

## STATISTICAL ASSESSMENT OF MULTI-VALUED SCREENING TESTS

MARLOS A. G. VIANA, ANDRÉ. ROGATKO, AND TIMOTHY R. REBBECK

ABSTRACT. When the outcome of a screening test is expressed by the probabilities of  $k$  possible outcomes among individuals with a certain physiologic condition and by the corresponding probabilities among individuals without the condition, the screening usefulness of the test depends on the relative likelihood that its result may properly alter the management of a given patient. New statistical methods are introduced to apply the screening usefulness concept to the assessment of combined or multi-valued tests. The method is applied to assess the usefulness of genotypes at cytochrome P450 IA1 and glutathione-S-transferase- $\mu$  as biomarkers of susceptibility to develop lung cancer. The argument and methods developed in the present paper should be widely applicable to the statistical assessment of screening tests for a wide range of physiologic conditions.

### 1. INTRODUCTION

The goal of screening tests is to identify individuals at risk of a certain physiologic or biologic condition. It is often desirable to assess the conditions under which a test will be useful for screening these individuals. Consider, for example, a genetic biomarker that may identify subjects at increased cancer risk. These markers have been proposed as tools that could be used to screen subjects at risk. Despite numerous associations of these markers with cancer, the usefulness of these biomarkers in screening tests has not been established. Among these cancer biomarkers are the genes involved in the metabolism of environmental carcinogens. These genes include members of the cytochrome P450 and glutathione-S-transferase supergene families.

In one example of this type, Kawajiri, Nakachi, Imai, Watanabe and Hayashi (1993) considered the joint effects of the cytochrome P450 IA1 (CYP1A1) and glutathione-S-transferase- $\mu$  (GST) on susceptibility of cancer. The results summarized in Table 1 show that the six genotypes formed by the two alleles at GST and three alleles at CYP1A1 were significantly associated with lung cancer susceptibility. The conventional statistical analysis shows that individuals who carry both CYP1A1-Val/Val and the GST-null genotypes were found to be at increased risk compared to those who carry the CYP1A1 Ile/Ile and GST-positive genotype (estimated relative risk = 5.44). Those authors also report that risks of specific cancer subtypes (e.g., squamous cell cancers) are also increased in individuals that carry certain genotypes.

The results presented by Kawajiri et al. (1993) and others using similar approaches are aimed to assess the usefulness of the test to *explain* the presence of a physiologic condition. The concepts and methods discussed in the next Section complements the standard case-control methodology and argue the conditions under which susceptible individuals may be identified in a screening (*pre-dictive*) strategy. As it turns out, correlation is a necessary, however not sufficient, condition to ensure good prediction.

---

*Date:* Electronic posting of August 2, 2004.

*1991 Mathematics Subject Classification.* 62F15, 62P10.

*Key words and phrases.* prevalence, screening tests, lung cancer.

We thank the referees and associate editors for valuable comments on an earlier version. The research for this paper was supported in part by a grant from the Illinois Eye Fund and NIH grants GM45979 and CA60798.

In the next section we discuss the statistical argument and methodology to assess a multiple-valued screening tests. Statistical methods to assess the usefulness of binary tests have been described previously by Viana and Farewell (1990, 1994). In particular, we show how the results of a case-control study designed to understand the etiology of a particular disease or physiologic condition, can provide information about disease prediction and identification of at-risk individuals. We use this methodology to assess the conditions under which CYP1A1 and GST may be useful screening tools for lung cancer risk (Section 3). While the example presented here considers genotypes as the risk factors of interest, the argument and corresponding method can be applied in studies of environmental exposures, demographic characteristics, and other biochemical of physiological traits.

## 2. DEFINITIONS AND BASIC RESULTS

The outcome  $T$  of a discrete screening test is a random variable defined for  $t = 1, \dots, k$ . For screening tests, the disease or physiologic condition is often defined as a binary random variable representing the presence or absence of a disease. Most such conditions are time-dependent random processes and their determination are relative to a gold standard, often an invasive procedure such as a tissue biopsy determining a tumor size and its cell morphology. In what follows, we indicate these complex processes by the generic term *disease*, assuming that at a given point in time such processes are well-defined and stable enough to characterize a physiologic condition of interest. We also assume that such characterization is a binary outcome, indicating the presence or absence of the disease.

