low in order to avoid large numbers of subjects undergoing unnecessary medical procedures. The risk threshold that yields an acceptable FPR must also be acceptable for individuals as a threshold for deciding for or against medical procedures. To maintain a very low FPR, the risk threshold may be very high in which case the decision rule would not be ethical. Reporting the risk threshold that yields specified $FPR=t$ is therefore

often important in practice and we denote the threshold by $R(\nu_F(t))$.

Unlike other predictiveness summary measures, $R(\nu_T(t))$ and $R(\nu_F(t))$ may not be suited to the task of comparing markers. It is not clear that a specific ordering of thresholds is always preferable. In the cystic fibrosis example, the risk threshold that yields TPR=0.85 is 0.27 when the calculation is based on lung function, but 0.32 when weight is used. Observe that another consideration is the corresponding false positive rate which is 0.50 for lung function and 0.72 for weight. If one wanted to control the false positive rate, at FPR=0.15 say, the corresponding risk thresholds are 0.54 for lung function and 0.51 for weight. Observe that the lung function based risk threshold is lower than that for weight when controlling the TPR but higher when controlling the FPR.

2.6 Risk reclassification measures

Several summary measures that rely on defined risk categories have been proposed recently. The context for their definition has been when comparing a baseline risk model with one that adds a novel marker to the baseline predictors using risk reclassification tables that involve 3 or more categories of risk. It is illuminating to consider these measures in our much simplified context, where only 2 risk categories defined by a single risk threshold p_L are of interest and when the baseline model involves no covariates at all so that the baseline risk is equal to ρ for all subjects. We discuss the more complex setting later.



Cook (2007) proposes the reclassification percent to summarize predictive information in a model. In our context, all subjects are considered high risk under the baseline model because $\rho > p_L$. The reclassification percent is therefore the proportion of subjects classified as low risk according to the risk model involving Y . This is exactly the summary index $R^{-1}(p_L)$ discussed earlier.

Pencina *et al.* (2008) criticize the reclassification percent because it does not distinguish between desirable risk reclassifications (up for cases and down for controls) and undesirable risk reclassifications (down for cases and up for controls). They propose the net reclassification improvement (NRI) summary statistic as an alternative. We use "up" and "down" to denote changes of one or more risk categories in the upward and downward directions, respectively, for a subject between their baseline and augmented risk values. The NRI is defined as

$$\text{NRI} = [P(\text{up}|D = 1) - P(\text{down}|D = 1)] - [P(\text{up}|D = 0) - P(\text{down}|D = 0)].$$

In our simple context it is easy to see that

$$\text{NRI} = \text{TPR}(p_L) - \text{FPR}(p_L)$$

where $\text{TPR}(p_L)$ and $\text{FPR}(p_L)$ were discussed earlier. We see that in the 2 category setting the NRI statistic is equal to Youden's index (Youden, 1950). Youden's index has been criticized because implicitly it weighs equally the consequences of classifying a case as low

risk, i.e. a case failing to receive intervention, and classifying a control as high risk, i.e. a control subjected to unnecessary intervention. Most often the costs and consequences of these mistakes will differ greatly for cases and controls. Therefore we recommend reporting the two components of the NRI separately, $\text{TPR}(p_L)$ and $\text{FPR}(p_L)$. Values were reported for the cystic fibrosis study above. The corresponding NRI values are $0.33=0.87-0.54$ for lung function and $0.08=0.93-0.85$ for weight.

2.7 Extensions and discussion

A key use of summary measures is to compare different risk models. One can quantify the difference in performance between two risk models by taking the difference between summary measures derived from the two models. In the cystic fibrosis example discussed here, the two risk models involve completely different markers. However, one could also entertain two models that involve some common predictors. The setting in which risk reclassification ideas have emerged, is where one model involves standard baseline predictors and the other includes a novel marker in addition to the baseline predictors. Taking the difference in summary measures for the two models is a sensible way of assessing improvement in performance in this context too.

Recall that when only 2 risk categories (low versus high) exist, Cook's reclassification percent is equal to $R^{-1}(p_L)$ when the baseline risk does not depend on baseline covariates. However, the reclassification percent is not equal to the difference of values for $R^{-1}(p_L)$

between the baseline and augmented models when the baseline model does involve covariates. In general, even when two models have exactly the same predictive performance, the reclassification percent is typically non-zero. In fact it has been shown to vary dramatically with correlations between predictors in one model versus another (Janes *et al.*, 2008). This measure therefore does not seem well suited for gauging the difference between predictive capacities of two models. Instead we suggest that one simply focus on the difference in proportions of subjects classified as low (or high) risk with the two models, i.e. differences in $R^{-1}(p_L)$.

We represented the NRI statistic as $\text{TPR}(p_L) - \text{FPR}(p_L)$ in the simple setting. It is easy to show that when two models involve covariates, the NRI statistic to compare the two models is the difference $(\text{TPR}_1(p_L) - \text{TPR}_2(p_L)) - (\text{FPR}_1(p_L) - \text{FPR}_2(p_L))$ where subscripts 1 and 2 are used to index the two models. In analogy with our earlier discussion, we recommend reporting the two comparative components separately, $\text{TPR}_1(p_L) - \text{TPR}_2(p_L)$ and $\text{FPR}_1(p_L) - \text{FPR}_2(p_L)$, rather than their difference, the NRI, because typically changes in TPR should be weighted differently than changes in FPR.

Summarizing data is difficult when more than two risk categories are involved. Statistics such as the NRI have been criticized because they do not distinguish between changes of one risk category and more than one risk category (Pepe *et al.*, 2008_b). In a similar vein, when 3 risk categories exist with specific treatment recommendations for each, misclassifying a case as being in the lowest risk level may be more serious than misclassifying him as in the middle category. Similarly, misclassifying a control as being in the highest risk level

may be more serious than misclassifying him as being in the middle category. Without specifying utilities associated with different types of misclassifications, any accumulation of data across risk categories is difficult to justify. For these settings we propose use of a vector of summary statistics distinguished by the risk thresholds. For example, suppose we consider three risk categories for the cystic fibrosis study defined by two thresholds $p_L = 0.25$ and $p_H = 0.75$. We could report: the proportions of subjects in the highest and lowest categories, $(1 - R^{-1}(p_H), R^{-1}(p_L))$; the proportions of cases and controls in each category, $(TPR(p_H), 1 - TPR(p_L))$ and $(FPR(p_H), 1 - FPR(p_L))$; and so forth.

Although statistical summaries that depend on clinically meaningful risk thresholds are appealing, the choice of risk thresholds is often uncertain. Different clinicians or policy makers may choose different risk categorizations. This argues for displaying the risk distributions as continuous curves since one can then read from them summary indices described here using any risk threshold of interest to the reader.

3 Threshold Independent Summary Measures

Classic measures that describe the predictive strength of a model can be interpreted as summary indices for the predictiveness curve. We describe the relationships next. These measures can compliment the display of risk distributions for several models when no specific risk thresholds are of key interest. In addition, formal hypothesis tests to compare predictiveness curves can be based on them.

3.1 Proportion of explained variation

The proportion of explained variation, also called R^2 , is the most popular measure of predictive power for continuous outcomes and is popular for binary outcomes too. It is most commonly defined as

$$\text{PEV} \equiv \frac{\text{var}(D) - E(\text{var}(D|Y))}{\text{var}(D)}.$$

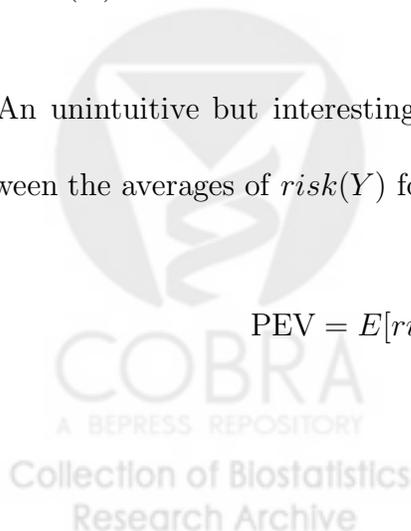
But it can also be written as

$$\text{PEV} = \text{var}(\text{risk}(Y))/\rho(1 - \rho),$$

because $\text{var}(D) = E(\text{var}(D|Y)) + \text{var}(E(D|Y))$ and $E(D|Y) = P(D = 1|Y) = \text{risk}(Y)$. PEV is a standardized measure of the variance in $\text{risk}(Y)$ since $\rho(1 - \rho)$ in the denominator is the risk variance for an ideal marker that predicts $\text{risk}(Y) = 1$ for cases and $\text{risk}(Y) = 0$ for controls. Hu *et al.* (2006) noted that PEV can also be written as the correlation between D and $\text{risk}(Y)$.

An unintuitive but interesting and simple interpretation for PEV is as the difference between the averages of $\text{risk}(Y)$ for cases and controls (Pepe *et al.*, 2008_b),

$$\text{PEV} = E[\text{risk}(Y)|D = 1] - E[\text{risk}(Y)|D = 0].$$



In summary for the cystic fibrosis data, PEV, calculated as 0.22 for the lung function measure and 0.05 for weight, can be interpreted as variances of risk distributions displayed in Figure 1 standardized by the ideal variance of $0.41 \times (1 - 0.41) = 0.24$, or as differences in means of distributions shown in Figure 2. In Figure 1, $\text{var}(\text{risk}(Y)) = 0.053$ for lung function and 0.012 for weight yielding 0.22 and 0.05 respectively when divided by 0.24. On the other hand in Figure 2, case and control mean risks are 0.54 and 0.32 for lung function while they are 0.44 and 0.39 for weight, again yielding $0.54 - 0.32 = 0.22$ and $0.44 - 0.39 = 0.05$ for the PEV values calculated as the differences in means.

Pencina *et al.* (2008) employ the PEV summary measure to gauge the improvement in risk prediction when clinically relevant risk thresholds do not exist. They do not recognize it as the proportion of explained variation but call it integrated discrimination improvement (IDI) and note that it has another interpretation as Youden's index integrated uniformly over (0,1):

$$\text{PEV} = \int YI(p)dp,$$

where $YI(p) = P(\text{risk}(Y) > p | D = 1) - P(\text{risk}(Y) > p | D = 0)$ is Youden's index for the binary decision rule that is positive when $\text{risk}(Y) > p$. In other words, PEV can also be interpreted as the difference between integrated TPR(p) and FPR(p) functions defined earlier.

In a commentary on the Pencina *et al.* paper, Ware and Cai (2008) suggest that IDI, denoted here by PEV, does not depend on the overall event rate, $\rho = P(D = 1)$. We disagree.

To illustrate, suppose we have a single marker with risk function $risk(Y)$ increasing in Y .

Then

$$\begin{aligned} PEV &= \int (P(risk(Y) > p|D = 1) - P(risk(Y) > p|D = 0))dp \\ &= \int (P(Y > y|D = 1) - P(Y > y|D = 0))\frac{\partial risk(Y)}{\partial Y}\Big|_{Y=y}dy \end{aligned}$$

Here, the conditional probabilities, $P(Y > y|D = 1)$ and $P(Y > y|D = 0)$, are independent of prevalence, ρ , but $\frac{\partial risk(Y)}{\partial Y}$ is a function of ρ . To demonstrate, consider a simple linear logistic regression model,

$$\text{logit}P(D = 1|Y) = \theta_0 + \theta_1 Y. \quad (3)$$

and note that

$$\frac{\partial risk(Y)}{\partial Y} = \frac{\partial P(D = 1|Y)}{\partial Y} = \theta_1 P(D = 1|Y)\{1 - P(D = 1|Y)\}.$$

Since $P(D = 1|Y) = \frac{P(Y|D=1)\frac{\rho}{1-\rho}}{P(Y|D=0)\frac{\rho}{1-\rho} + \{1 + \frac{P(Y|D=1)\frac{\rho}{1-\rho}}{P(Y|D=0)\frac{\rho}{1-\rho}}\}}$, clearly varies with ρ , so does its derivative. Figure 4 shows the relationship between PEV and ρ for a marker Y that is standard normally distributed in controls and normally distributed with mean 1 and variance 1 in cases. The risk is a simple linear logistic risk function (equation (3)). As ρ increases from 0 to 1, we see that PEV increases then decreases with maximum occurring at $\rho = 0.5$. Janssens *et al.* (2006) also demonstrated dependence of PEV on ρ through a simulation

study.

Figure 4

The proportion of explained variation has been defined in other ways, notably based on notions of log likelihood (deviance). Gail and Pfeiffer (2005) note that these can also be calculated from the risk distribution. However, Zheng and Agresti (2000) make the point that these summary measures are difficult to interpret and we concur wholeheartedly. Therefore, we do not pursue them further in this paper but note that methods for inference could be developed in analogy with those we develop here for PEV.

3.2 Total Gain

Total gain, proposed by Bura and Gastwirth (2001) is defined as ,

$$\text{TG} = \int_0^1 |R(\nu) - \rho| d\nu. \quad (4)$$

This is the area sandwiched between the predictiveness curve and the horizontal line at ρ , which is the predictiveness curve for a completely uninformative marker assigning $\text{risk}(Y) = \rho$ to all subjects. TG is appealing because it can be visualized directly from the predictiveness curve. For a perfect risk prediction model, the predictiveness curve is a step function rising from 0 to 1 at $\nu = 1 - \rho$. The corresponding TG is $2\rho(1 - \rho)$.

Other interpretations can be made for TG. Huang and Pepe (2008_b) have shown that TG is equivalent to the Kolmogorov-Smirnov measure of distance between risk distributions for cases and controls. This is an ROC summary index (Pepe, 2003, page 80):

$$\text{TG} = 2\rho(1 - \rho)\sup_t\{\text{ROC}(t) - t\} \quad (5)$$

$$= 2\rho(1 - \rho)\max_p\{\text{TPR}(p) - \text{FPR}(p)\}, \quad (6)$$

In fact we can write this more simply.

Result

$$\text{TG} = 2\rho(1 - \rho)\{\text{TPR}(\rho) - \text{FPR}(\rho)\} \quad (7)$$

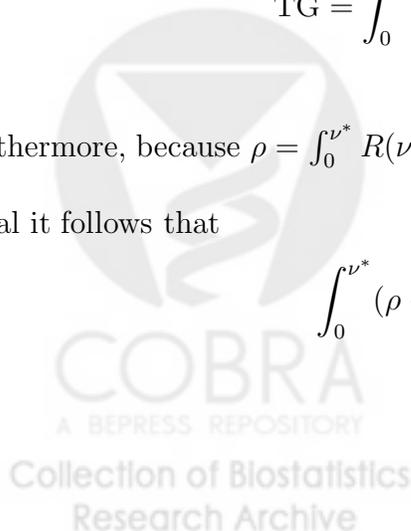
Proof

Let ν^* be the point where $R(\nu^*) = \rho$. We have

$$\text{TG} = \int_0^{\nu^*} (\rho - R(\nu))d\nu + \int_{\nu^*}^1 (R(\nu) - \rho)d\nu.$$

Furthermore, because $\rho = \int_0^{\nu^*} R(\nu)d\nu + \int_{\nu^*}^1 R(\nu)d\nu$ and $\rho = \int_0^1 \rho d\nu$, setting these two terms equal it follows that

$$\int_0^{\nu^*} (\rho - R(\nu))d\nu = \int_{\nu^*}^1 (R(\nu) - \rho)d\nu.$$



Therefore TG can be written as

$$\begin{aligned} \text{TG} &= 2 \int_{\nu^*}^1 (R(\nu) - \rho) d\nu \\ &= 2 \int_{\nu^*}^1 R(\nu) d\nu - 2\rho(1 - \nu^*) \\ &= 2\rho\text{TPR}(\rho) - 2\rho(1 - R^{-1}(\rho)) \end{aligned}$$

because $\text{TPR}(\rho) = \int_{\nu^*}^1 R(\nu) d\nu / \rho$. Moreover, since $1 - R^{-1}(\rho) = \rho\text{TPR}(\rho) + (1 - \rho)\text{FPR}(\rho)$, the above representation of TG can be further simplified to $2\rho(1 - \rho)\{\text{TPR}(\rho) - \text{FPR}(\rho)\}$.

This representation of TG is useful for estimation and for deriving asymptotic distribution theory. Interestingly, by equating (6) and (7), we find that the maximum value of $\text{TPR}(p)$ - $\text{FPR}(p)$ occurs at the risk threshold $p = \rho$. Another short proof follows by taking its derivative. In particular, since

$$\text{TPR}(R(\nu)) - \text{FPR}(R(\nu)) = \frac{\int_{\nu}^1 R(u) du}{\rho} + \frac{\int_{\nu}^1 (1 - R(u)) du}{1 - \rho}, \quad (8)$$

taking the derivative of the right side with respect to ν and setting it to 0, we have

$$\left\{1 - \frac{R(\nu)}{\rho}\right\} - \left\{1 - \frac{1 - R(\nu)}{1 - \rho}\right\} = 0$$

at the solution. That is, the solution is at $R(\nu) = \rho$. In the same illustrative setting used above, where $\rho = 0.2$, Y is standard normal in controls and normal with mean 1 and variance 1 in cases, we see from Figure 5 how $\text{TPR}(p)$ - $\text{FPR}(p)$ varies with p . The maximum value,

0.39, is achieved at $p = 0.2$, i.e. at $p = \rho$.

Figure 5

Another appealing feature of TG is that after it is standardized by $2\rho(1 - \rho)$, the total gain for a perfect marker, it is functionally independent of ρ . Let's use $\overline{\text{TG}}$ to denote standardized total gain

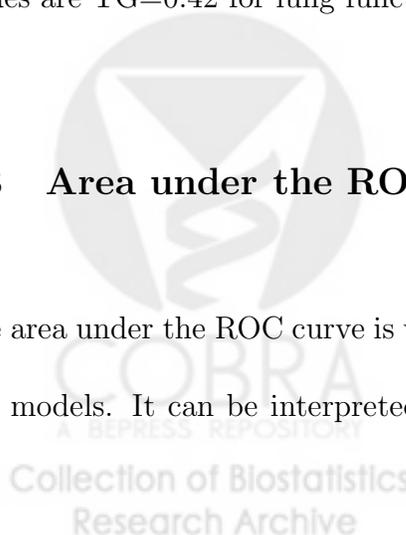
$$\overline{\text{TG}} \equiv \frac{\text{TG}}{2\rho(1 - \rho)}$$

so that $\overline{\text{TG}} \in [0, 1]$. We will focus on $\overline{\text{TG}}$ here. It is independent of disease prevalence because of its interpretation as the Kolmogorov-Smirnov ROC summary index. Moreover, based on the results above, $\overline{\text{TG}}$ is simply interpreted as the difference between the proportions of cases and controls with risks above the average, $\rho = P(D = 1) = E(\text{risk}(Y))$.

In the cystic fibrosis example, TG based on lung function is 0.20, while TG based on weight is 0.09. Since the overall event rate is $\rho=41\%$, the corresponding standardized TG values are $\overline{\text{TG}}=0.42$ for lung function and $\overline{\text{TG}} =0.20$ for weight.

3.3 Area under the ROC curve and further discussion

The area under the ROC curve is widely used to summarize and compare predictive markers and models. It can be interpreted simply as the probability of correctly ordering subjects

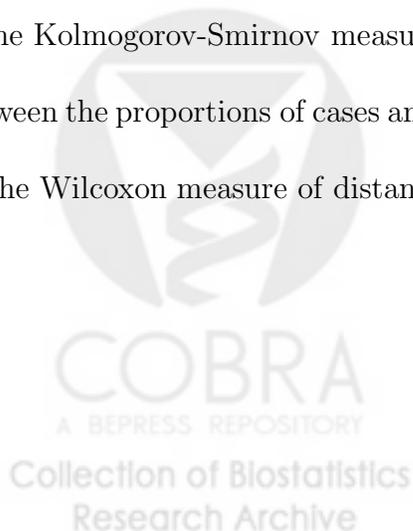


with and without events using $risk(Y)$:

$$AUC = P(risk(Y_1) > risk(Y_2) | D_1 = 1, D_2 = 0)$$

However, it has been criticized widely for having little relevance to clinical practice (Cook, 2007; Pepe and Janes, 2008; Pepe *et al.*, 2007). In particular, the task facing the clinician in practice is not to order risks for two individuals. Part of the appeal of the AUC, however, lies in the fact that it depends neither on prevalence, ρ , nor on risk thresholds. Yet in the context of risk prediction within a specific clinical population, these attributes may be weaknesses. In particular, when specific risk thresholds are of interest, the ROC curve hides them. In Figure 3, we plot the ROC curves for risk based on lung function and on weight. The AUC values are 0.771 and 0.639, respectively.

Interestingly all of the measures discussed here can be thought of as the mathematical distance between risk distributions for cases and controls (Figure 2) measured in different ways. The PEV is the difference in the means of case and control risk distributions. The \overline{TG} is the Kolmogorov-Smirnov measure and we have shown that this is equal to the difference between the proportions of cases and controls with risks larger than ρ . The AUC is equivalent to the Wilcoxon measure of distance between risk distributions for cases and controls.



4 Estimation of Summary Measures

We now turn to estimation of summary indices from data. We focus on the scenario where Y is a single continuous marker. We also allow Y to be a predefined combination of multiple markers. For example, the score may be derived from a training dataset and our task is to evaluate the combination score using a test dataset.

We use the following notation: Y , Y_D and $Y_{\bar{D}}$ are marker measurements from the general, case and control populations, respectively. Let F , F_D and $F_{\bar{D}}$ be the corresponding distribution functions and let f , f_D and $f_{\bar{D}}$ be the density functions. We assume the risk, $risk(Y) = P(D = 1|Y)$, is monotone increasing in Y . Under this assumption we have $R(\nu) = P\{D = 1|Y = F^{-1}(\nu)\}$. Thus the curve $R(\nu)$ vs. ν is the same as the curve $risk(Y)$ vs $F(Y)$ and the predictiveness curve can be obtained by first estimating the risk model $risk(Y)$, and then the marker distribution $F(Y)$. Let Y_{D_i} , $i = 1, \dots, n_D$ be the n_D independent identically distributed observations from cases, and $Y_{\bar{D}_i}$, $i = 1, \dots, n_{\bar{D}}$ be the $n_{\bar{D}}$ independent identically distributed observations from controls. We write Y_i , $i = 1, \dots, n$ for $\{Y_{D_1}, \dots, Y_{D_{n_D}}, Y_{\bar{D}_1}, \dots, Y_{\bar{D}_{n_{\bar{D}}}}\}$ where $n = n_D + n_{\bar{D}}$.

Suppose the risk model is $risk(Y) = P(D = 1|Y) = G(\theta, Y)$, where

$$\text{logit}\{G(\theta, Y)\} = \theta_0 + h(\theta_1, Y),$$

and h is some monotone increasing function of Y . This is a very general formulation. As a

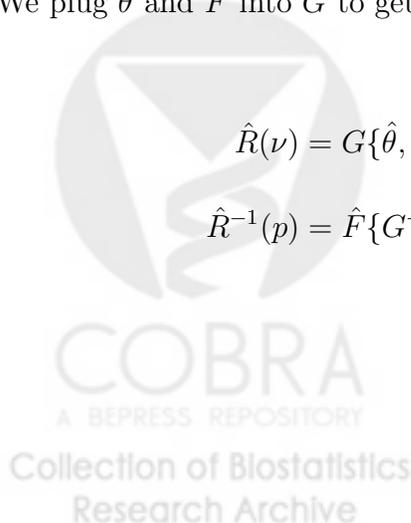
special case, $\text{logit}\{G(\theta, Y)\}$ could be as simple as $\theta_0 + \theta_1 Y$ with $\theta_1 > 0$, the ordinary linear logistic model. We consider estimation first under a cohort or cross sectional design and later discuss case-control designs for which the logistic regression formulation is particularly helpful.

4.1 Cohort Design

Suppose we have n independent identically distributed observations (Y_i, D_i) from the population. Maximum likelihood estimates of θ can be obtained, denoted by $\hat{\theta}$, as well as empirical estimates of F , F_D , $F_{\bar{D}}$, and ρ , denoted by \hat{F} , \hat{F}_D , $\hat{F}_{\bar{D}}$, and $\hat{\rho}$. We use these to calculate estimated summary indices. Summary measures that involve risk thresholds are the risk quantile, $R(\nu)$, the population proportion with risk below p , $R^{-1}(p)$, cases and controls with risks above p , $\text{TPR}(p)$ and $\text{FPR}(p)$, event rates in risk strata, $\text{PPV}(p)$ and $1\text{-NPV}(p)$, and the risk thresholds yielding specified TPR or FPR, $R(\nu_T(t))$ and $R(\nu_F(t))$.

We plug $\hat{\theta}$ and \hat{F} into G to get estimators of $R(\nu)$, and $R^{-1}(p)$:

$$\begin{aligned}\hat{R}(\nu) &= G\{\hat{\theta}, \hat{F}^{-1}(\nu)\} \quad \text{for } \nu \in (0, 1), \\ \hat{R}^{-1}(p) &= \hat{F}\{G^{-1}(\hat{\theta}, p)\} \quad \text{for } p \in \{R(\nu) : \nu \in (0, 1)\}.\end{aligned}$$



Estimates of cases and controls with risks above p are:

$$\hat{\text{TPR}}(p) = 1 - \hat{F}_D\{G^{-1}(\hat{\theta}, p)\} \quad \text{for } p \in \{R(\nu) : \nu \in (0, 1)\},$$

$$\hat{\text{FPR}}(p) = 1 - \hat{F}_D\{G^{-1}(\hat{\theta}, p)\} \quad \text{for } p \in \{R(\nu) : \nu \in (0, 1)\}.$$

We write the event rates in risk strata in terms of $\text{TPR}(p)$ and $\text{FPR}(p)$ to facilitate their estimation:

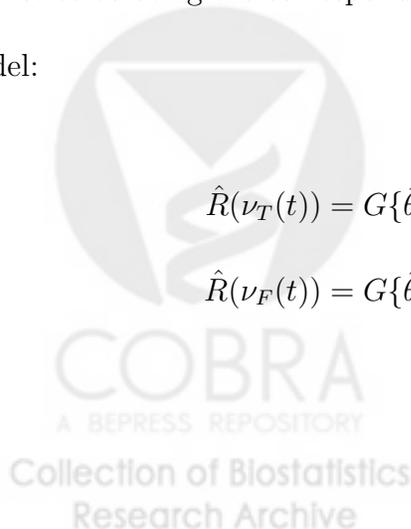
$$\hat{\text{PPV}}(p) = \frac{\hat{\rho} \hat{\text{TPR}}(p)}{1 - \hat{\rho} \hat{\text{FPR}}(p)} \quad \text{for } p \in \{R(\nu) : \nu \in (0, 1)\}$$

$$1 - \hat{\text{NPV}}(p) = 1 - \frac{1 - \hat{\rho} \hat{\text{FPR}}(p)}{\hat{\rho} \hat{\text{TPR}}(p)} \quad \text{for } p \in \{R(\nu) : \nu \in (0, 1)\}$$

In a cohort study, these estimates are equal to the empirical proportions of cases amongst those with estimated risks above and below p . However, the formulations here are valid in a case-control study too. Finally risk thresholds yielding specified TPR or FPR are obtained by first calculating the corresponding quantile of Y and then plugging it into the fitted risk model:

$$\hat{R}(\nu_T(t)) = G\{\hat{\theta}, \hat{F}_D^{-1}(\nu_T(t))\} \quad \text{for } \text{TPR} = 1 - \nu_T(t),$$

$$\hat{R}(\nu_F(t)) = G\{\hat{\theta}, \hat{F}_D^{-1}(\nu_F(t))\} \quad \text{for } \text{FPR} = 1 - \nu_T(t)$$



Summary measures that do not involve specific risk thresholds are proportion of explained variation, PEV, standardized total gain, $\overline{\text{TG}}$, and area under the ROC curve, AUC. Recall that PEV is the difference between mean risk in cases and in controls. Sample means of estimated risks yield an estimator of PEV:

$$\text{P}\hat{\text{E}}\text{V} = \int G(\hat{\theta}, Y) d\hat{F}_D(Y) - \int G(\hat{\theta}, Y) d\hat{F}_{\bar{D}}(Y).$$

On the other hand, $\overline{\text{TG}}$, can be expressed as the difference between the proportion of cases and controls with risks less than ρ . We write:

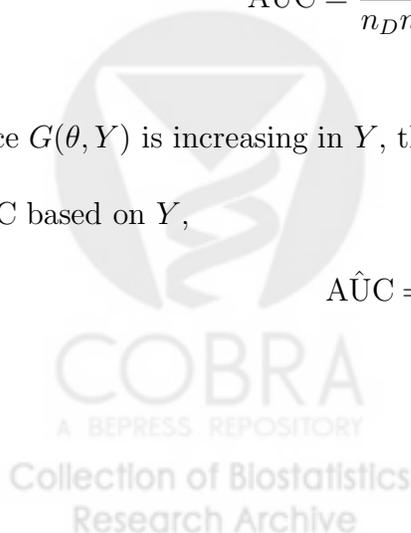
$$\hat{\overline{\text{TG}}} = \{\hat{F}_D(G^{-1}(\hat{\theta}, \hat{\rho})) - \hat{F}_{\bar{D}}(G^{-1}(\hat{\theta}, \hat{\rho}))\}.$$

Finally AUC is estimated as the proportion of case-control pairs where the estimated risk for the case exceeds that of the control

$$\text{A}\hat{\text{U}}\text{C} = \frac{1}{n_D n_{\bar{D}}} \sum_{i=1}^{n_D} \sum_{j=1}^{n_{\bar{D}}} I(G(\hat{\theta}, Y_{Di}) > G(\hat{\theta}, Y_{\bar{D}j})).$$

Since $G(\theta, Y)$ is increasing in Y , this is the same as the standard empirical estimator of the AUC based on Y ,

$$\text{A}\hat{\text{U}}\text{C} = \frac{1}{n_D n_{\bar{D}}} \sum_{i=1}^{n_D} \sum_{j=1}^{n_{\bar{D}}} I(Y_{Di} > Y_{\bar{D}j}).$$



4.2 Case-Control Design

Case-control studies are often conducted in the early phases of marker development (Pepe *et al.*, 2001; Baker *et al.*, 2002). Compared to cohort studies, they are smaller and more cost efficient. Since early phase studies dominate biomarker research, it is crucial that estimates of statistical measures of performance accommodate case-control designs. In this section, we describe estimation under a case-control design assuming that an estimate of prevalence, $\hat{\rho}$ is available. The value $\hat{\rho}$ may be derived either from a cohort which is independent from the case-control sample, or from the parent cohort within which the case-control sample is nested. As a special case one can assume ρ is known or fixed without sampling variability. In determining populations where risk markers may or may not be useful, predictiveness curves could be evaluated for various specified fixed values of ρ .

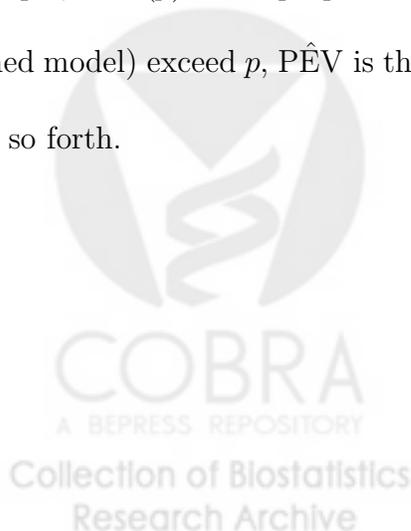
In case-control studies, we sample fixed numbers of cases and controls, n_D and $n_{\bar{D}}$, respectively. As a consequence, the intercept of the logistic risk model is not estimable. But by adjusting the intercept, we can still estimate the true risk in the population. In particular let S indicate case-control sampling. In the case-control study the risk model can be written as

$$\text{logit}\{G(\theta_S, Y)\} = \theta_{0S} + h(\theta_{1S}, Y),$$

where $\theta_{0S} = \theta_0 - \log \frac{n_{\bar{D}} \rho}{n_D (1-\rho)}$ and $\theta_{1S} = \theta_1$, and θ_0 and θ_1 are population based intercept and slope. Therefore, having calculated maximum likelihood estimates for θ_{0S} and θ_1 from the case-control study, we use $\hat{\theta}_{0S} + \log(\frac{n_D}{n_D} \frac{\hat{\rho}}{1-\hat{\rho}})$ to estimate the population intercept θ_0 .

The marker distribution in the population, F , cannot be estimated directly because of the case-control sampling design. However, since case and control samples are representative, empirical estimates of F_D and $F_{\bar{D}}$ are valid which we have denoted by \hat{F}_D and $\hat{F}_{\bar{D}}$. Therefore we estimate F with $\hat{F} = \hat{\rho}\hat{F}_D + (1 - \hat{\rho})\hat{F}_{\bar{D}}$.

Estimates of the predictiveness summary measures can then be obtained by plugging corresponding values for $\hat{\theta}$, \hat{F} , \hat{F}_D , $\hat{F}_{\bar{D}}$ and $\hat{\rho}$ into the expressions given earlier. These estimates are called semiparametric “empirical” estimates by Huang and Pepe (2008_a) because F_D and $F_{\bar{D}}$ are estimated empirically. The semiparametric likelihood framework also allows one to estimate F_D and $F_{\bar{D}}$ using maximum likelihood (Qin and Zhang, 1997, Zhang, 2000). Huang and Pepe (2008_a) compared the performance of semiparametric “empirical” estimators of the predictiveness curve with semiparametric maximum likelihood estimators. Gains in efficiency by using maximum likelihood are typically small. We use empirical estimators of F_D and $F_{\bar{D}}$ here, because this approach is intuitive and easy to implement. Moreover, they estimate important estimable quantities even when the risk model is misspecified. For example, $\hat{\text{TPR}}(p)$ is the proportion of cases whose calculated risks (calculated under the assumed model) exceed p , $\hat{\text{PEV}}$ is the difference in mean calculated risk for cases and controls, and so forth.



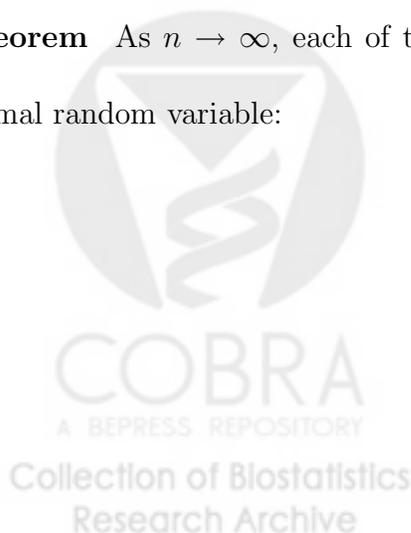
5 Asymptotic Distribution Theory

In this section, we present asymptotic distribution theory for all of the summary measures defined in previous sections. Results for pointwise estimators of $R(\nu)$ and $R^{-1}(p)$ were previously reported by Huang *et al.* (2007) and Huang and Pepe (Biometrika, in press), but for completeness we restate them here. Theory for the empirical estimator of AUC is not reported here since it is well established (Pepe, 2003, page 105). Derivations of our results are provided in the Appendix. In addition, in the Appendix, we detail the components of the asymptotic variance expressions separately for case-control and cohort study designs.

Assume the following conditions hold:

- (1) $G(s, Y)$ is a differentiable function with respect to s and Y at $s = \theta$, $Y = F^{-1}(\nu)$.
- (2) $G^{-1}(s, p)$ is continuous, and $\partial G^{-1}(s, p)/\partial s$ exists at $s = \theta$.

Theorem As $n \rightarrow \infty$, each of the following random variables converges to a mean zero normal random variable:



(i) $\sqrt{n}(\hat{R}^{-1}(p) - R^{-1}(p))$, with variance

$$\begin{aligned} \sigma_1(p)^2 &= \text{var}(\sqrt{n}(\hat{F}(G^{-1}(\theta, p)) - F(G^{-1}(\theta, p)))) + \left(\frac{\partial R^{-1}(p)}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R^{-1}(p)}{\partial \theta}\right) \\ &\quad + 2\left(\frac{\partial R^{-1}(p)}{\partial \theta}\right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}(G^{-1}(\theta, p)) - F(G^{-1}(\theta, p)))); \end{aligned}$$

(ii) $\sqrt{n}(T\hat{P}R(p) - TPR(p))$, with variance

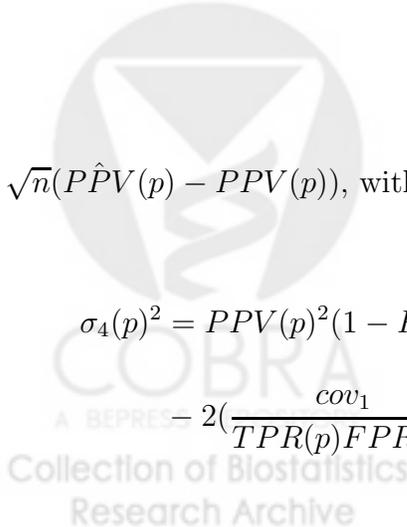
$$\begin{aligned} \sigma_2(p)^2 &= \text{var}(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))) + \left(\frac{\partial TPR(p)}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial TPR(p)}{\partial \theta}\right) \\ &\quad - 2\left(\frac{\partial TPR(p)}{\partial \theta}\right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))); \end{aligned}$$

(iii) $\sqrt{n}(F\hat{P}R(p) - FPR(p))$, with variance

$$\begin{aligned} \sigma_3(p)^2 &= \text{var}(\sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\theta, p)) - F_{\bar{D}}(G^{-1}(\theta, p)))) + \left(\frac{\partial FPR(p)}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial FPR(p)}{\partial \theta}\right) \\ &\quad - 2\left(\frac{\partial FPR(p)}{\partial \theta}\right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\theta, p)) - F_{\bar{D}}(G^{-1}(\theta, p)))); \end{aligned}$$

(iv) $\sqrt{n}(P\hat{P}V(p) - PPV(p))$, with variance

$$\begin{aligned} \sigma_4(p)^2 &= PPV(p)^2(1 - PPV(p))^2 \left\{ \left(\frac{\sigma_2(p)^2}{TPR(p)^2} + \frac{\sigma_3(p)^2}{FPR(p)^2} + \frac{\sigma_\rho^2}{\rho^2(1 - \rho)^2} \right) \right. \\ &\quad \left. - 2\left(\frac{cov_1}{TPR(p)FPR(p)} - \frac{cov_2}{TPR(p)\rho(1 - \rho)} + \frac{cov_3}{FPR(p)\rho(1 - \rho)} \right) \right\} \end{aligned}$$



where σ_ρ^2 is the asymptotic variance of $\sqrt{n}(\hat{\rho} - \rho)$ and

$$cov_1 = cov(\sqrt{n}(T\hat{P}R(p) - TPR(p)), \sqrt{n}(F\hat{P}R(p) - FPR(p))) \quad (9)$$

$$cov_2 = cov(\sqrt{n}(T\hat{P}R(p) - TPR(p)), \sqrt{n}(\hat{\rho} - \rho)) \quad (10)$$

$$cov_3 = cov(\sqrt{n}(F\hat{P}R(p) - FPR(p)), \sqrt{n}(\hat{\rho} - \rho)); \quad (11)$$

(v) $\sqrt{n}(N\hat{P}V(p) - NPV(p))$, with variance

$$\begin{aligned} \sigma_5(p)^2 = & NPV(p)^2(1 - NPV(p))^2 \left\{ \left(\frac{\sigma_5(p)^2}{(1 - TPR(p))^2} + \frac{\sigma_6(p)^2}{(1 - FPR(p))^2} + \frac{\sigma_\rho^2}{\rho^2(1 - \rho)^2} \right) \right. \\ & \left. - 2 \left(\frac{cov_1}{(1 - TPR(p))(1 - FPR(p))} + \frac{cov_2}{(1 - TPR(p))\rho(1 - \rho)} - \frac{cov_3}{(1 - FPR(p))\rho(1 - \rho)} \right) \right\}; \end{aligned}$$

(vi) $\sqrt{n}(\hat{R}(\nu) - R(\nu))$, with variance

$$\begin{aligned} \sigma_6(\nu)^2 = & \left(\frac{\partial R(\nu)}{\partial \nu} \right)^2 var(\sqrt{n}(\hat{F}(F^{-1}(\nu)) - \nu)) + \left(\frac{\partial R(\nu)}{\partial \theta} \right)^T var(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R(\nu)}{\partial \theta} \right) \\ & - 2 \left(\frac{\partial R(\nu)}{\partial \nu} \right) cov(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}(F^{-1}(\nu)) - \nu)) \left(\frac{\partial R(\nu)}{\partial \theta} \right); \end{aligned}$$



(vii) $\sqrt{n}(\hat{R}(\nu_T(t)) - R(\nu_T(t)))$ where $\text{TPR}=1 - \nu_T(t)$ is pre-specified, with variance

$$\begin{aligned} \sigma_7(t)^2 &= \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^2 \text{var}(\sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)) + \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right) \\ &\quad - 2\left(\frac{\partial R(\nu_T(t))}{\partial t}\right) \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right); \end{aligned}$$

(viii) $\sqrt{n}(\hat{R}(\nu_F(t)) - R(\nu_F(t)))$ where $\text{FPR}=1 - \nu_F(t)$ is pre-specified, with variance

$$\begin{aligned} \sigma_8(t)^2 &= \left(\frac{\partial R(\nu_F(t))}{\partial t}\right)^2 \text{var}(\sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_F(t))) - t)) + \left(\frac{\partial R(\nu_F(t))}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R(\nu_F(t))}{\partial \theta}\right) \\ &\quad - 2\left(\frac{\partial R(\nu_F(t))}{\partial t}\right) \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_F(t))) - t)) \left(\frac{\partial R(\nu_F(t))}{\partial \theta}\right); \end{aligned}$$

(ix) $\sqrt{n}(P\hat{E}V - PEV)$, with variance

$$\begin{aligned} \sigma_9^2 &= \frac{\text{var}(G(\theta, Y_D))}{n_D/n} + \frac{\text{var}(G(\theta, Y_{\bar{D}}))}{n_{\bar{D}}/n} + \left(\int \frac{\partial G(\theta, y)}{\partial \theta} dF_D(y) - \int \frac{\partial G(\theta, y)}{\partial \theta} dF_{\bar{D}}(y)\right)^T \times \\ &\quad \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \times \left(\int \frac{\partial G(\theta, y)}{\partial \theta} dF_D(y) - \int \frac{\partial G(\theta, y)}{\partial \theta} dF_{\bar{D}}(y)\right) \\ &\quad + 2\left(\int \frac{\partial G(\theta, y)}{\partial \theta} dF_D(y) - \int \frac{\partial G(\theta, y)}{\partial \theta} dF_{\bar{D}}(y)\right)^T \times \\ &\quad \left\{ \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n} \int G(\theta, Y) d(\hat{F}_D(Y) - F_D(Y))) \right. \\ &\quad \left. - \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n} \int G(\theta, Y) d(\hat{F}_{\bar{D}}(Y) - F_{\bar{D}}(Y))) \right\}; \end{aligned}$$

(x) $\sqrt{n}(\widehat{TG} - \overline{TG})$, with variance

$$\sigma_{10}^2 = \Sigma_1 - 2\Sigma_{1,2} + \Sigma_2,$$

where

$$\Sigma_1 = \text{var}(\sqrt{n}(T\hat{P}R(\hat{\rho}) - TPR(\rho))), \quad (12)$$

$$\Sigma_2 = \text{var}(\sqrt{n}(F\hat{P}R(\hat{\rho}) - FPR(\rho))), \quad (13)$$

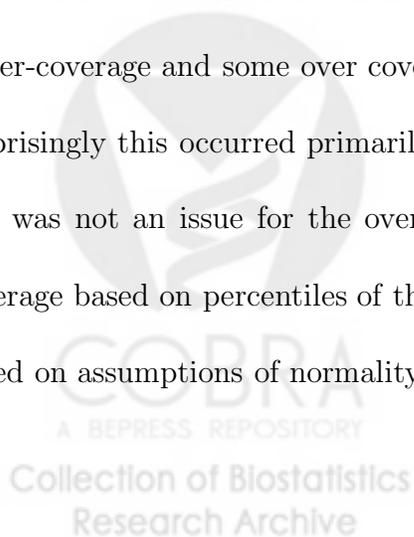
$$\Sigma_{1,2} = \text{cov}(\sqrt{n}(T\hat{P}R(\hat{\rho}) - TPR(\rho)), \sqrt{n}(F\hat{P}R(\hat{\rho}) - FPR(\rho))). \quad (14)$$

6 Simulation Studies

We performed simulation studies to investigate the validity of using large sample theory for making inference in finite sample studies, and to compare it with inference using bootstrap resampling. Data were simulated under a linear logistic risk model. Specifically we employed a population prevalence of $\rho = 0.2$ and generated marker data according to $Y_{\bar{D}} \sim N(0, 1)$ and $Y_D \sim N(1, 1)$. The correct form for $G(\theta, Y)$ was employed in fitting the risk model, namely a linear logistic model. For each simulated dataset, estimates of summary indices were calculated and their corresponding variances were estimated using the analytic formulae from the asymptotic theory. Variance estimates were also calculated using bootstrap resampling. Sample sizes ranged from 100 to 2000 and 5000 simulations were conducted for each scenario.

Simulation studies were conducted for case-control study designs as well as for cohort study designs. For the case-control scenario, we simulated nested case-control samples within the main study cohort employing equal numbers of cases and controls and with size of the cohort equal to 5 times that of the case-control study. The estimator $\hat{\rho}$ is calculated from the main study cohort and sampling variability in summary estimates due to $\hat{\rho}$ is acknowledged in making inference. Separate resampling of cases and controls was done for the nested case-control scenarios.

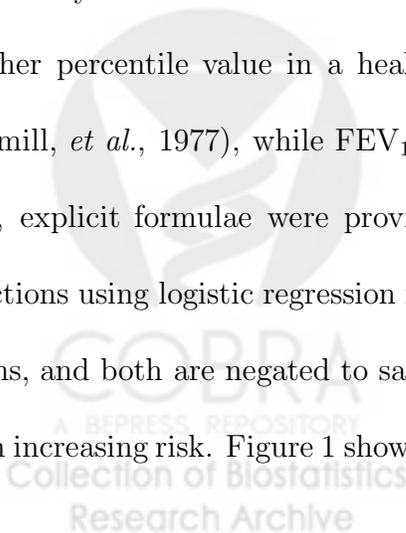
Table 1-3 displayed results for the summary indices under cohort study designs, while Table 4-6 are corresponding results under case-control study designs. Indeed, extensive simulation results for estimates of points on the predictiveness curve, $R(\nu)$ and $R^{-1}(p)$, were already reported by Huang and Pepe (2008_a). For completeness, we also included results for these two estimates under our simulation. We found little bias in the estimated values. Moreover estimated standard deviations based on asymptotic theory agree well with the actual standard deviations and with those estimated from bootstrap resampling. Coverage probabilities were excellent when sample sizes were moderate to large. We observed some under-coverage and some over coverage with small sample sizes ($n = n_D + n_{\bar{D}} = 100$). Not surprisingly this occurred primarily at the boundaries of the case and control distributions and was not an issue for the overall summary measures, PEV, $\overline{\text{TG}}$ and AUC. Generally, coverage based on percentiles of the bootstrap distribution are somewhat better than those based on assumptions of normality, but the difference shrinks for larger n .



7 The Cystic Fibrosis Data

Cystic fibrosis is an inherited chronic disease that affects the lungs and digestive system of people. A defective gene and its protein product cause the body to produce unusually thick, sticky mucus which clogs the lungs and leads to life-threatening lung infections, and also obstructs the pancreas and stops natural enzymes from helping the body break down and absorb food. The main culminating event that leads to death is acute pulmonary exacerbations, i.e. lung infection requiring intravenous antibiotics.

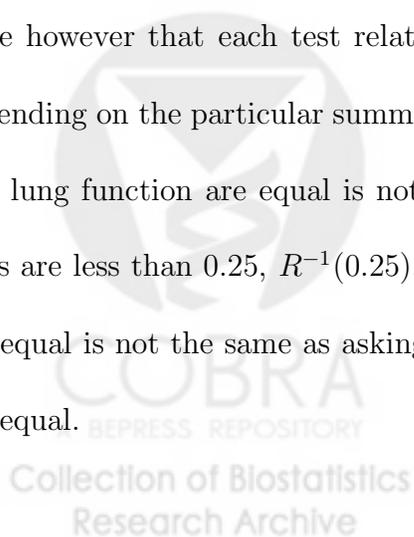
The data for analysis here is from the Cystic Fibrosis Registry, a database maintained by the Cystic Fibrosis Foundation that contains annually updated information on over 20,000 people diagnosed with CF and living in the USA. In order to illustrate our methodology, we consider FEV_1 , a measure of lung function, and weight, a measure of nutritional status, as measured in 1995 to predict occurrence of pulmonary exacerbation in 1996. Data from 12,802 patients 6 years of age and older are analyzed. 5245 subjects (41%) had at least one pulmonary exacerbation. A child's weight is standardized for age and gender by reporting his/her percentile value in a healthy population of children of the same age and gender (Hamill, *et al.*, 1977), while FEV_1 is standardized for age, gender and height in a different way, explicit formulae were provided by Knudson *et al.* (1983). We modelled the risk functions using logistic regression models with weight and FEV_1 entered the model as linear terms, and both are negated to satisfy the assumption that increasing values are associated with increasing risk. Figure 1 shows the predictiveness curves for the entire cohort and Figure



2 shows the risk distributions separately for cases (those who had a pulmonary exacerbation) and for controls.

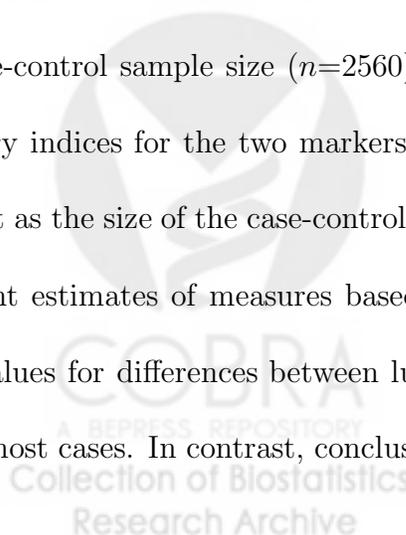
First, we use the entire cohort to estimate predictiveness summary measures for weight and lung function. Table 3 shows the point estimates discussed earlier in sections 2 and 3. Here we provide confidence intervals based on asymptotic distribution theory and on bootstrap resampling. Observe that standard deviations are all small and that corresponding confidence intervals are very tight. Bootstrap confidence intervals are almost identical to those based on asymptotic theory.

We used the summary indices as the basis for hypothesis tests to formally compare the predictive capacities of FEV_1 and weight. The difference between estimates of the indices was calculated and standardized using a bootstrap estimate of the variance of the difference. By comparing these test statistics with quantiles of the standard normal distribution, p -values were calculated. We see that differences between lung function and weight as predictive markers are statistically significant, no matter what summary index is employed. Note however that each test relates to a different question about predictive performance, depending on the particular summary index on which it is based. Asking if PEVs for weight and lung function are equal is not the same as asking if the proportion of subjects whose risks are less than 0.25, $R^{-1}(0.25)$, are equal. Asking if PEVs for weight and lung function are equal is not the same as asking if the AUCs for risks associated with weight and FEV_1 are equal.



Next, we randomly sampled 1280 cases and 1280 controls from the entire cohort to form a nested case-control study sample that is about 1/5 th the size of the cohort. Table 4 presents results that use data on FEV₁ and weight from the case-control subset and the estimate of the overall incidence of pulmonary exacerbation from the entire cohort, $\hat{\rho} = 0.41$. Estimates of summary indices are very close to the cohort estimates but corresponding confidence intervals are much wider. Nevertheless conclusions remain the same. This demonstrates that in a study where predictive markers are costly to obtain, the nested case-control design could yield considerable cost savings.

Predictiveness summary measures, such as $R(\nu)$ and $R^{-1}(p)$, are based on a single point on the predictiveness curve. Others, such as true and false positive rates and event rates, accumulate differences over part of the curve. Measures such as PEV, TG and AUC accumulate differences over the entire curve. One might conjecture that measures that accumulate differences would often be more powerful for testing if the predictiveness curves are equal. To investigate, we varied the case-control sample size and evaluated p -values associated with differences between the various summary measures. From Table 4, with a reasonably large case-control sample size ($n=2560$), we concluded that differences between almost all summary indices for the two markers are significantly different. However we see from Table 5 that as the size of the case-control study varies from medium ($n=500$) to small ($n=100$), the point estimates of measures based on specific thresholds become much more variable and p -values for differences between lung function and weight become statistically insignificant in most cases. In contrast, conclusions about the superiority of lung function as a predictive



marker remained firm when PEV, TG or AUC were used as the basis of hypothesis tests about equality of curves, even with very small sample sizes ($n=100$).

8 Concluding Remarks

This paper presents some new clinically relevant measures that quantify the predictive performance of a risk marker. New measures formally defined include $\text{TPR}(p)$, $\text{FPR}(p)$, $\text{PPV}(p)$, $\text{NPV}(p)$, $R(\nu_T)$, $R(\nu_F)$, although several of these are already used informally in the applied literature. We have previously argued for use of $R(\nu)$ and $R^{-1}(p)$ in practice. In addition we have provided new intuitive interpretations for some existing predictive measures, including the popular proportion of explained variation, PEV which is called the IDI by Pencina *et al.* (2008), the standardized total gain, $\overline{\text{TG}}$, and recently proposed risk reclassification measures, namely the NRI and the risk reclassification percent. We demonstrated that all of these measures are functions of the risk distribution, also known as the predictiveness curve. A fundamental initial step in the assessment of any risk model is to evaluate if risks calculated according to the model reflect the probabilities $P(D = 1|Y)$. The predictiveness curve can also be useful in making this assessment (Pepe *et al.*, 2008) and is complemented by use of the Hosmer-Lemeshow statistic (Hosmer and Lemeshow, 1989). In our cystic fibrosis example, the two linear logistic risk models, one for lung function and one for weight, both yield Hosmer-Lemeshow test p -values > 0.05 , indicating that they fit the observed data well.

A second contribution of this paper is to provide distribution theory for estimators of

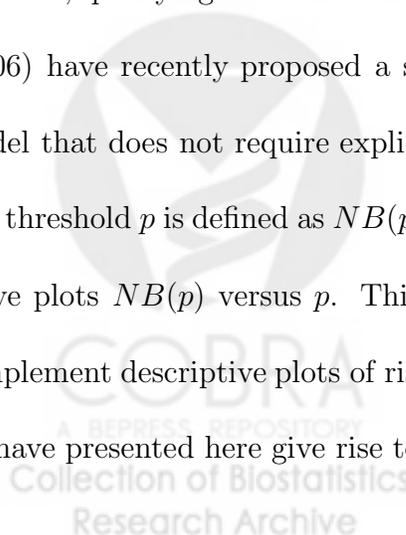
the summary indices. Such has not been available for most of the measures heretofore, including the popular PEV measure. Our methods can now be used to construct confidence intervals for the summary indices. Simulation studies indicate that the methods are valid for application in practice with finite samples.

We also demonstrated in an example how summary indices can be used to make formal rigorous comparisons between markers. Such has only been possible previously for comparisons based on the AUC or on point estimates of the predictiveness curve, $R(\nu)$ and $R^{-1}(p)$ (Huang *et al.*, 2007; Huang and Pepe, 2008).

Our methods accommodate cohort or case-control study designs. The latter are particularly important in the early phases of biomarker development when biomarker assays are expensive or procurement of biological samples is difficult (Pepe *et al.*, 2001). In such settings nested case-control studies are much more feasible (Baker *et al.*, 2002; Pepe *et al.*, 2008_d). Our methodology is currently restricted to risk models that include a single marker or a predefined combination of markers. In practice studies often involve development of a marker combination and assessment of its performance. Compelling arguments have been provided in the literature for splitting a dataset into training and test subsets (Simon, 2006; Ransohoff, 2004). In these circumstances our methods apply to evaluation with the test data of the combination developed with the training data. It would be worthwhile however to explore use of cross validation techniques for simultaneous development and assessment of risk models using the summary indices we have described.

Which summary index should be recommended for use in practice? In our opinion, a summary index should not replace the display of the risk distributions (Figures 1 and 2) but should serve only to complement them. The choice of summary indices to report should be driven by the scientific objectives of the study. For example, if the objective is to risk stratify the population according to some risk threshold, below which treatment is not indicated and above which treatment is indicated, the corresponding proportions of the population that fall into the two risk strata, $R^{-1}(p_L)$ and $1 - R^{-1}(p_L)$ would be key performance measures to report. Yet additional measures would also be of interest in this setting and should be reported including $TPR(p_L)$, $FPR(p_L)$, $PPV(p_L)$ and $NPV(p_L)$. When no risk thresholds have been defined as clinically relevant, PEV or \overline{TG} or AUC could complement the displays of risk distributions and serve as the basis of test statistics to test for the equality of differences between case and control risk distribution curves.

The final stages of evaluating a model for use in practice should incorporate notions of costs and benefits (i.e. utilities) that may be associated with decisions based on $risk(Y)$. However, specifying costs and benefits is typically very difficult in practice. Vickers and Elkin (2006) have recently proposed a standardized measure of expected utility for a prediction model that does not require explicit specifications of costs and benefits. The net benefit at risk threshold p is defined as $NB(p) = \rho TPR(p) + (1 - \rho) FPR(p)p / (1 - p)$, and their decision curve plots $NB(p)$ versus p . This weighted average of true and false positive rates could complement descriptive plots of risk distributions. Moreover, the methods for inference that we have presented here give rise to methods for inference about decision curves too.



References

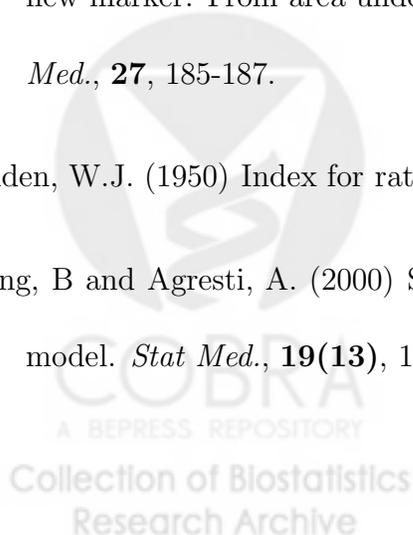
- Baker, S.G., Kramer, B.S. and Srivastava, S. (2002) Markers for early detection of cancer: statistical guidelines for nested case-control studies. *BMC Med. Res. Methodol.*, **2**, 4.
- Bura, E. and Gastwirth, J. L. (2001) The binary regression quantile plot: assessing the importance of predictors in binary regression visually. *Biometrical J.*, **43**, 5-21.
- Cook, N.R. (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, **115**, 928-935.
- Cook, N.R., Buring, J.E. and Ridker, P.M. (2006) The effect of including c-reactive protein in cardiovascular risk prediction models for women. *Ann. Intern. Med.*, **145**, 21-29.
- Expert Panel on Detection EaToHBCiA. (2001) Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *J. Am. Med. Assoc.*, **285(19)**, 2486-2497.
- Gail, M.H. and Pfeiffer, R.M. (2005) On criteria for evaluating models of absolute risk. *Biostatistics*, **6(2)**, 227-239.
- Green, D.M. and Swets, J.A. (1996) *Signal detection theory and psychophysics*. New York: Wiley.
- Gu, W. and Pepe, M.S. (2009) Estimating the capacity for improvement in risk prediction with a marker. *Biostatistics*, **10(1)**, 172-186.

- Gu, W. and Pepe, M.S. (2009_a) Measures to summarize and compare the predictive capacity of markers. *UW Biostatistics Working Paper Series*.
- Hamill, P.V., Drizd, T.A., Johnson, C.L., Reed, R.B. and Roche, A.F. (1977) NCHS growth curves for children birth-8 years. United States, Vital Health Statistics 11, pp. 1-74. Washington, DC.
- Hosmer, D.W. and Lemeshow, S. (1989) *Applied Logistic Regression* (section 5.2.2). New York: Wiley.
- Hu, B., Palta, M. and Shao, J. (2006) Properties of R^2 statistics for logistic regression. *Statist. Med.*, **25(8)**, 1383-1395.
- Huang, Y. and Pepe, M.S. (2008_a) Semiparametric and nonparametric methods for evaluating risk prediction markers in case-control studies. *Biometrika* (in press).
- Huang, Y. and Pepe, M.S. (2008_b) A parametric ROC model based approach for evaluating the predictiveness of continuous markers in case-control studies. *Biometrics*, doi: 10.1111/j.1541-0420.2005.00454.x
- Huang, Y., Pepe, M.S. and Feng, Z. (2007) Evaluating the predictiveness of a continuous marker. *Biometrics*, **63(4)**, 1181-1188.
- Hunink, M., Glasziou, P., Siegel, J., Weeks, J., Pliskin, J., Elstein, A. and Weinstein, M. (2006) *Decision making in health and medicine*. Cambridge University Press.
- Janes, H., Pepe, M.S. and Gu, W. (2008) Assessing the value of risk predictions using risk stratification tables. *Ann. Intern. Med.*, **149(10)**, 751-760.

- Janssens, A.C.J.W., Aulchenko, Y.S., Elefante, S., Borsboom, G.J.J.M, Steyerberg, E.W. and van Duijn, C.M. (2006) Predictive testing for complex diseases using multiple genes: Fact or fiction? *Genet. Med.*, **8(7)**, 395-400.
- Knudson, R.J., Lebowitz, M.D., Holberg, C.J. and Burrows, B. (1983) Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am. Rev. Respir. Dis.*, **127**, 725-734.
- Pencina, M.J., D'Agostino Sr., R.B., D'Agostino Jr., R.B. and Vasan, R.S. (2008) Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statist. Med.*, **27**, 157-172.
- Pepe, M.S. (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press.
- Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D., Thompson, M.L., Thornquist, M., Winget, M. and Yasui, Y. (2001) Phases of biomarker development for early detection of cancer. *J. Natl. Cancer Inst.*, **93**, 1054-1061.
- Pepe, M.S., Feng, Z. and Gu, W. (2008_b) Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond'. *Statist. Med.*, **27**, 173-181.
- Pepe, M.S., Feng, Z., Huang, Y., Longton, G.M., Prentice, R., Thompson, I.M. and Zheng, Y. (2008_a) Integrating the predictiveness of a marker with its performance as a classifier. *Am. J. Epidemiol.*, **167(3)**, 362-368.

- Pepe, M.S. and Janes, H. (2008_c) Gauging the performance of SNPs, biomarkers and clinical factors for predicting risk of breast cancer. *J. Natl. Cancer Inst.*, **100(14)**, 978-979.
- Pepe, M.S., Feng Z, Janes H, Bossuyt P and Potter J. (2008_d) Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design *Journal of the National Cancer Institute*, **100(20)**, 1432-1438
- Pepe, M.S., Janes, H. and Gu, W. (2007) Letter by Pepe et al regarding the article, "Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction". *Circulation*, **116**, e132.
- Prentice, R.L. and Pyke, R. (1979) Logistic disease incidence models and case-control studies. *Biometrika*, **66(3)**, 403-411.
- Qin, J. (1998) Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, **85(3)**, 619-630.
- Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.*, **66(3)**, 403-411.
- Qin, J. and Zhang, J. (1997) Empirical likelihood and general estimating equations. *Ann. Statist.*, **22(1)**, 300-325.
- Qin, J. and Zhang, J. (2003) Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika*, **84(3)**, 609-618.
- Ransohoff, D.F. (2004) Rules of evidence for cancer molecular-marker discovery and validation. *Nat. Rev. Cancer*, **4(4)**, 309-314.

- Ridker, P.M., Hennekens, C.H., Buring, J.E. and Rifai, N. (2000). C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N. Engl. Med.*, **342**, 836-843.
- Ridker, P.M., Paynter, N.P., Rifai, N., Gaziano, J.M. and Cook, N.R. (2008) C-reactive protein and parental history improve global cardiovascular risk prediction: The risk score for men. *Circulation*, **118**, 2243-2251.
- Simon, R. (2006) Roadmap for developing and validating therapeutically relevant genomic classifiers. *J. Clin. Oncol.*, **23(29)**, 7332-7341.
- Stern, R.H. (2008) Evaluating new cardiovascular risk factors for risk stratification. *J. Clin. Hypertens.*, **10(6)**, 485-488.
- Vickers, A.J. and Elkin, E.B. (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making*, **26(6)**, 565-574.
- Ware, J.H. and Cai T. (2008) Comments on ‘Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond’. *Statist. Med.*, **27**, 185-187.
- Youden, W.J. (1950) Index for rating diagnostic tests. *Cancer*, **3**, 3235.
- Zheng, B and Agresti, A. (2000) Summarizing the predictive power of a generalized linear model. *Stat Med.*, **19(13)**, 1771-1781.



Appendix: Asymptotic Theory

To simplify notation, we suppose the risk model is logistic linear in Y :

$$\text{logit}\{G(\theta, Y)\} = \theta_0 + \theta_1 Y.$$

I. Cohort Design

In a cohort study the log likelihood function is

$$l(\theta|Y_1, \dots, Y_n) = \sum_{i=1}^{n_D} \log \frac{\exp(\theta_0 + \theta_1 Y_i)}{1 + \exp(\theta_0 + \theta_1 Y_i)} + \sum_{i=1}^{n_{\bar{D}}} \log \frac{1}{1 + \exp(\theta_0 + \theta_1 Y_i)}. \quad (15)$$

Let $\hat{\theta}_0, \hat{\theta}_1$ be the maximum likelihood estimators (MLE) of θ_0, θ_1 based on (15). The following results are well known.

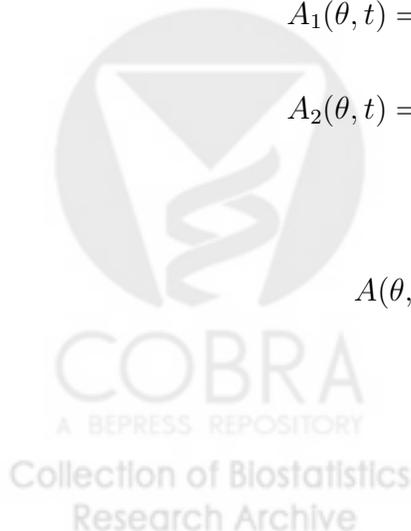
Result 1 Let

$$A_0(\theta, t) = \int_{-\infty}^t \frac{\exp(\theta_0 + \theta_1 y)}{(1 + \exp(\theta_0 + \theta_1 y))^2} dF(y)$$

$$A_1(\theta, t) = \int_{-\infty}^t \frac{y \exp(\theta_0 + \theta_1 y)}{(1 + \exp(\theta_0 + \theta_1 y))^2} dF(y)$$

$$A_2(\theta, t) = \int_{-\infty}^t \frac{y^2 \exp(\theta_0 + \theta_1 y)}{(1 + \exp(\theta_0 + \theta_1 y))^2} dF(y)$$

$$A(\theta, t) = \begin{bmatrix} A_0(\theta, t) & A_1(\theta, t)^T \\ A_1(\theta, t) & A_2(\theta, t) \end{bmatrix},$$



and $A(\theta) = A(\theta, \infty)$. If $A(\theta)^{-1}$ exists,

$$\sqrt{n} \begin{bmatrix} \hat{\theta}_0 - \theta_0 \\ \hat{\theta}_1 - \theta_1 \end{bmatrix} \rightarrow_d N(0, A(\theta)^{-1}).$$

We can also write $\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell_{\theta}(Y_i) + o_p(1)$, where $\ell_{\theta}(Y_i) = A(\theta)^{-1} i_{\theta}(Y_i)$, $i = 1, \dots, n$. And

$$i_{\theta}(Y_i) = \begin{bmatrix} \partial l(\theta|Y_i)/\partial \theta_0 \\ \partial l(\theta|Y_i)/\partial \theta_1 \end{bmatrix} = \begin{bmatrix} D_i - \exp(\theta_0 + \theta_1 Y_i)/(1 + \exp(\theta_0 + \theta_1 Y_i)) \\ D_i Y_i - Y_i \exp(\theta_0 + \theta_1 Y_i)/(1 + \exp(\theta_0 + \theta_1 Y_i)) \end{bmatrix}.$$

Result 2 As $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\rho} - \rho) \rightarrow_d N(0, \rho(1 - \rho)),$$

$$\sqrt{n}(\hat{F}(t) - F(t)) \rightarrow_d N(0, F(t)(1 - F(t))),$$

$$\sqrt{n}(\hat{F}_D(t) - F_D(t)) \rightarrow_d N(0, F_D(t)(1 - F_D(t))/\eta),$$

$$\sqrt{n}(\hat{F}_{\bar{D}}(t) - F_{\bar{D}}(t)) \rightarrow_d N(0, F_{\bar{D}}(t)(1 - F_{\bar{D}}(t))/(1 - \eta)),$$

where $\eta \equiv n_D/n$.



Lemma 1 Let

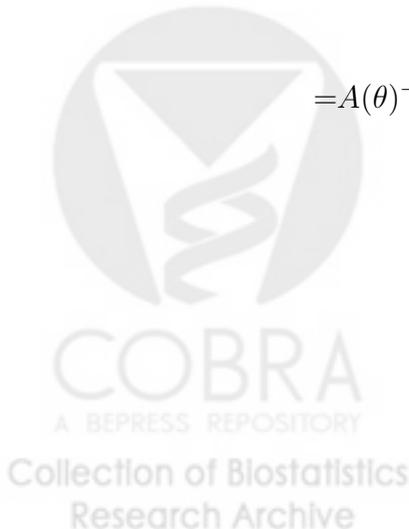
$$\begin{aligned}
 B_{D,0}(t) &= \int_{-\infty}^t \frac{1}{1 + \exp(\theta_0 + \theta_1 y)} dF_D(y) \\
 B_{D,1}(t) &= \int_{-\infty}^t \frac{y}{1 + \exp(\theta_0 + \theta_1 y)} dF_D(y) \\
 B_{\bar{D},0}(t) &= \int_{-\infty}^t \frac{\exp(\theta_0 + \theta_1 y)}{1 + \exp(\theta_0 + \theta_1 y)} dF_{\bar{D}}(y) \\
 B_{\bar{D},1}(t) &= \int_{-\infty}^t \frac{y \exp(\theta_0 + \theta_1 y)}{1 + \exp(\theta_0 + \theta_1 y)} dF_{\bar{D}}(y),
 \end{aligned}$$

and use $B_{D,0}$, $B_{D,1}$, $B_{\bar{D},0}$ and $B_{\bar{D},1}$ for the limits at $t = \infty$.

Then we have

$$\begin{aligned}
 & \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}(t) - F(t))) \\
 &= A(\theta)^{-1} \begin{bmatrix} \rho B_{D,0}(t) - (1 - \rho) B_{\bar{D},0}(t) \\ \rho B_{D,1}(t) - (1 - \rho) B_{\bar{D},1}(t) \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
 & \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(t) - F_D(t))) \\
 &= A(\theta)^{-1} \begin{bmatrix} B_{D,0}(t) - F_D(t) B_{D,0} \\ B_{D,1}(t) - F_D(t) B_{D,1} \end{bmatrix}
 \end{aligned}$$



$$\begin{aligned}
& cov(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_{\bar{D}}(t) - F_{\bar{D}}(t))) \\
&= A(\theta)^{-1} \begin{bmatrix} -B_{\bar{D},0}(t) + F_{\bar{D}}(t)B_{\bar{D},0} \\ -B_{\bar{D},1}(t) + F_{\bar{D}}(t)B_{\bar{D},1} \end{bmatrix}
\end{aligned}$$

Proof:

We prove the first result and proofs of the other two results follow from similar arguments.

$$\begin{aligned}
& cov(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}(t) - F(t))) \\
&= cov\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n A(\theta)^{-1} \dot{l}_{\theta}(Y_i), \frac{1}{\sqrt{n}} \sum_{i=1}^n (I(Y_i \leq t) - F(t))\right) \\
&= cov(A(\theta)^{-1} \dot{l}_{\theta}(Y), I(Y \leq t) - F(t)) \\
&= A(\theta)^{-1} E(\dot{l}_{\theta}(Y) \times I(Y \leq t)) \\
&= A(\theta)^{-1} \{ \rho E(\dot{l}_{\theta}(Y_D) \times I(Y_D \leq t)) + (1 - \rho) E(\dot{l}_{\theta}(Y_{\bar{D}}) \times I(Y_{\bar{D}} \leq t)) \} \\
&= A(\theta)^{-1} \begin{bmatrix} \rho B_{D,0}(t) - (1 - \rho) B_{\bar{D},0}(t) \\ \rho B_{D,1}(t) - (1 - \rho) B_{\bar{D},1}(t) \end{bmatrix}
\end{aligned}$$

Proof of Theorem items (i), (ii) and (iii) We show the proof for item (i). The proofs for items (ii) and (iii) follow similar arguments.

$$\begin{aligned}
& \sqrt{n}(\hat{R}^{-1}(p) - R^{-1}(p)) = \sqrt{n}(\hat{F}(G^{-1}(\hat{\theta}, p)) - F(G^{-1}(\theta, p))) \\
&= \sqrt{n}(\hat{F}(G^{-1}(\hat{\theta}, p)) - F(G^{-1}(\hat{\theta}, p))) + \sqrt{n}(F(G^{-1}(\hat{\theta}, p)) - F(G^{-1}(\theta, p))) + R_n,
\end{aligned}$$

where

$$R_n = \sqrt{n}(\hat{F}(G^{-1}(\hat{\theta}, p)) - \hat{F}(G^{-1}(\theta, p))) - \sqrt{n}(F(G^{-1}(\hat{\theta}, p)) - F(G^{-1}(\theta, p))) = o_p(1)$$

by equicontinuity of the process $\sqrt{n}(\hat{F} - F)$. Earlier results and the delta method then imply:

$$\begin{aligned} \sigma_1(p)^2 &= \text{var}(\sqrt{n}(\hat{R}^{-1}(p) - R^{-1}(p))) \\ &= \text{var}(\sqrt{n}(\hat{F}(G^{-1}(\theta, p)) - F(G^{-1}(\theta, p)))) + \text{var}(\sqrt{n}(F(G^{-1}(\hat{\theta}, p)) - F(G^{-1}(\theta, p)))) \\ &\quad + 2\text{cov}\sqrt{n}(\hat{F}(G^{-1}(\theta, p)) - F(G^{-1}(\theta, p))), \sqrt{n}(F(G^{-1}(\hat{\theta}, p)) - F(G^{-1}(\theta, p))) \quad (16) \\ &= R^{-1}(p)(1 - R^{-1}(p)) + \left(\frac{\partial R^{-1}(p)}{\partial \theta}\right)^T A(\theta)^{-1} \left(\frac{\partial R^{-1}(p)}{\partial \theta}\right) \\ &\quad + 2\left(\frac{\partial R^{-1}(p)}{\partial \theta}\right)^T A(\theta)^{-1} \begin{bmatrix} \rho B_{D,0}(G^{-1}(\theta, p)) - (1 - \rho)B_{\bar{D},0}(G^{-1}(\theta, p)) \\ \rho B_{D,1}(G^{-1}(\theta, p)) - (1 - \rho)B_{\bar{D},1}(G^{-1}(\theta, p)) \end{bmatrix}. \end{aligned}$$

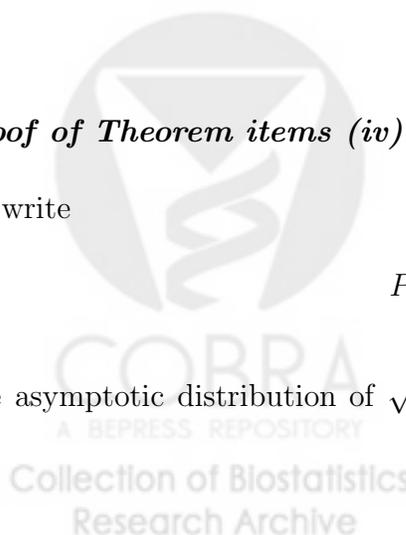
The last equality follows from Result 2 (for variance of \hat{F}), Result 1 (for variance of $\hat{\theta}$) and Lemma 1 (for covariance of $(\hat{F}, \hat{\theta})$).

Proof of Theorem items (iv) and (v)

We write

$$P\hat{P}V(p) = \frac{\hat{\rho}}{1 - \hat{\rho}} \frac{T\hat{P}R(p)}{F\hat{P}R(p)},$$

The asymptotic distribution of $\sqrt{n}(\hat{\rho} - \rho)$ is given in Result 2 as are the distributions of



$\sqrt{n}(T\hat{P}R(p) - TPR(p))$ and $\sqrt{n}(F\hat{P}R(p) - FPR(p))$ because:

$$\begin{aligned}\sqrt{n}(T\hat{P}R(p) - TPR(p)) &= \sqrt{n}((1 - \hat{F}_D(G^{-1}(\hat{\theta}, p))) - (1 - F_D(G^{-1}(\theta, p)))) \\ &= -\sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))).\end{aligned}$$

And similarly,

$$\sqrt{n}(F\hat{P}R(p) - FPR(p)) = -\sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))).$$

In the following, we calculate the covariances between $(\sqrt{n}(T\hat{P}R(p) - TPR(p)), \sqrt{n}(\hat{\rho} - \rho))$,

$(\sqrt{n}(F\hat{P}R(p) - FPR(p)), \sqrt{n}(\hat{\rho} - \rho))$ and $(\sqrt{n}(T\hat{P}R(p) - TPR(p)), \sqrt{n}(F\hat{P}R(p) - FPR(p)))$.

The asymptotic variance of $\sqrt{n}(P\hat{P}V(p) - PPV(p))$, $\sigma_4(p)^2$, then follows from the delta method.

Consider the covariance between $\sqrt{n}(T\hat{P}R(p) - TPR(p))$ and $\sqrt{n}(F\hat{P}R(p) - FPR(p))$:

$$\begin{aligned}cov_1 &= cov(\sqrt{n}(T\hat{P}R(p) - TPR(p)), \sqrt{n}(F\hat{P}R(p) - FPR(p))) \\ &= cov(\sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p)))) \\ &= cov(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))) + \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))) + o_p(1), \\ &\quad \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))) + \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))) + o_p(1)),\end{aligned}$$

Because \hat{F}_D and $\hat{F}_{\bar{D}}$ are uncorrelated, the above covariance can be written as

$$\begin{aligned}
cov_1 &= cov(\sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(F_{\bar{D}}(G^{-1}(\hat{\theta}, p)) - F_{\bar{D}}(G^{-1}(\theta, p)))) \\
&\quad + cov(\sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\theta, p)) - F_{\bar{D}}(G^{-1}(\theta, p))), \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p)))) \\
&\quad + cov(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(F_{\bar{D}}(G^{-1}(\hat{\theta}, p)) - F_{\bar{D}}(G^{-1}(\theta, p)))) \\
&= (\frac{\partial TPR(p)}{\partial \theta})^T var(\sqrt{n}(\hat{\theta} - \theta)) (\frac{\partial FPR(p)}{\partial \theta}) \\
&\quad - (\frac{\partial TPR(p)}{\partial \theta})^T cov(\sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\theta, p)) - F_{\bar{D}}(G^{-1}(\theta, p))), \sqrt{n}(\hat{\theta} - \theta)) \\
&\quad - (\frac{\partial FPR(p)}{\partial \theta})^T cov(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\theta} - \theta)) \tag{17} \\
&= (\frac{\partial TPR(p)}{\partial \theta})^T A(\theta)^{-1} (\frac{\partial FPR(p)}{\partial \theta}) \\
&\quad - (\frac{\partial TPR(p)}{\partial \theta})^T A(\theta)^{-1} \begin{bmatrix} -B_{\bar{D},0}(G^{-1}(\theta, p)) + (1 - FPR(p))B_{\bar{D},0} \\ -B_{\bar{D},1}(G^{-1}(\theta, p)) + (1 - FPR(p))B_{\bar{D},1} \end{bmatrix} \\
&\quad - (\frac{\partial FPR(p)}{\partial \theta})^T A(\theta)^{-1} \begin{bmatrix} B_{D,0}(G^{-1}(\theta, p)) - (1 - TPR(p))B_{D,0} \\ B_{D,1}(G^{-1}(\theta, p)) - (1 - TPR(p))B_{D,1} \end{bmatrix} \tag{18}
\end{aligned}$$

The last equality uses Result 1 (for variance of $\hat{\theta}$) and Lemma 1 (for covariance of $(\hat{F}_D, \hat{\theta})$ and $(\hat{F}_{\bar{D}}, \hat{\theta})$).



The second covariance (equation (10)) is between $\sqrt{n}(\hat{\rho} - \rho)$ and $\sqrt{n}(T\hat{P}R(p) - TPR(p))$:

$$\begin{aligned}
cov_2 &= cov(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(T\hat{P}R(p) - TPR(p))) \\
&= -cov(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))) + \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p)))) \\
&= -cov(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p)))) \\
&\quad -cov(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))) \\
&\equiv -(A_n + B_n),
\end{aligned}$$

where $A_n = cov(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))))$ and $B_n = cov(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))))$.

Observe that

$$\begin{aligned}
\sqrt{n}(\hat{\rho} - \rho) &= \sqrt{n}\left(\int G(\hat{\theta}, Y)d\hat{F}(Y) - \int G(\theta, Y)dF(Y)\right) \\
&= \sqrt{n}\left(\int (G(\hat{\theta}, Y) - G(\theta, Y))dF(Y) + \int G(\theta, Y)d(\hat{F}(Y) - F(Y))\right) + H_n \\
&= \left(\int \frac{\partial R(\nu)}{\partial \theta}d\nu\right)\sqrt{n}(\hat{\theta} - \theta) + \sqrt{n}\int G(\theta, Y)d(\hat{F}(Y) - F(Y)) + H_n,
\end{aligned}$$

where $R(\nu) \equiv G(\theta, Y)$ and $H_n \equiv \sqrt{n}\int (G(\hat{\theta}, Y) - G(\theta, Y))d(\hat{F}(Y) - F(Y)) = \frac{1}{\sqrt{n}}\int \sqrt{n}(G(\hat{\theta}, Y) - G(\theta, Y))d(\sqrt{n}(\hat{F}(Y) - F(Y)))$. Both $\sqrt{n}(G(\hat{\theta}, Y) - G(\theta, Y))$ and $\sqrt{n}(\hat{F}(Y) - F(Y))$ are bounded in probability and therefore H_n converges to 0 as $n \rightarrow \infty$.

We next derive expressions for A_n and B_n .

$$\begin{aligned}
A_n &= \text{cov}(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p)))) \\
&= \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T \text{cov}(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(\hat{\theta} - \theta)) \\
&= \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T \left\{ \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \int \frac{\partial R(\nu)}{\partial\theta} d\nu + \text{cov}\left(\sqrt{n} \int G(\theta, Y) d(\hat{F}(Y) - F(Y)), \sqrt{n}(\hat{\theta} - \theta)\right) \right\} \\
&= \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T \left\{ \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \int \frac{\partial R(\nu)}{\partial\theta} d\nu + \text{cov}\left(1/\sqrt{n} \sum_{i=1}^n G(\theta, Y_i), 1/\sqrt{n} \sum_{i=1}^n A(\theta)^{-1} \dot{l}_\theta(Y_i)\right) \right\} \\
&= \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T A(\theta)^{-1} \left(\int \frac{\partial R(\nu)}{\partial\theta} d\nu \right) + \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T A(\theta)^{-1} \text{cov}(G(\theta, Y), \dot{l}_\theta(Y)) \\
&= \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T A(\theta)^{-1} \left(\int \frac{\partial R(\nu)}{\partial\theta} d\nu \right) \\
&\quad + \left(\frac{\partial(1 - TPR(p))}{\partial\theta}\right)^T A(\theta)^{-1} \left[\begin{array}{l} \rho \int \frac{G(\theta, y)}{1 + \exp(\theta_0 + \theta_1 y)} dF_D(y) + (1 - \rho) \int \frac{-G(\theta, Y) \exp(\theta_0 + \theta_1 y)}{1 + \exp(\theta_0 + \theta_1 y)} dF_{\bar{D}}(y) \\ \rho \int \frac{yG(\theta, y)}{1 + \exp(\theta_0 + \theta_1 y)} dF_D(y) + (1 - \rho) \int \frac{-yG(\theta, Y) \exp(\theta_0 + \theta_1 y)}{1 + \exp(\theta_0 + \theta_1 y)} dF_{\bar{D}}(y) \end{array} \right]
\end{aligned} \tag{19}$$

And B_n is

$$\begin{aligned}
B_n &= \text{cov}(\sqrt{n}(\hat{\rho} - \rho), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))) \\
&= \left(\int \frac{\partial R(\nu)}{\partial\theta} d\nu \right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))) \\
&\quad \text{cov}\left(\int G(\theta, Y) d(\sqrt{n}(\hat{F}(Y) - F(Y))), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))\right) \\
&= \left(\int \frac{\partial R(\nu)}{\partial\theta} d\nu \right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))) \\
&\quad + \text{cov}\left(1/\sqrt{n} \sum_{i=1}^n G(\theta, Y_i), \sqrt{n}/n_D \sum_{i=1}^{n_D} I(Y_{Di} \leq G^{-1}(\theta, p))\right)
\end{aligned}$$

$$\begin{aligned}
&= \left(\int \frac{\partial R(\nu)}{\partial \theta} d\nu \right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p)))) \\
&\quad + \left(\int_{-\infty}^{G^{-1}(\theta, p)} G(\theta, Y) dF_D(Y) - F_D(G^{-1}(\theta, p)) \int G(\theta, Y) dF_D(Y) \right), \tag{20}
\end{aligned}$$

where $\text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))))$ is given by Lemma 1.

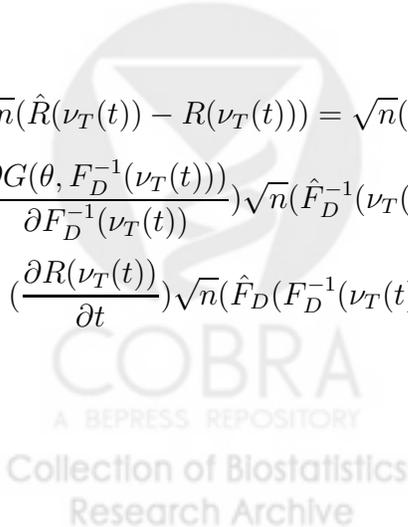
Combining the two terms yields a value for cov_2 . The derivation of cov_3 follows from a similar argument.

The proof of item (v) of the Theorem uses exactly the same techniques.

Proof of Theorem items (vi), (vii) and (viii) We prove Theorem item (vii) in the following. Proofs of (vi) and (viii) are similar. The following proof is based on Huang and Pepe (2008_a) where they derived the asymptotic distribution of $\sqrt{n}(\hat{R}(\nu) - R(\nu))$ when $R(\nu)$ is estimated under a case-control design.

When the value of TPR is $1 - \nu_T(t)$, by a Taylor series expansion,

$$\begin{aligned}
&\sqrt{n}(\hat{R}(\nu_T(t)) - R(\nu_T(t))) = \sqrt{n}(G(\hat{\theta}, \hat{F}_D^{-1}(\nu_T(t))) - G(\theta, F_D^{-1}(\nu_T(t)))) \\
&= \left(\frac{\partial G(\theta, F_D^{-1}(\nu_T(t)))}{\partial F_D^{-1}(\nu_T(t))} \right) \sqrt{n}(\hat{F}_D^{-1}(\nu_T(t)) - F_D^{-1}(\nu_T(t))) + \left(\frac{\partial G(\theta, F^{-1}(\nu_T(t)))}{\partial \theta} \right)^T \sqrt{n}(\hat{\theta} - \theta) + o_p(1) \\
&= - \left(\frac{\partial R(\nu_T(t))}{\partial t} \right) \sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t) + \left(\frac{\partial R(\nu_T(t))}{\partial \theta} \right)^T \sqrt{n}(\hat{\theta} - \theta) + o_p(1)
\end{aligned}$$



It follows that the asymptotic variance is

$$\begin{aligned}
\sigma_7(t)^2 &= \text{var}(\sqrt{n}(\hat{R}(\nu_T(t)) - R(\nu_T(t)))) \\
&= \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^2 \text{var}(\sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)) + \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right) \\
&\quad - 2\left(\frac{\partial R(\nu_T(t))}{\partial t}\right) \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right). \tag{21}
\end{aligned}$$

The variance of $\sqrt{n}(\hat{\theta} - \theta)$ and of $\sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)$ are provided in Result 2, and their covariance is provided in Lemma 1. Putting them all together, we have the following result,

$$\begin{aligned}
\sigma_7(t)^2 &= \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^2 \nu_T(t)(1 - \nu_T(t))/\eta + \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^T A(\theta)^{-1} \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right) \\
&\quad - 2\left(\frac{\partial R(\nu_T(t))}{\partial t}\right) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right)^T A(\theta)^{-1} \begin{bmatrix} B_{D,0}(F_D^{-1}(\nu_T(t))) - \nu_T(t)B_{\bar{D},0}(F_D^{-1}(\nu_T(t))) \\ B_{D,1}(F_D^{-1}(\nu_T(t))) - \nu_T(t)B_{\bar{D},1}(F_D^{-1}(\nu_T(t))) \end{bmatrix}
\end{aligned}$$

Proof of Theorem item (ix)

$$\begin{aligned}
&\sqrt{n}(P\hat{E}V - PEV) \\
&= \sqrt{n}\left\{\int G(\hat{\theta}, Y)d\hat{F}_D - \int G(\hat{\theta}, Y)d\hat{F}_{\bar{D}} - \left(\int G(\theta, Y)dF_D - \int G(\theta, Y)dF_{\bar{D}}\right)\right\} \\
&= \sqrt{n}\left\{\int G(\hat{\theta}, Y)d\hat{F}_D - \int G(\theta, Y)dF_D - \left(\int G(\hat{\theta}, Y)d\hat{F}_{\bar{D}} - \int G(\theta, Y)dF_{\bar{D}}\right)\right\} \\
&= \left\{\sqrt{n}\left(\int G(\theta, Y)d(\hat{F}_D - F_D)\right) + \sqrt{n}\left(\int (G(\hat{\theta}, Y) - G(\theta, Y))dF_D\right)\right\} \\
&\quad - \left\{\sqrt{n}\left(\int G(\theta, Y)d(\hat{F}_{\bar{D}} - F_{\bar{D}})\right) + \sqrt{n}\left(\int (G(\hat{\theta}, Y) - G(\theta, Y))dF_{\bar{D}}\right)\right\} + P_n
\end{aligned}$$

$$\equiv (A_n + B_n) - (C_n + D_n) + P_n,$$

where $P_n = \sqrt{n} \int (G(\hat{\theta}, Y) - G(\theta, Y)) d(\hat{F}_D - F_D) + \sqrt{n} \int (G(\hat{\theta}, Y) - G(\theta, Y)) d(\hat{F}_{\bar{D}} - F_{\bar{D}})$. It is easy to see that P_n converges to zero as $n \rightarrow \infty$ since $\sqrt{n}(G(\hat{\theta}, Y) - G(\theta, Y))$ is bounded in probability and $\hat{F}_D - F_D$ (or $\hat{F}_{\bar{D}} - F_{\bar{D}}$) converges in probability to 0. We define $A_n \equiv \sqrt{n}(\int G(\theta, Y) d(\hat{F}_D - F_D))$, $B_n \equiv \sqrt{n}(\int (G(\hat{\theta}, Y) - G(\theta, Y)) dF_D)$, $C_n \equiv \sqrt{n}(\int G(\theta, Y) d(\hat{F}_{\bar{D}} - F_{\bar{D}}))$ and $D_n \equiv \sqrt{n}(\int (G(\hat{\theta}, Y) - G(\theta, Y)) dF_{\bar{D}})$.

Now we have,

$$\begin{aligned} & \text{var}(A_n + B_n) = \text{var}(A_n) + \text{var}(B_n) + 2\text{cov}(A_n, B_n) \\ &= \text{var}\left(\frac{1}{\sqrt{\eta}} \frac{1}{\sqrt{n_D}} \sum_{i=1}^{n_D} G(\theta, Y_{Di})\right) + \text{var}\left(\left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right)^T \sqrt{n}(\hat{\theta} - \theta)\right) \\ & \quad + 2\left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right)^T \text{cov}\left(\sqrt{n}\left(\int G(\theta, Y) d(\hat{F}_D - F_D)\right), \sqrt{n}(\hat{\theta} - \theta)\right) \\ &= \text{var}(G(\theta, Y_D))/\eta + \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right)^T A(\theta)^{-1} \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right) \\ & \quad + 2\left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right)^T \text{cov}(G(\theta, Y_D), A(\theta)^{-1} \dot{l}_\theta(Y_D)) \\ &= \text{var}(G(\theta, Y_D))/\eta + \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right)^T A(\theta)^{-1} \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right) \\ & \quad + 2\left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D(Y)\right)^T A(\theta)^{-1} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}, \end{aligned} \tag{22}$$

where

$$\begin{aligned}
 \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} &\equiv \begin{bmatrix} \int G(\theta, Y)(\partial l(\theta|Y_D)/\partial\theta_0)dF_D(Y) - \int G(\theta, Y)dF_D(Y) \int (\partial l(\theta|Y_D)/\partial\theta_0)dF_D(Y) \\ \int G(\theta, Y)(\partial l(\theta|Y_D)/\partial\theta_1)dF_D(Y) - \int G(\theta, Y)dF_D(Y) \int (\partial l(\theta|Y_D)/\partial\theta_0)dF_D(Y) \end{bmatrix} \\
 &= \begin{bmatrix} \int \frac{G(\theta, Y)}{1+\exp(\theta_0+\theta_1 Y)}dF_D(Y) - \int G(\theta, Y)dF_D(Y) \int \frac{1}{1+\exp(\theta_0+\theta_1 Y)}dF_D(Y) \\ \int \frac{YG(\theta, Y)}{1+\exp(\theta_0+\theta_1 Y)}dF_D(Y) - \int G(\theta, Y)dF_D(Y) \int \frac{Y}{1+\exp(\theta_0+\theta_1 Y)}dF_D(Y) \end{bmatrix}. \quad (23)
 \end{aligned}$$

From a similar argument,

$$\begin{aligned}
 &var(C_n + D_n) = var(C_n) + var(D_n) + 2cov(C_n, D_n) \\
 &= var(G(\theta, Y_{\bar{D}}))/(1 - \eta) + \left(\int \frac{\partial G(\theta, Y)}{\partial\theta}dF_{\bar{D}} \right)^T A(\theta)^{-1} \left(\int \frac{\partial G(\theta, Y)}{\partial\theta}dF_{\bar{D}} \right) \\
 &\quad + 2 \left(\int \frac{\partial G(\theta, Y)}{\partial\theta}dF_{\bar{D}} \right)^T A(\theta)^{-1} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}, \quad (24)
 \end{aligned}$$

where

$$\begin{aligned}
 \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} &\equiv \begin{bmatrix} - \int G(\theta, Y)(\partial l(\theta|Y_{\bar{D}})/\partial\theta_0)dF_{\bar{D}}(Y) + \int G(\theta, Y)dF_{\bar{D}}(Y) \int (\partial l(\theta|Y_{\bar{D}})/\partial\theta_0)dF_{\bar{D}}(Y) \\ - \int G(\theta, Y)(\partial l(\theta|Y_{\bar{D}})/\partial\theta_1)dF_{\bar{D}}(Y) + \int G(\theta, Y)dF_{\bar{D}}(Y) \int (\partial l(\theta|Y_{\bar{D}})/\partial\theta_1)dF_{\bar{D}}(Y) \end{bmatrix} \\
 &= \begin{bmatrix} - \int \frac{G(\theta, Y)\exp(\theta_0+\theta_1 Y)}{1+\exp(\theta_0+\theta_1 Y)}dF_{\bar{D}}(Y) + \int G(\theta, Y)dF_{\bar{D}}(Y) \int \frac{\exp(\theta_0+\theta_1 Y)}{1+\exp(\theta_0+\theta_1 Y)}dF_{\bar{D}}(Y) \\ - \int \frac{YG(\theta, Y)\exp(\theta_0+\theta_1 Y)}{1+\exp(\theta_0+\theta_1 Y)}dF_{\bar{D}}(Y) + \int G(\theta, Y)dF_{\bar{D}}(Y) \int \frac{Y\exp(\theta_0+\theta_1 Y)}{1+\exp(\theta_0+\theta_1 Y)}dF_{\bar{D}}(Y) \end{bmatrix}. \quad (25)
 \end{aligned}$$

Because \hat{F}_D and $\hat{F}_{\bar{D}}$ are independent the covariance between A_n and C_n is zero. Observe

also that from previous derivations we have

$$\begin{aligned}
& cov(A_n + B_n, C_n + D_n) = cov(A_n, D_n) + cov(B_n, C_n) + cov(B_n, D_n) \\
& = \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D \right)^T A(\theta)^{-1} \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_{\bar{D}} \right) \\
& \quad + \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_D \right)^T A(\theta)^{-1} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} + \left(\int \frac{\partial G(\theta, Y)}{\partial \theta} dF_{\bar{D}} \right)^T A(\theta)^{-1} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \quad (26)
\end{aligned}$$

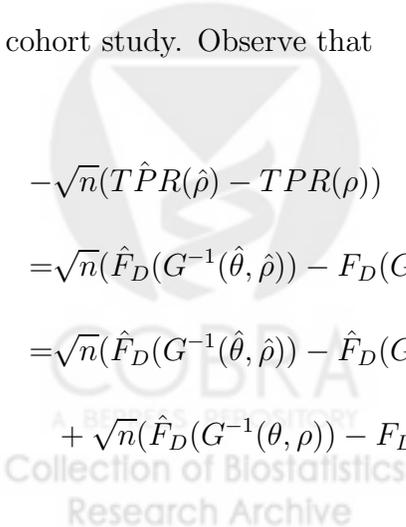
The asymptotic variance of $\sqrt{n}(\hat{PEV} - PEV)$, σ_9^2 , can be obtained by combining $var(A_n + B_n)$, $var(C_n + D_n)$ and $cov(A_n + B_n, C_n + D_n)$.

Proof of Theorem item (x)

$$\begin{aligned}
\sqrt{n}(\widehat{TG} - \overline{TG}) &= \sqrt{n}\{(T\hat{P}R(\hat{\rho}) - F\hat{P}R(\hat{\rho})) - (TPR(\rho) - FPR(\rho))\} \\
&= \sqrt{n}(T\hat{P}R(\hat{\rho}) - TPR(\rho)) - \sqrt{n}(F\hat{P}R(\hat{\rho}) - FPR(\rho)).
\end{aligned}$$

The result in the Theorem follows. Now we derive expressions for the variance components in a cohort study. Observe that

$$\begin{aligned}
& -\sqrt{n}(T\hat{P}R(\hat{\rho}) - TPR(\rho)) \\
& = \sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, \hat{\rho})) - F_D(G^{-1}(\theta, \rho))) \\
& = \sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, \hat{\rho})) - \hat{F}_D(G^{-1}(\theta, \rho))) - \sqrt{n}(F_D(G^{-1}(\hat{\theta}, \hat{\rho})) - F_D(G^{-1}(\theta, \rho))) \\
& \quad + \sqrt{n}(\hat{F}_D(G^{-1}(\theta, \rho)) - F_D(G^{-1}(\theta, \rho))) + \sqrt{n}(F_D(G^{-1}(\hat{\theta}, \hat{\rho})) - F_D(G^{-1}(\theta, \rho)))
\end{aligned}$$



$$\begin{aligned}
&= \sqrt{n}(\hat{F}_D(G^{-1}(\theta, \rho)) - F_D(G^{-1}(\theta, \rho))) + f_D(G^{-1}(\theta, \rho)) \frac{\partial G^{-1}(\theta, \rho)}{\partial \theta} \sqrt{n}(\hat{\theta} - \theta) \\
&\quad + f_D(G^{-1}(\theta, \rho)) \frac{\partial G^{-1}(\theta, \rho)}{\partial \rho} \sqrt{n}(\hat{\rho} - \rho) + o_p(1) \\
&\equiv A_n + B_n + C_n + o_p(1),
\end{aligned}$$

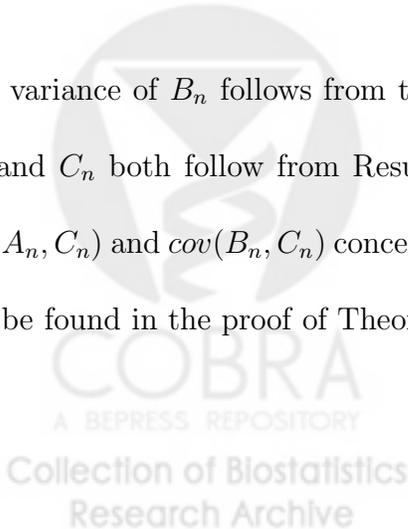
where we define

$$\begin{aligned}
A_n &\equiv \sqrt{n}(\hat{F}_D(G^{-1}(\theta, \rho)) - F_D(G^{-1}(\theta, \rho))), \\
B_n &\equiv f_D(G^{-1}(\theta, \rho)) \frac{\partial G^{-1}(\theta, \rho)}{\partial \theta} \sqrt{n}(\hat{\theta} - \theta), \\
C_n &\equiv f_D(G^{-1}(\theta, \rho)) \frac{\partial G^{-1}(\theta, \rho)}{\partial \rho} \sqrt{n}(\hat{\rho} - \rho).
\end{aligned}$$

and note that $\sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, \hat{\rho})) - \hat{F}_D(G^{-1}(\theta, \rho))) - \sqrt{n}(F_D(G^{-1}(\hat{\theta}, \hat{\rho})) - F_D(G^{-1}(\theta, \rho))) \rightarrow 0$ as $n \rightarrow \infty$ due to the equicontinuity of the process.

$$\begin{aligned}
\Sigma_1 &\equiv \text{var}(\sqrt{n}(T\hat{P}R(\hat{\rho}) - TPR(\rho))) \\
&= \text{var}(A_n) + \text{var}(B_n) + \text{var}(C_n) + \text{cov}(A_n, B_n) + \text{cov}(A_n, C_n) + \text{cov}(B_n, C_n) \quad (27)
\end{aligned}$$

The variance of B_n follows from that of $\sqrt{n}(\hat{\theta} - \theta)$ given in Result 1, and the variances of A_n and C_n both follow from Result 2. $\text{cov}(A_n, B_n)$ follows from Lemma 1. Furthermore, $\text{cov}(A_n, C_n)$ and $\text{cov}(B_n, C_n)$ concern the covariance between $(\hat{F}_D, \hat{\rho})$ and $(\hat{\theta}, \hat{\rho})$, both of which can be found in the proof of Theorem item (iv), cov_2 (see equation (19) and (20)).



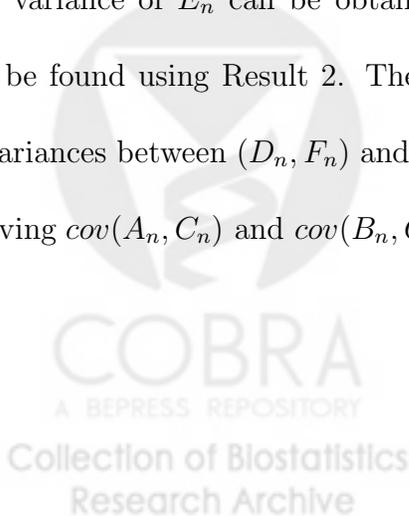
Similarly, we have

$$\begin{aligned}
& -\sqrt{n}(F\hat{P}R(\hat{\rho}) - FPR(\rho)) \\
& =\sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\hat{\theta}, \hat{\rho})) - F_{\bar{D}}(G^{-1}(\theta, \rho))) \\
& =\sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\theta, \rho)) - F_{\bar{D}}(G^{-1}(\theta, \rho))) + f_{\bar{D}}(G^{-1}(\theta, \rho))\frac{\partial G^{-1}(\theta, \rho)}{\partial \theta}\sqrt{n}(\hat{\theta} - \theta) \\
& \quad + f_{\bar{D}}(G^{-1}(\theta, \rho))\frac{\partial G^{-1}(\theta, \rho)}{\partial \rho}\sqrt{n}(\hat{\rho} - \rho) + o_p(1) \\
& \equiv D_n + E_n + F_n + o_p(1). \tag{28}
\end{aligned}$$

The variance of $\sqrt{n}(F\hat{P}R(\hat{\rho}) - FPR(\rho))$, Σ_2 (equation (13)), depends on the variances and covariances of the three terms

$$\begin{aligned}
D_n & \equiv \sqrt{n}(\hat{F}_{\bar{D}}(G^{-1}(\theta, \rho)) - F_{\bar{D}}(G^{-1}(\theta, \rho))), \\
E_n & \equiv f_{\bar{D}}(G^{-1}(\theta, \rho))\frac{\partial G^{-1}(\theta, \rho)}{\partial \theta}\sqrt{n}(\hat{\theta} - \theta), \\
F_n & \equiv f_{\bar{D}}(G^{-1}(\theta, \rho))\frac{\partial G^{-1}(\theta, \rho)}{\partial \rho}\sqrt{n}(\hat{\rho} - \rho).
\end{aligned}$$

The variance of E_n can be obtained by using Result 1, and the variances of D_n and F_n can be found using Result 2. The covariance between D_n and E_n follows from Lemma 1. Covariances between (D_n, F_n) and (E_n, F_n) follow from an argument similar to that used in deriving $cov(A_n, C_n)$ and $cov(B_n, C_n)$ (equation (19) and (20)).



The asymptotic variance of $\sqrt{n}(\widehat{TG} - \overline{TG})$, σ_{10}^2 , is therefore

$$\sigma_{10}^2 = \Sigma_1 + \Sigma_2 - 2\Sigma_{1,2},$$

where $\Sigma_{1,2}$ is

$$\begin{aligned} \Sigma_{1,2} &\equiv \text{cov}(\sqrt{n}(T\hat{P}R(\hat{\rho}) - TPR(\rho)), \sqrt{n}(F\hat{P}R(\hat{\rho}) - FPR(\rho))) \\ &= \text{cov}(A_n + B_n + C_n, D_n + E_n + F_n) \\ &= \text{cov}(A_n, E_n) + \text{cov}(A_n, F_n) + \text{cov}(B_n, D_n) + \text{cov}(B_n, E_n) + \text{cov}(B_n, F_n) \\ &\quad + \text{cov}(C_n, D_n) + \text{cov}(C_n, E_n) + \text{cov}(C_n, F_n). \end{aligned} \tag{29}$$

All of these covariance terms can be obtained using corresponding Results and Lemmas: $\text{cov}(A_n, E_n)$ and $\text{cov}(B_n, D_n)$ from Lemma 1; $\text{cov}(B_n, E_n)$ from Result 1; $\text{cov}(C_n, F_n)$ from Result 2. $\text{cov}(A_n, F_n)$ and $\text{cov}(B_n, F_n)$ concern the covariance between $(\hat{F}_D, \hat{\rho})$ and $(\hat{\theta}, \hat{\rho})$ and expressions have been derived in the proof of Theorem item (iv) above, cov_2 (equation (19) and (20)), while $\text{cov}(C_n, D_n)$ and $\text{cov}(C_n, E_n)$ concern the covariance between $(\hat{F}_D, \hat{\rho})$ and $(\hat{\theta}, \hat{\rho})$ and are derived using a similar argument.

II. Case-Control Design

Let $\hat{\rho}$ be the estimate of disease prevalence ρ from a cohort independent of the case-control sample, or the parent cohort within which the case-control sample is nested. Assume the

size of the cohort is λ times the size of the case-control sample. Denote

$$\hat{F}^*(t) = \rho \hat{F}_D(t) + (1 - \rho) \hat{F}_{\bar{D}}(t)$$

$$\hat{\theta}^* = \begin{bmatrix} \hat{\theta}_0^* \\ \hat{\theta}_1^* \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{0s} + \log\left(\frac{n_{\bar{D}}}{n_D} \frac{\rho}{1-\rho}\right) \\ \hat{\theta}_{1s} \end{bmatrix}.$$

The following results are well established.

Result 3 Let

$$A_0(t) = \int_{-\infty}^t \frac{\exp(\theta_0^* + \theta_1^* y)}{(1 + k \exp(\theta_0^* + \theta_1^* y))} dF_{\bar{D}}(y)$$

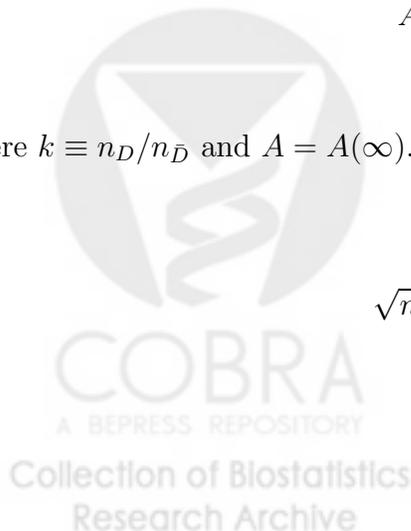
$$A_1(t) = \int_{-\infty}^t \frac{y \exp(\theta_0^* + \theta_1^* y)}{(1 + k \exp(\theta_0^* + \theta_1^* y))} dF_{\bar{D}}(y)$$

$$A_2(t) = \int_{-\infty}^t \frac{y^2 \exp(\theta_0^* + \theta_1^* y)}{(1 + k \exp(\theta_0^* + \theta_1^* y))} dF_{\bar{D}}(y)$$

$$A(t) = \begin{bmatrix} A_0(t) & A_1(t)^T \\ A_1(t) & A_2(t) \end{bmatrix},$$

where $k \equiv n_D/n_{\bar{D}}$ and $A = A(\infty)$. If A^{-1} exists,

$$\sqrt{n} \begin{bmatrix} \hat{\theta}_0^* - \theta_0^* \\ \hat{\theta}_1^* - \theta_1^* \end{bmatrix} \rightarrow_d N(0, \Sigma^{-1}),$$



where

$$\Sigma = \frac{1+k}{k} \left\{ A^{-1} - \begin{bmatrix} 1+k & 0 \\ 0 & 0 \end{bmatrix} \right\}.$$

A proof can be found in Prentice and Pyke (1979), Qin and Zhang (1997) and Zhang (2000).

The next set of results, Results 4-7, have been proved by Huang and Pepe (2008_a).

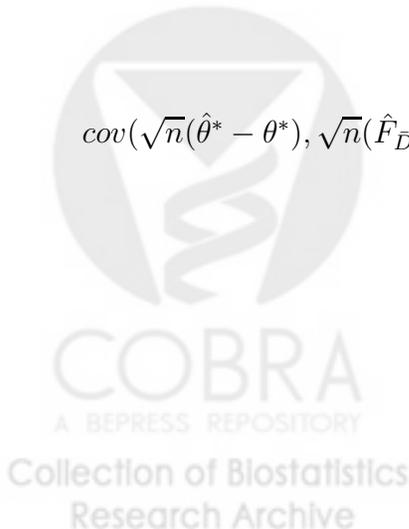
Result 4 As $n \rightarrow \infty$, $\sqrt{n}(\hat{F}^*(t) - F^*(t))$ converges to a normal random variable with mean 0 and variance

$$\sigma_{F^*}^2 = (1-\rho)^2(1+k)F_{\bar{D}}(t)(1-F_{\bar{D}}(t)) + \rho^2 \frac{1+k}{k} F_D(t)(1-F_D(t)).$$

Result 5

$$\text{cov}(\sqrt{n}(\hat{\theta}^* - \theta^*), \sqrt{n}(\hat{F}_D(t) - F_D(t))) = \frac{n}{n_D} \left\{ A^{-1} \begin{bmatrix} A_0(t) \\ A_1(t) \end{bmatrix} - \begin{bmatrix} F_D(t) \\ 0 \end{bmatrix} \right\},$$

$$\text{cov}(\sqrt{n}(\hat{\theta}^* - \theta^*), \sqrt{n}(\hat{F}_{\bar{D}}(t) - F_{\bar{D}}(t))) = \frac{n}{n_{\bar{D}}} \left\{ -A^{-1} \begin{bmatrix} A_0(t) \\ A_1(t) \end{bmatrix} + \begin{bmatrix} F_{\bar{D}}(t) \\ 0 \end{bmatrix} \right\}$$



$$\begin{aligned} \text{cov}(\sqrt{n}(\hat{\theta}^* - \theta^*), \sqrt{n}(\hat{F}^*(t) - F(t))) &= \frac{n}{n_D}(1 - \rho) \left\{ -A^{-1} \begin{bmatrix} A_0(t) \\ A_1(t) \end{bmatrix} + \begin{bmatrix} F_{\bar{D}}(t) \\ 0 \end{bmatrix} \right\} \\ &+ \frac{n}{n_D} \rho \left\{ A^{-1} \begin{bmatrix} A_0(t) \\ A_1(t) \end{bmatrix} - \begin{bmatrix} F_D(t) \\ 0 \end{bmatrix} \right\}. \end{aligned}$$

Result 6

$$\begin{aligned} \text{var}(\sqrt{n}(\hat{\rho} - \rho)) &= \rho(1 - \rho)/\lambda, \\ \text{var}(\sqrt{n}(\hat{\theta} - \theta)) &= \begin{bmatrix} \frac{1}{\lambda\rho(1-\rho)} & 0 \\ 0 & 0 \end{bmatrix} + \text{var}(\sqrt{n}(\hat{\theta}^* - \theta)), \\ \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{\rho} - \rho)) &= \begin{bmatrix} 1/\lambda \\ 0 \end{bmatrix}. \end{aligned}$$

Result 7

$$\begin{aligned} \text{var}(\sqrt{n}(\hat{F}(t) - F(t))) &= (F_D(t) - F_{\bar{D}}(t))^2 \rho(1 - \rho)/\lambda + \text{var}(\sqrt{n}(\hat{F}^*(t) - F(t))), \\ \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}(t) - F(t))) &= \begin{bmatrix} \frac{F_D(t) - F_{\bar{D}}(t)}{\lambda} \\ 0 \end{bmatrix} + \text{cov}(\sqrt{n}(\hat{\theta}^* - \theta), \sqrt{n}(\hat{F}^*(t) - F(t))), \\ \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(t) - F_D(t))) &= \text{cov}(\sqrt{n}(\hat{\theta}^* - \theta^*), \sqrt{n}(\hat{F}_D(t) - F_D(t))), \\ \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_{\bar{D}}(t) - F_{\bar{D}}(t))) &= \text{cov}(\sqrt{n}(\hat{\theta}^* - \theta^*), \sqrt{n}(\hat{F}_{\bar{D}}(t) - F_{\bar{D}}(t))). \end{aligned}$$

Most of the proofs in the following are exactly the same as those developed for a cohort study. Therefore we do not repeat the proofs that are the same. However, expressions for the components of the asymptotic variances that are different are provided. We will frequently refer to expressions in Results 4-7.

Proof of Theorem item (i), (ii) and (iii)

The proof is the same as the proof provided for cohort studies. Based on equation (16),

$$\begin{aligned} \sigma_1(p)^2 = & \text{var}(\sqrt{n}(\hat{F}(G^{-1}(\theta, p)) - F(G^{-1}(\theta, p)))) + \left(\frac{\partial R^{-1}(p)}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R^{-1}(p)}{\partial \theta}\right) \\ & + 2\left(\frac{\partial R^{-1}(p)}{\partial \theta}\right)^T \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}(G^{-1}(\theta, p)) - F(G^{-1}(\theta, p)))). \end{aligned}$$

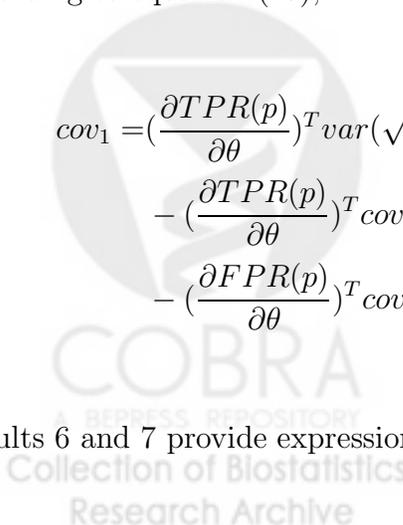
Expressions for the three individual components can all be found in Results 6 and 7. Proofs for items (ii) and (iii) of the Theorem follow similar arguments.

Proof of Theorem items (iv) and (v)

According to equation (17),

$$\begin{aligned} \text{cov}_1 = & \left(\frac{\partial TPR(p)}{\partial \theta}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial FPR(p)}{\partial \theta}\right) \\ & - \left(\frac{\partial TPR(p)}{\partial \theta}\right)^T \text{cov}(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\theta} - \theta)) \\ & - \left(\frac{\partial FPR(p)}{\partial \theta}\right)^T \text{cov}(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\theta} - \theta)). \end{aligned}$$

Results 6 and 7 provide expressions for the three individual terms.



However, the expressions for cov_2 and cov_3 are different from those under a cohort study design,

$$\begin{aligned}
cov_2 &= cov(\sqrt{n}(T\hat{P}R(p) - TPR(p)), \sqrt{n}(\hat{\rho} - \rho)) \\
&= -cov(\sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\rho} - \rho)) \\
&= -cov(\sqrt{n}(\hat{F}_D(G^{-1}(\theta, p)) - F_D(G^{-1}(\theta, p))) + \sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\rho} - \rho)) \\
&= -cov(\sqrt{n}(F_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\rho} - \rho)) \\
&= \left(\frac{\partial TPR(p)}{\partial \theta}\right)^T \times \begin{bmatrix} 1/\lambda \\ 0 \end{bmatrix}
\end{aligned}$$

Similarly,

$$\begin{aligned}
cov_3 &= cov(\sqrt{n}(F\hat{P}R(p) - FPR(p)), \sqrt{n}(\hat{\rho} - \rho)) \\
&= -cov(\sqrt{n}(\hat{F}_D(G^{-1}(\hat{\theta}, p)) - F_D(G^{-1}(\theta, p))), \sqrt{n}(\hat{\rho} - \rho)) \\
&= \left(\frac{\partial FPR(p)}{\partial \theta}\right)^T \times \begin{bmatrix} 1/\lambda \\ 0 \end{bmatrix}
\end{aligned}$$

Item (v) of the Theorem follows from a similar argument.

Proof of Theorem (vi), (vii) and (viii) These all follow similar arguments. We use (vii) to illustrate.

According to equation (21),

$$\begin{aligned} \sigma_7(t)^2 &= \text{var}(\sqrt{n}(\hat{R}(\nu_T(t)) - R(\nu_T(t)))) \\ &= \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^2 \text{var}(\sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)) + \left(\frac{\partial R(\nu_T(t))}{\partial t}\right)^T \text{var}(\sqrt{n}(\hat{\theta} - \theta)) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right) \\ &\quad - 2\left(\frac{\partial R(\nu_T(t))}{\partial t}\right) \text{cov}(\sqrt{n}(\hat{\theta} - \theta), \sqrt{n}(\hat{F}_D(F_D^{-1}(\nu_T(t))) - t)) \left(\frac{\partial R(\nu_T(t))}{\partial \theta}\right). \end{aligned}$$

The result follows by plugging in corresponding expressions from Result 2, 6 and 7. Proofs of items (vi) and (viii) follow similar arguments.

Proof of Theorem item (ix) Proof of Theorem (ix) is exactly the same as the proofs for a cohort study. Equations (22), (24) and (26) defined expressions for the components of the asymptotic variance of $\sqrt{n}(P\hat{E}V - PEV)$, σ_9^2 . We only need to substitute $l(\theta)$ in the definition of $\begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$ and $\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}$ (equation (23) and (25)) with the likelihood function based on case-control data, which are defined by Prentice and Pyke (1979), Qin and Zhang (1997) and Zhang (2000).

Proof of Theorem item (x) According to equations (27), (28) and (29), the three components of $\text{var}(\sqrt{n}(\hat{T}G - \overline{T}G))$, σ_{10}^2 , are

$$\Sigma_1 = \text{var}(A_n) + \text{var}(B_n) + \text{var}(C_n) + \text{cov}(A_n, B_n) + \text{cov}(A_n, C_n) + \text{cov}(B_n, C_n);$$

$$\begin{aligned}\Sigma_2 &= \text{var}(D_n) + \text{var}(E_n) + \text{var}(F_n) + \text{cov}(D_n, E_n) + \text{cov}(D_n, F_n) + \text{cov}(E_n, F_n); \\ \Sigma_{1,2} &= \text{cov}(A_n, E_n) + \text{cov}(A_n, F_n) + \text{cov}(B_n, D_n) + \text{cov}(B_n, E_n) + \text{cov}(B_n, F_n) \\ &\quad + \text{cov}(C_n, D_n) + \text{cov}(C_n, E_n) + \text{cov}(C_n, F_n).\end{aligned}$$

The following Results provide corresponding expressions for each of the individual components:

Result 2: $\text{var}(A_n)$ and $\text{var}(D_n)$.

Result 3: $\text{cov}(B_n, E_n)$.

Result 6: $\text{var}(B_n)$, $\text{var}(C_n)$, $\text{var}(E_n)$, $\text{cov}(B_n, C_n)$, $\text{var}(F_n)$, $\text{cov}(E_n, F_n)$, $\text{cov}(B_n, F_n)$, $\text{cov}(C_n, E_n)$ and $\text{cov}(C_n, F_n)$.

Result 7: $\text{cov}(A_n, B_n)$, $\text{cov}(D_n, E_n)$, $\text{cov}(B_n, D_n)$ and $\text{cov}(A_n, E_n)$.

Furthermore, $\text{cov}(A_n, C_n) = \text{cov}(D_n, F_n) = \text{cov}(A_n, F_n) = \text{cov}(C_n, D_n) = 0$.



Table 1: Results of simulations to evaluate the application of inference based on asymptotic distribution theory and bootstrap resampling to finite sample studies. The study employs cohort design with prevalence 0.2. Marker data for controls are standard normally distributed and for cases is normally distributed with mean 1 and variance 1. Shown are results for $R(v)$, $R^{-1}(p)$, $R(v(\text{TPR}))$ and $R(v(\text{FPR}))$.

		$v=0.1$	$v=0.5$	$v=0.9$	$p=0.1$	$p=0.35$	$p=0.6$	TPR=.9	TPR=.5	TPR=.1	FPR=.9	FPR=.5	FPR=.1
		$R(v)=$			$R^{-1}(p)=$			$R(v(\text{TPR}))=$			$R(v(\text{FPR}))=$		
		0.045	0.154	0.427	0.321	0.839	0.972	0.103	0.292	0.598	0.040	0.132	0.353
% Bias													
n=100		7.15	-1.56	-0.26	3.47	-0.08	-0.85	15.84	5.77	-3.30	11.17	-2.75	-3.43
n=500		1.59	0.06	-0.14	0.62	0.04	-0.18	2.95	0.89	-0.78	2.80	-0.25	-0.60
n=2000		0.21	-0.24	0.09	0.04	0.02	-0.04	0.84	0.30	-0.32	0.79	-0.10	-0.14
Standard deviation													
n=100	observed	0.027	0.045	0.104	0.142	0.074	0.033	0.037	0.086	0.153	0.028	0.043	0.068
	asymptotic	0.027	0.044	0.100	0.136	0.077	0.037	0.036	0.089	0.154	0.026	0.041	0.065
	bootstrap	0.028	0.044	0.104	0.140	0.084	0.050	0.041	0.092	0.152	0.027	0.037	0.064
n=500	observed	0.013	0.021	0.044	0.064	0.035	0.015	0.016	0.037	0.072	0.012	0.019	0.029
	asymptotic	0.012	0.020	0.045	0.063	0.035	0.016	0.016	0.038	0.072	0.012	0.019	0.029
	bootstrap	0.012	0.020	0.046	0.065	0.035	0.016	0.017	0.038	0.073	0.013	0.018	0.028
n=2000	observed	0.006	0.010	0.023	0.032	0.017	0.008	0.008	0.018	0.038	0.006	0.009	0.015
	asymptotic	0.006	0.010	0.023	0.032	0.018	0.008	0.008	0.019	0.037	0.006	0.009	0.015
	bootstrap	0.006	0.010	0.023	0.032	0.018	0.008	0.008	0.019	0.037	0.006	0.009	0.016
95% coverage probability													
n=100	asymptotic	88.7	93.0	92.1	89.6	92.8	90.0	94.8	93.7	87.2	87.6	92.7	93.9
	bootstrap—N	89.4	91.4	95.0	90.2	95.4	96.8	98.2	96.4	88.8	89.3	90.6	92.5
	bootstrap—P	92.0	93.0	96.2	92.6	96.0	96.0	96.8	97.6	93.4	93.8	93.4	95.8
n=500	asymptotic	94.0	94.6	94.4	93.2	94.7	93.1	94.5	94.7	92.4	93.9	94.5	94.4
	bootstrap—N	93.8	94.8	94.8	94.8	95.8	91.8	96.4	95.8	94.0	94.2	92.8	94.8
	bootstrap—P	95.0	94.4	94.2	95.6	95.6	95.0	97.2	95.8	95.8	93.6	96.8	95.4
n=2000	asymptotic	94.9	95.4	95.2	94.4	94.7	95.2	94.8	95.0	94.3	94.7	95.2	94.6
	bootstrap—N	95.0	96.2	95.6	94.2	94.6	94.6	96.0	94.8	96.0	94.8	94.8	95.2
	bootstrap—P	93.8	96.2	95.0	95.0	94.4	95.0	95.6	94.8	95.6	95.0	94.6	94.6

* bootstrap-N: confidence intervals are based on the normal assumption of bootstrapped values

bootstrap-P: confidence intervals are based on percentile of bootstrapped values

Research Archive

Table 2: Results of simulations to evaluate the application of inference based on asymptotic distribution theory and bootstrap resampling to finite sample studies. Data were generated as described for Table 1. Shown are results for $TPR(p)$, $FPR(p)$, $PPV(p)$ and $NPV(p)$.

		$p=0.1$	$p=0.35$	$p=0.6$									
		TPR(p)=			FPR(p)=			PPV(p)=			NPV(p)=		
		0.905	0.395	0.098	0.622	0.103	0.011	0.267	0.490	0.691	0.941	0.856	0.814
% Bias													
n=100		-0.20	-1.09	29.45	-1.48	-0.21	18.82	3.53	-0.05	-14.08	-0.75	0.42	0.73
n=500		0.01	-0.89	5.48	-0.10	-0.58	3.69	0.70	-0.02	-1.88	0.03	0.08	0.15
n=2000		0.01	0.05	1.12	0.02	-0.05	1.83	0.04	-0.05	-0.19	0.02	0.02	0.06
Standard deviation													
n=100	observed	0.074	0.174	0.116	0.158	0.052	0.016	0.053	0.115	0.211	0.045	0.032	0.037
	asymptotic	0.073	0.161	0.117	0.151	0.054	0.019	0.052	0.117	0.213	0.045	0.033	0.036
	bootstrap	0.092	0.168	0.128	0.154	0.067	0.033	0.052	0.117	0.210	0.045	0.030	0.034
n=500	observed	0.029	0.077	0.054	0.073	0.023	0.007	0.022	0.042	0.135	0.015	0.014	0.016
	asymptotic	0.030	0.075	0.054	0.071	0.024	0.008	0.022	0.041	0.132	0.015	0.015	0.016
	bootstrap	0.032	0.078	0.053	0.072	0.024	0.007	0.021	0.040	0.136	0.016	0.014	0.016
n=2000	observed	0.015	0.038	0.027	0.037	0.012	0.003	0.011	0.199	0.058	0.007	0.008	0.008
	asymptotic	0.015	0.038	0.027	0.036	0.012	0.004	0.011	0.020	0.057	0.007	0.007	0.008
	bootstrap	0.015	0.039	0.027	0.037	0.012	0.004	0.011	0.020	0.056	0.007	0.007	0.008
95% coverage probability													
n=100	asymptotic	93.0	90.2	87.6	88.3	93.5	91.1	95.2	96.5	96.4	91.2	93.0	93.2
	bootstrap—N	95.2	93.0	95.2	90.6	97.2	98.2	94.8	96.6	94.6	90.4	92.4	92.3
	bootstrap—P	98.0	97.0	96.8	94.8	98.2	98.6	95.0	96.6	95.0	92.4	93.4	93.5
n=500	asymptotic	93.3	93.8	91.9	93.3	94.9	93.8	94.6	95.1	96.4	92.8	94.6	95.0
	bootstrap—N	96.2	94.4	93.6	95.2	94.6	93.6	94.0	94.0	95.9	93.6	92.4	94.3
	bootstrap—P	97.4	95.0	96.2	95.8	94.8	97.0	94.0	95.4	95.2	94.8	93.2	95.1
n=2000	asymptotic	94.7	94.9	94.2	94.4	94.9	95.3	94.7	94.7	96.0	94.6	95.1	94.6
	bootstrap—N	94.8	96.0	93.8	93.6	95.0	94.6	95.2	95.1	95.3	95.1	95.1	93.8
	bootstrap—P	95.8	96.0	94.4	94.4	94.4	95.2	95.0	94.8	95.8	94.6	95.2	94.7

Table 3: Results of simulations to evaluate the application of inference based on asymptotic distribution theory and bootstrap resampling to finite sample studies. Data were generated as described for Table 1. Shown are results for PEV, standardized total gain, \overline{TG} and AUC.

		PEV=0.154	\overline{TG} =0.383	AUC=0.760
% Bias				
n=100		2.68	1.42	0.25
n=500		1.24	0.48	0.16
n=2000		-0.05	-0.02	0.02
Standard deviation				
n=100	observed	0.078	0.105	0.061
	asymptotic	0.076	0.120	0.060
	Bootstrap	0.078	0.123	0.058
n=500	observed	0.034	0.048	0.027
	asymptotic	0.034	0.051	0.026
	bootstrap	0.034	0.053	0.028
n=2000	observed	0.018	0.025	0.013
	asymptotic	0.017	0.026	0.013
	bootstrap	0.017	0.026	0.014
95% coverage probability				
n=100	asymptotic	91.8	95.1	92.5
	bootstrap—N	90.8	96.8	94.3
	bootstrap—P	93.4	96.4	93.6
n=500	asymptotic	93.9	95.2	94.4
	bootstrap—N	93.4	95.8	95.1
	bootstrap—P	93.6	95.8	94.7
n=2000	asymptotic	94.7	95.5	96.2
	bootstrap—N	95.4	95.3	95.7
	bootstrap—P	95.0	95.6	96.0



Table 4: Results of simulations to evaluate the application of inference based on asymptotic distribution theory and bootstrap resampling to finite sample studies. The study design employs case-control sampling from a parent cohort with prevalence 0.2. Marker data for controls are standard normally distributed and for cases is normally distributed with mean 1 and variance 1. The case-control subset is 1/5 the size of the parent cohort and is randomly selected. Shown are results for $R(\nu)$, $R^{-1}(p)$, $R(\nu(\text{TPR}))$ and $R(\nu(\text{FPR}))$.

	$\nu=0.1$	$\nu=0.5$	$\nu=0.9$	$p=0.1$	$p=0.35$	$p=0.6$	TPR=.9	TPR=.5	TPR=.1	FPR=.9	FPR=.5	FPR=.1
		$R(\nu)=$			$R^{-1}(p)=$			$R(\nu(\text{TPR}))=$			$R(\nu(\text{FPR}))=$	
	0.045	0.154	0.427	0.321	0.839	0.972	0.103	0.292	0.598	0.040	0.132	0.353
% Bias												
n=100	3.39	-1.55	0.86	0.72	0.35	-0.50	6.63	4.11	-0.19	8.72	-1.59	-3.77
n=500	0.55	-0.29	0.38	0.35	0.03	-0.07	1.31	0.78	-0.08	1.95	-0.52	-0.72
n=2000	0.09	-0.06	0.14	0.04	0.01	-0.02	0.37	0.17	0.01	0.67	-0.06	-0.19
Standard deviation												
n=100	observed	0.022	0.027	0.072	0.099	0.047	0.020	0.058	0.117	0.022	0.028	0.045
	asymptotic	0.021	0.028	0.071	0.097	0.050	0.018	0.059	0.119	0.021	0.030	0.047
	bootstrap	0.020	0.022	0.065	0.094	0.046	0.020	0.057	0.117	0.020	0.026	0.044
n=500	observed	0.009	0.012	0.032	0.045	0.021	0.012	0.009	0.025	0.057	0.009	0.013
	asymptotic	0.009	0.012	0.031	0.044	0.021	0.013	0.008	0.025	0.056	0.009	0.013
	bootstrap	0.009	0.011	0.029	0.042	0.019	0.012	0.008	0.022	0.055	0.009	0.012
n=2000	observed	0.005	0.006	0.015	0.024	0.011	0.006	0.004	0.012	0.028	0.004	0.006
	asymptotic	0.005	0.006	0.015	0.022	0.011	0.006	0.004	0.012	0.028	0.005	0.007
	bootstrap	0.005	0.005	0.014	0.021	0.010	0.006	0.004	0.011	0.028	0.005	0.006
95% coverage probability												
n=100	asymptotic	91.7	94.2	94.0	92.8	94.9	91.9	88.0	94.9	90.8	90.7	94.1
	bootstrap—N	90.4	88.2	91.2	91.2	91.2	92.4	95.0	93.6	90.4	90.4	89.8
	bootstrap—P	93.6	90.0	92.4	92.4	93.4	94.0	96.4	91.8	93.4	92.2	93.6
n=500	asymptotic	94.1	95.4	95.1	93.8	94.9	94.9	92.5	95.0	94.4	94.3	94.8
	bootstrap—N	93.3	92.9	93.5	93.6	94.0	94.2	91.4	93.2	93.0	93.6	93.8
	bootstrap—P	93.9	93.1	94.1	94.7	95.4	94.8	93.0	91.6	94.8	94.6	94.2
n=2000	asymptotic	94.9	95.2	94.4	94.8	95.2	95.2	94.3	94.6	94.5	94.7	95.3
	bootstrap—N	94.4	94.6	94.2	95.2	95.4	95.4	93.2	93.8	93.6	94.6	94.8
	bootstrap—P	94.6	95.4	95.6	95.2	94.9	95.4	93.8	94.8	93.6	94.6	95.2

Table 5: Results of simulations to evaluate the application of inference based on asymptotic distribution theory and bootstrap resampling to finite sample studies. Data are generated as described for Table 3. Shown are results for TPR(p), FPR(p), PPV(p) and NPV(p).

		$p=0.1$	$p=0.35$	$p=0.6$										
		TPR(p)=			FPR(p)=			PPV(p)=			NPV(p)=			
		0.905	0.395	0.098	0.622	0.103	0.011	0.267	0.490	0.691	0.941	0.856	0.814	
% Bias														
	n=100	0.22	-0.69	21.42	-0.52	-4.09	3.32	2.07	3.16	-18.73	-0.21	0.14	0.39	
	n=500	0.05	-0.27	4.26	-0.22	-0.41	2.38	0.51	0.75	-0.11	-0.03	0.05	0.09	
	n=2000	0.01	-0.19	1.18	-0.19	-0.26	0.79	0.09	0.11	0.66	0.00	0.01	0.02	
Standard deviation														
	n=100	observed	0.037	0.117	0.090	0.120	0.037	0.019	0.037	0.101	0.189	0.032	0.023	0.020
		asymptotic	0.039	0.114	0.091	0.115	0.041	0.018	0.037	0.113	0.205	0.025	0.022	0.020
		bootstrap	0.037	0.110	0.089	0.112	0.038	0.016	0.038	0.112	0.200	0.028	0.022	0.018
	n=500	observed	0.017	0.053	0.041	0.054	0.017	0.007	0.017	0.042	0.135	0.010	0.010	0.009
		asymptotic	0.016	0.052	0.042	0.052	0.017	0.008	0.016	0.041	0.141	0.009	0.010	0.009
		bootstrap	0.015	0.050	0.041	0.050	0.015	0.007	0.016	0.045	0.132	0.011	0.010	0.008
	n=2000	observed	0.008	0.027	0.021	0.027	0.008	0.004	0.008	0.020	0.065	0.005	0.005	0.005
		asymptotic	0.008	0.026	0.021	0.026	0.008	0.004	0.008	0.020	0.065	0.005	0.005	0.005
		bootstrap	0.007	0.025	0.020	0.025	0.008	0.004	0.008	0.021	0.066	0.005	0.005	0.005
95% coverage probability														
	n=100	asymptotic	92.9	92.5	89.2	92.6	96.7	89.0	94.2	97.5	95.8	90.3	93.3	94.7
		bootstrap—N	94.4	90.6	90.4	91.0	92.2	86.0	93.0	97.2	93.2	94.2	92.8	92.4
		bootstrap—P	97.4	92.2	93.2	92.4	92.6	89.4	92.8	98.2	90.5	95.8	93.6	94.4
	n=500	asymptotic	93.4	94.1	93.1	93.7	96.0	94.2	94.2	95.3	94.6	92.8	94.1	94.9
		bootstrap—N	92.0	93.8	92.2	92.4	93.4	92.4	93.2	96.0	95.2	95.6	92.2	93.6
		bootstrap—P	94.4	94.2	93.0	93.4	93.6	94.0	93.4	97.4	96.2	96.8	92.2	94.8
	n=2000	asymptotic	95.6	93.7	94.4	94.2	95.5	95.1	94.4	95.3	95.1	94.0	94.1	94.8
		bootstrap—N	95.8	94.2	93.4	93.8	94.7	94.2	95.0	94.8	96.2	95.0	94.2	94.6
		bootstrap—P	95.0	94.8	93.2	94.2	95.7	95.7	94.4	96.6	95.6	96.4	94.6	95.2

Table 6: Results of simulations to evaluate the application of inference based on asymptotic distribution theory to finite sample studies. Data are generated as described for Table 3. Shown are results for PEV, standardized total gain, \overline{TG} and AUC.

		PEV=0.154	\overline{TG} =0.383	AUC=0.760
% Bias				
n=100		7.09	0.97	0.05
n=500		1.24	0.56	0.06
n=2000		0.62	-0.13	-0.01
Standard deviation				
n=100	observed	0.064	0.092	0.047
	asymptotic	0.071	0.095	0.047
	Bootstrap	0.064	0.096	0.047
n=500	observed	0.029	0.041	0.021
	asymptotic	0.031	0.042	0.021
	bootstrap	0.028	0.042	0.021
n=2000	observed	0.015	0.021	0.011
	asymptotic	0.016	0.021	0.011
	bootstrap	0.016	0.021	0.011
95% coverage probability				
n=100	asymptotic	95.7	94.8	94.3
	bootstrap—N	93.2	95.2	93.0
	bootstrap—P	92.8	96.8	94.2
n=500	asymptotic	95.8	95.0	94.3
	bootstrap—N	95.4	94.4	95.0
	bootstrap—P	94.4	94.2	94.6
n=2000	asymptotic	96.5	95.2	94.8
	bootstrap—N	96.4	94.6	95.0
	bootstrap—P	95.4	95.0	95.0



Table 7: Point estimates and 95% confidence intervals for the summary indices using FEV₁ and weight as markers of risk for subsequent pulmonary exacerbation in patients with cystic fibrosis. Results based on the entire cohort.

		Estimate	Standard deviation Asymptotic	95% confidence interval		p-value
				Asymptotic	Percentile Bootstrap	
R(0.9)						
	FEV ₁	0.76	0.007	(0.745, 0.773)	(0.746, 0.773)	<0.001
	weight	0.52	0.006	(0.503, 0.537)	(0.503, 0.527)	
R(0.1)						
	FEV ₁	0.14	0.005	(0.129, 0.148)	(0.129, 0.148)	<0.001
	weight	0.24	0.007	(0.221, 0.251)	(0.223, 0.252)	
R⁻¹(0.25)						
	FEV ₁	0.32	0.010	(0.305, 0.344)	(0.303, 0.342)	<0.001
	weight	0.11	0.009	(0.095, 0.131)	(0.095, 0.134)	
R⁻¹(0.75)						
	FEV ₁	0.89	0.007	(0.875, 0.905)	(0.874, 0.906)	<0.001
	weight	1	0	(1, 1)	(1, 1)	
R(□(TPR=0.85))						
	FEV ₁	0.27	0.007	(0.253, 0.280)	(0.254, 0.279)	<0.001
	weight	0.32	0.007	(0.304, 0.330)	(0.305, 0.332)	
R(□(TPR=0.55))						
	FEV ₁	0.53	0.008	(0.511, 0.540)	(0.512, 0.538)	<0.001
	weight	0.47	0.005	(0.456, 0.477)	(0.457, 0.477)	
R(□(FPR=0.15))						
	FEV ₁	0.54	0.008	(0.525, 0.556)	(0.524, 0.554)	<0.001
	weight	0.51	0.006	(0.503, 0.527)	(0.503, 0.527)	
R(□(FPR=0.45))						
	FEV ₁	0.29	0.006	(0.282, 0.304)	(0.283, 0.303)	<0.001
	weight	0.43	0.005	(0.424, 0.443)	(0.424, 0.443)	
TPR(0.25)						
	FEV ₁	0.87	0.007	(0.853, 0.880)	(0.852, 0.879)	<0.001
	weight	0.93	0.007	(0.920, 0.947)	(0.921, 0.946)	
TPR(0.75)						
	FEV ₁	0.21	0.014	(0.185, 0.239)	(0.186, 0.240)	<0.001
	weight	0	0	(0, 0)	(0, 0)	
FPR(0.25)						
	FEV ₁	0.54	0.013	(0.517, 0.569)	(0.519, 0.570)	<0.001
	Weight	0.85	0.011	(0.832, 0.877)	(0.831, 0.875)	
FPR(0.75)						
	FEV ₁	0.039	0.004	(0.032, 0.046)	(0.033, 0.045)	<0.001
	weight	0	0	(0, 0)	(0, 0)	
PPV(0.25)						
	FEV ₁	0.53	0.005	(0.515, 0.535)	(0.513, 0.536)	<0.001
	weight	0.43	0.002	(0.427, 0.435)	(0.428, 0.437)	
PPV(0.75)						
	FEV ₁	0.79	0.011	(0.768, 0.811)	(0.768, 0.814)	<0.001
	weight	0	0	(0, 0)	(0, 0)	
NPV(0.25)						
	FEV ₁	0.83	0.005	(0.820, 0.842)	(0.819, 0.840)	<0.001
	Weight	0.76	0.011	(0.747, 0.780)	(0.746, 0.778)	
NPV(0.75)						
	FEV ₁	0.64	0.003	(0.630, 0.644)	(0.630, 0.645)	<0.001
	weight	0.59	0	(0.590, 0.590)	(0.590, 0.590)	

PEV	FEV ₁	0.22	0.005	(0.202, 0.229)	(0.203, 0.230)	<0.001
	weight	0.05	0.003	(0.041, 0.056)	(0.041, 0.056)	
TG	FEV ₁	0.20	0.006	(0.189, 0.207)	(0.186, 0.208)	<0.001
	weight	0.09	0.005	(0.080, 0.099)	(0.082, 0.101)	
TG	FEV ₁	0.42	0.008	(0.407, 0.440)	(0.407, 0.439)	<0.001
	weight	0.20	0.009	(0.183, 0.218)	(0.182, 0.218)	
AUC	FEV ₁	0.77	0.004	(0.762, 0.779)	(0.762, 0.780)	<0.001
	weight	0.64	0.005	(0.630, 0.649)	(0.630, 0.649)	



Table 8: Point estimates and 95% confidence intervals for the summary indices using FEV₁ and weight as markers of risk for subsequent pulmonary exacerbation in patients with cystic fibrosis. Results are based on prevalence estimated from the entire cohort and marker data from a randomly selected case-control subset with 1,280 cases and 1,280 controls.

		Estimate	Standard deviation Asymptotic	95% confidence interval		p-value
				Asymptotic	Percentile Bootstrap	
R(0.9)	FEV ₁	0.76	0.014	(0.731, 0.785)	(0.735, 0.788)	<0.001
	weight	0.52	0.011	(0.494, 0.535)	(0.496, 0.534)	
R(0.1)	FEV ₁	0.14	0.009	(0.119, 0.155)	(0.117, 0.152)	<0.001
	weight	0.23	0.015	(0.204, 0.263)	(0.197, 0.253)	
R⁻¹(0.25)	FEV ₁	0.35	0.014	(0.318, 0.375)	(0.314, 0.367)	<0.001
	weight	0.12	0.020	(0.086, 0.157)	(0.089, 0.151)	
R⁻¹(0.75)	FEV ₁	0.88	0.013	(0.850, 0.903)	(0.850, 0.903)	<0.001
	weight	1	0	(1, 1)	(1, 1)	
R(□(TPR=0.85))	FEV ₁	0.26	0.008	(0.247, 0.277)	(0.250, 0.278)	<0.001
	weight	0.31	0.008	(0.299, 0.330)	(0.300, 0.331)	
R(□(TPR=0.55))	FEV ₁	0.55	0.014	(0.518, 0.573)	(0.522, 0.573)	<0.001
	weight	0.47	0.009	(0.458, 0.492)	(0.455, 0.490)	
R(□(FPR=0.15))	FEV ₁	0.53	0.010	(0.513, 0.551)	(0.517, 0.550)	0.332
	weight	0.52	0.011	(0.498, 0.541)	(0.500, 0.543)	
R(□(FPR=0.45))	FEV ₁	0.29	0.010	(0.267, 0.308)	(0.266, 0.306)	<0.001
	weight	0.43	0.005	(0.420, 0.441)	(0.421, 0.442)	
TPR(0.25)	FEV ₁	0.86	0.006	(0.848, 0.875)	(0.850, 0.874)	<0.001
	weight	0.93	0.011	(0.911, 0.954)	(0.915, 0.950)	
TPR(0.75)	FEV ₁	0.25	0.027	(0.195, 0.306)	(0.198, 0.310)	<0.001
	weight	0	0	(0, 0)	(0, 0)	
FPR(0.25)	FEV ₁	0.51	0.021	(0.465, 0.552)	(0.476, 0.568)	<0.001
	Weight	0.84	0.027	(0.794, 0.888)	(0.792, 0.885)	
FPR(0.75)	FEV ₁	0.04	0.005	(0.025, 0.045)	(0.027, 0.045)	<0.001
	weight	0	0	(0, 0)	(0, 0)	
PPV(0.25)	FEV ₁	0.54	0.011	(0.519, 0.561)	(0.517, 0.558)	<0.001
	weight	0.43	0.007	(0.423, 0.447)	(0.422, 0.445)	
PPV(0.75)	FEV ₁	0.80	0.018	(0.796, 0.868)	(0.792, 0.870)	<0.001
	weight	0	0	(0, 0)	(0, 0)	
NPV(0.25)	FEV ₁	0.84	0.008	(0.819, 0.854)	(0.819, 0.851)	<0.001
	Weight	0.77	0.016	(0.739, 0.804)	(0.738, 0.806)	
NPV(0.75)						

PEV	FEV ₁	0.65	0.008	(0.633, 0.666)	(0.633, 0.669)	<0.001
	weight	0.59	0	(0.590, 0.590)	(0.590, 0.590)	
TG	FEV ₁	0.21	0.017	(0.181, 0.248)	(0.184, 0.245)	<0.001
	weight	0.04	0.008	(0.022, 0.054)	(0.025, 0.053)	
TG	FEV ₁	0.20	0.009	(0.182, 0.216)	(0.182, 0.216)	<0.001
	weight	0.09	0.009	(0.069, 0.104)	(0.070, 0.104)	
AUC	FEV ₁	0.41	0.019	(0.368, 0.442)	(0.370, 0.444)	<0.001
	weight	0.21	0.020	(0.171, 0.251)	(0.171, 0.252)	
AUC	FEV ₁	0.768	0.009	(0.749, 0.786)	(0.750, 0.787)	<0.001
	weight	0.649	0.011	(0.625, 0.669)	(0.628, 0.670)	



Table 9: Point estimates and 95% confidence intervals for the summary indices using FEV₁ and weight as markers of risk for subsequent pulmonary exacerbation in patients with cystic fibrosis. Results are based on prevalence estimated from the entire cohort and marker data from a randomly selected case-control subset with equal numbers of cases and controls. The total number of cases and controls is denoted by n . Confidence intervals and p -values were based on bootstrap resampling.

		Est. (95% CI)		p-value
		-FEV1	-weight	
R(0.1)	n=500	0.15 (0.100, 0.203)	0.28 (0.212, 0.371)	<0.001
	n=200	0.11 (0.049, 0.179)	0.22 (0.107, 0.335)	0.04
	n=100	0.17 (0.077, 0.283)	0.16 (0.041, 0.299)	0.89
R(0.9)	n=500	0.80 (0.736, 0.862)	0.53 (0.470, 0.590)	<0.001
	n=200	0.69 (0.574, 0.795)	0.49 (0.401, 0.578)	<0.001
	n=100	0.66 (0.502, 0.835)	0.48 (0.337, 0.619)	0.004
R⁻¹(0.25)	n=500	0.29 (0.173, 0.373)	0.12 (0.022, 0.216)	<0.001
	n=200	0.32 (0.125, 0.453)	0.14 (0, 0.286)	0.02
	n=100	0.37 (0.126, 0.516)	0.10 (0, 0.331)	0.09
R⁻¹(0.75)	n=500	0.87 (0.811, 0.934)	1 (1, 1)	<0.001
	n=200	0.85 (0.770, 0.960)	1 (1, 1)	0.005
	n=100	0.86 (0.738, 0.966)	1 (0, 1)	0.61
R(v(TPR=0.85))	n=500	0.25 (0.203, 0.304)	0.32 (0.283, 0.368)	0.002
	n=200	0.27 (0.209, 0.346)	0.36 (0.273, 0.443)	0.008
	n=100	0.25 (0.167, 0.366)	0.32 (0.227, 0.433)	0.10
R(v(TPR=0.55))	n=500	0.53 (0.468, 0.614)	0.48 (0.426, 0.537)	0.04
	n=200	0.56 (0.456, 0.702)	0.46 (0.375, 0.552)	0.05
	n=100	0.61 (0.454, 0.828)	0.51 (0.387, 0.634)	0.16
R(v(FPR=0.15))	n=500	0.58 (0.508, 0.626)	0.50 (0.439, 0.562)	0.006
	n=200	0.55 (0.460, 0.632)	0.50 (0.414, 0.591)	0.25
	n=100	0.48 (0.330, 0.607)	0.55 (0.404, 0.654)	0.42
R(v(FPR=0.45))	n=500	0.28 (0.230, 0.344)	0.42 (0.379, 0.472)	<0.001
	n=200	0.29 (0.204, 0.378)	0.43 (0.357, 0.500)	<0.001
	n=100	0.33 (0.211, 0.444)	0.44 (0.327, 0.539)	0.03
TPR(0.25)	n=500	0.89 (0.837, 0.922)	0.93 (0.865, 0.974)	0.09
	n=200	0.90 (0.845, 0.972)	1 (0, 1)	0.58
	n=100	0.88 (0.731, 1.000)	0.96 (0, 1)	0.81
TPR(0.75)	n=500	0.20 (0.092, 0.320)	0 (0, 0)	<0.001
	n=200	0.26 (0.027, 0.460)	0 (0, 0)	0.08
	n=100	0.22 (0, 0.451)	0 (0, 1)	0.44
FPR(0.25)	n=500	0.58 (0.474, 0.807)	0.82 (0.695, 0.956)	0.003

	n=200	0.57 (0.386, 0.816)	0.83 (0.553, 0.990)	0.03
	n=100	0.72 (0.384, 0.980)	1 (0, 1)	0.53
FPR(0.75)	n=500	0.05 (0.023, 0.072)	0 (0, 0)	<0.001
	n=200	0.05 (0.004, 0.096)	0 (0, 0)	0.32
	n=100	0.06 (0, 0.128)	0 (0, 1)	0.74
PPV(0.25)	n=500	0.50 (0.454, 0.543)	0.44 (0.408, 0.461)	0.01
	n=200	0.49 (0.437, 0.575)	0.43 (0.409, 0.461)	0.05
	n=100	0.46 (0.410, 0.613)	0.40 (0.396, 0.501)	0.29
PPV(0.75)	n=500	0.76 (0.626, 0.857)	0 (0, 0)	<0.001
	n=200	0.72 (0.563, 0.917)	0 (0, 0)	<0.001
	n=100	0.76 (0.555, 0.943)	0 (0, 0)	<0.001
NPV(0.25)	n=500	0.85 (0.807, 0.885)	0.76 (0.696, 0.840)	0.02
	n=200	0.85 (0.753, 0.909)	0.76 (0.448, 0.865)	0.42
	n=100	0.80 (0.598, 0.907)	0.74 (0, 0.859)	0.77
NPV(0.75)	n=500	0.62 (0.593, 0.653)	0.59 (0.590, 0.590)	0.11
	n=200	0.62 (0.590, 0.675)	0.59 (0.590, 0.590)	0.21
	n=100	0.60 (0.587, 0.681)	0.59 (0.590, 0.590)	0.83
PEV	n=500	0.22 (0.158, 0.293)	0.04 (0.016, 0.082)	<0.001
	n=200	0.22 (0.119, 0.337)	0.07 (0.020, 0.148)	0.004
	n=100	0.31 (0.163, 0.480)	0.08 (0.012, 0.215)	0.006
TG	n=500	0.19 (0.142, 0.222)	0.08 (0.030, 0.111)	<0.001
	n=200	0.23 (0.159, 0.281)	0.09 (0, 0.150)	0.001
	n=100	0.29 (0.181, 0.353)	0.09 (-0.007, 0.198)	0.002
\overline{TG}	n=500	0.45 (0.39, 0.54)	0.21 (0.12, 0.29)	<0.001
	n=200	0.48 (0.35, 0.59)	0.25 (0.11, 0.37)	0.004
	n=100	0.56 (0.38, 0.74)	0.24 (0.06, 0.42)	0.004
AUC	n=500	0.76 (0.72, 0.80)	0.63 (0.58, 0.68)	<0.001
	n=200	0.76 (0.69, 0.82)	0.64 (0.56, 0.71)	0.002
	n=100	0.83 (0.75, 0.91)	0.70 (0.61, 0.79)	0.02

Figure 1: Predictiveness curves for FEV₁ (solid curve) and weight (dashed curve) as predictors of the risk of having at least one pulmonary exacerbation in the following year in children with cystic fibrosis. The horizontal line indicates the overall proportion of the population with an event, $\rho = 41\%$. Using the low risk threshold, $p_L = 0.25$, 11% of subjects are classified as low risk according to weight while 32% are classified as low risk according to FEV₁.

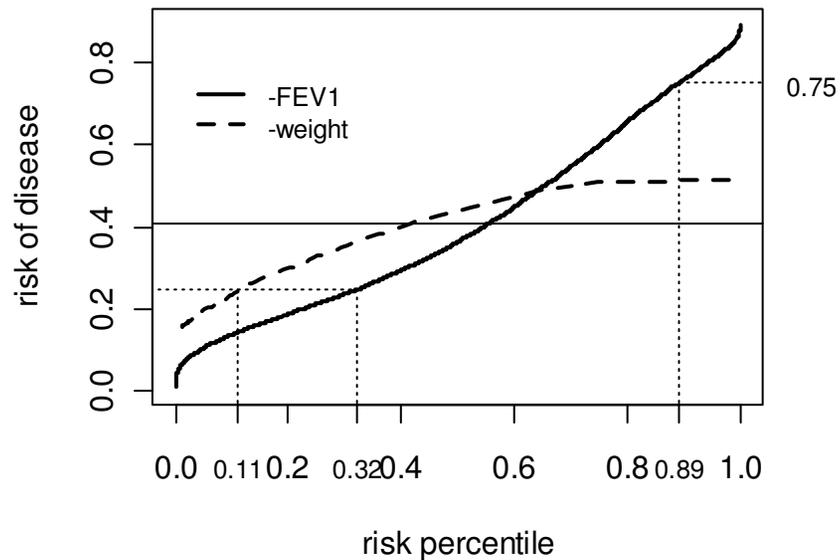


Figure 2: Cumulative distributions of risk based on FEV₁ and weight in predicting the risk of having at least one pulmonary exacerbation in the following year in children with cystic fibrosis. Distributions are shown separately for subjects who had events (cases, solid curve) and for subjects who did not (controls, dashed curve). According to FEV₁, 13% of cases and 46% of controls are classified as low risk, while only 7% of cases and 15% of controls are assigned low risk status according to weight.

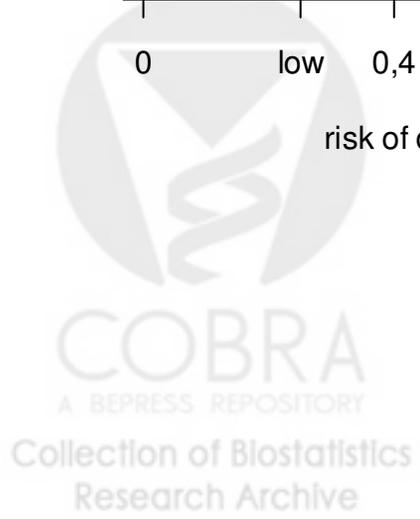
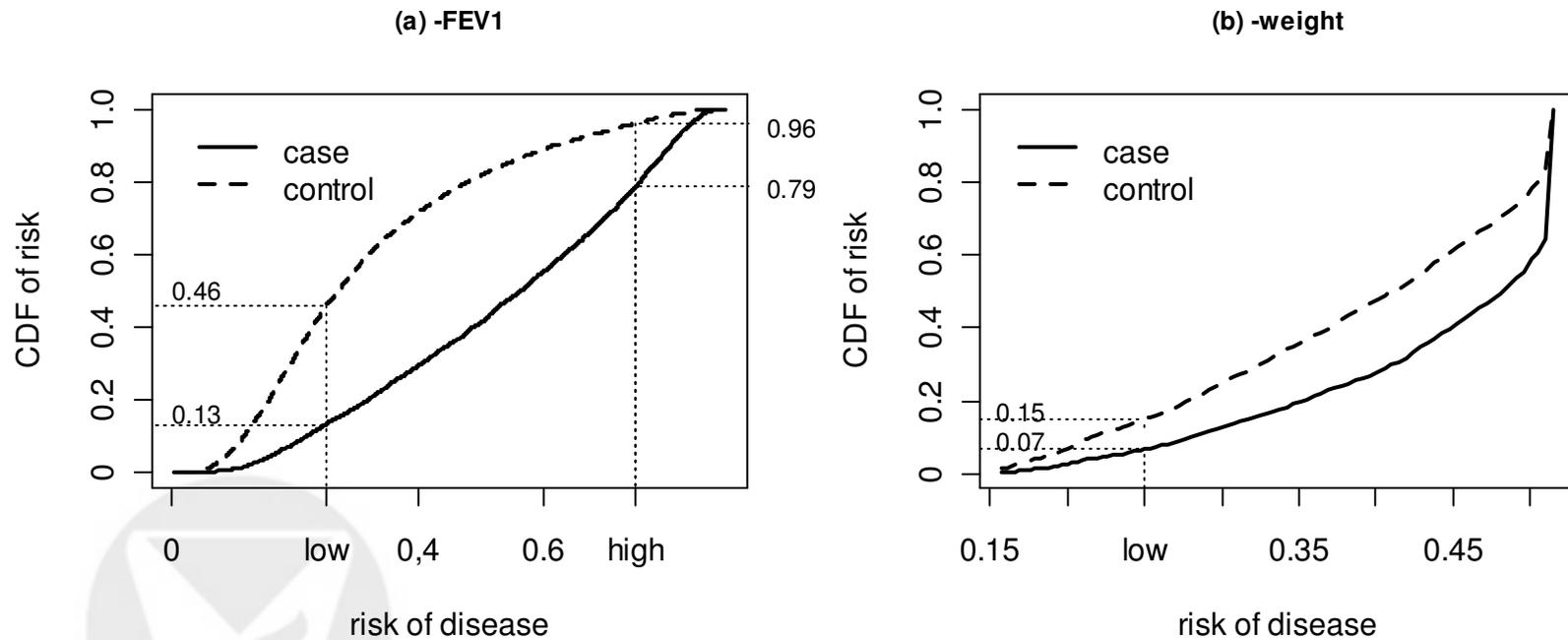


Figure 3: ROC curves for FEV₁ (solid curve) and weight (dashed curve) as predictors of risk of having at least one pulmonary exacerbation in the following year in children with cystic fibrosis. The solid and filled circles are the true and false positive rates corresponding to the low risk threshold $p_L = 0.25$. The areas under the ROC curve are 0.771 for FEV₁ and 0.639 for weight.

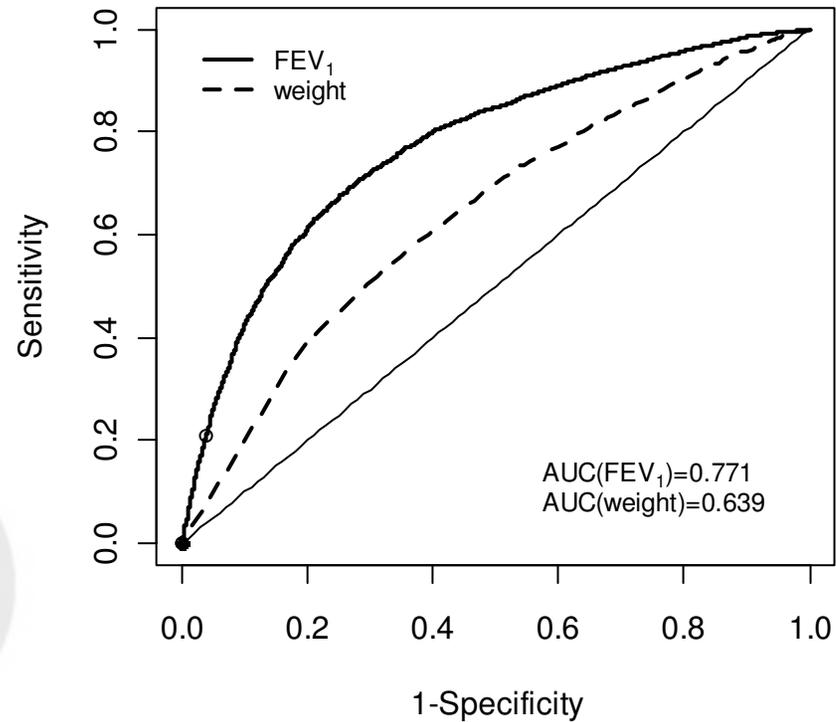


Figure 4: Relationship between the proportion of explained variation, PEV, and the prevalence ρ . A linear logistic risk model with controls standard normally distributed and cases normally distributed with mean 1 and variance 1 was used to generate the data. Maximum PEV occurs at $\rho=0.5$.

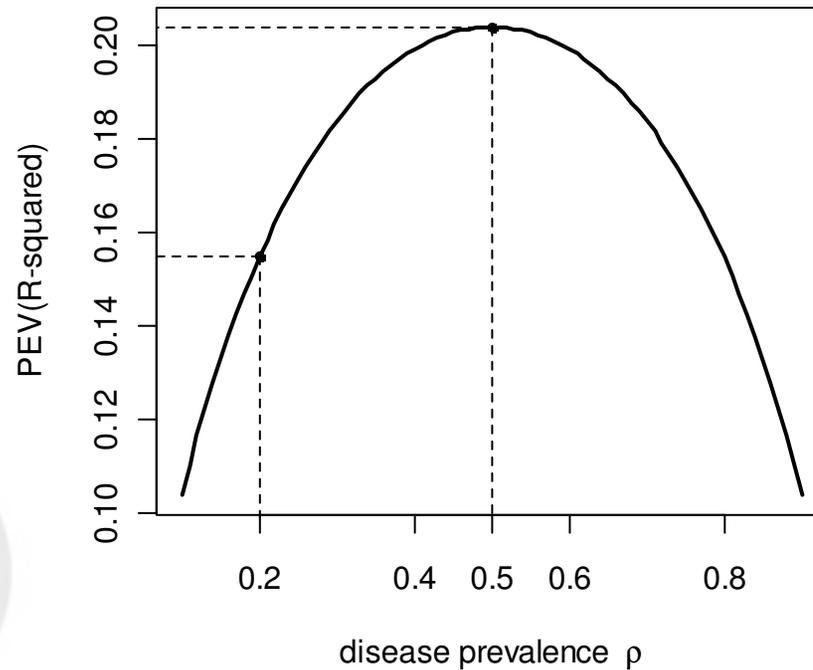


Figure 5: Association between $\{TPR(p) - FPR(p)\}$ and p . A linear logistic risk model with controls standard normally distributed and cases normally distributed with mean 1 and variance 1 was used to generate the data. Overall prevalence of event $\rho=0.2$. Maximum value, also known as the Kolmogorov-Smirnov distance, occurs at $p=\rho$.