The sensitivity of a screening test is described by the conditional probabilities  $\eta_t$  of test outcome  $T = t$  among individuals with the disease, whereas the test specificity is described by the conditional probabilities  $\theta_t$  among individuals without the disease or physiologic condition. In addition to the sensitivity and specificity probability distributions, we consider the conditional probabilities  $pv_t$  of disease being present among individuals testing  $T = t$ . The marginal distributions are denoted by  $(\pi, 1 - \pi)$ , where  $\pi$  is the probability of disease being present, and by  $p = (p_1, \dots, p_k)$ , where  $p_t$  is the probability of the event  $T = t$ . Clearly, there is a one-to-one correspondence between the parametric spaces

$$\Omega_T = \{p, pv_1, 1 - pv_1, \dots, pv_k, 1 - pv_k\}, \quad \Omega_D = \{\pi, 1 - \pi, \eta_1, \dots, \eta_k, \theta_1, \dots, \theta_k\}.$$

Either space leads to the joint distribution of  $(T, D)$ . In general, a statistical problem associated with discrete screening tests falls closer to one of these two conditional configurations. In the present paper, we follow the case-control [e.g., Marshall (1988), Zelen and Parker (1986)], or sampling-screening paradigm  $\Omega_D$  [e.g., Sox (1990), Dawid (1976), Boys and Dunsmore (1987), Viana and Ramakrishnan (1992)] in which all parameters in  $\Omega_D$  but the prevalence distribution are estimable from the case-control data.

As an example, consider the lung cancer data from Table 1 and the single screening effect of the three alleles at CYP1A1. The test  $T$  (genotyping at CYP1A1) has  $k = 3$  possible outcomes: Ile/Ile ( $T = 1$ ), Ile/Val ( $T = 2$ ) and Val/Val ( $T = 3$ ). The estimated sensitivity ( $\eta$ ) and specificity ( $\theta$ ) distributions, and corresponding likelihood ratio  $\lambda$  with components  $\lambda_t = \theta_t/\eta_t$  defined for all  $t$

such that  $\eta_t > 0$ , are

$$\begin{aligned}\hat{\eta} &= \frac{1}{181}(161, 87, 34) = (0.57, 0.30, 0.13), \\ \hat{\theta} &= \frac{1}{358}(233, 108, 17) = (0.65, 0.30, .05), \\ \hat{\lambda} &= (1.14, 1.00, 0.33).\end{aligned}$$

The sensitivity and specificity distributions are related to the predictive values pv via the usual Bayes formula. First note that the posterior odds  $O_t$  on the disease when a test result  $T = t$  is observed is a function of  $\pi$ , given by

$$O_t = \frac{1}{\lambda_t} O, \quad t = 1, \dots, k, \quad (2.1)$$

where  $O = \frac{\pi}{1-\pi}$  and often is called the prior (or marginal) odds on the disease. Consequently, the functional relation between  $O_t$  and the predictive value  $pv_t$  is given by  $pv_t = O_t/(1 + O_t)$ . If, for a given group of individuals, the odds on having lung cancer are of 1 to 4, then the resulting posterior odds on the presence of the disease are estimated by

$$\hat{O}_1 = 0.219, \quad \hat{O}_2 = 0.25, \quad \hat{O}_3 = 0.650.$$

Note that whatever the test outcome, the patient management remains the same, since the test predicts the absence of the disease. Should, however, the odds on the disease be of 2 to 3, then

$$\hat{O}_1 = 0.584, \quad \hat{O}_2 = 0.666, \quad \hat{O}_3 = 1.733,$$

which may now properly alter the management of the patient: The test outcome Val/Val ( $T = 3$ ) points to the presence of the disease, against the opposite management when either one of the two other test results obtain.

To assess the usefulness of a multiple-valued screening test, a comparison is made between  $O_s$  and  $O_t$  at each prevalence value  $\pi$ . Since the odds of 1 to 1 is equivalent to 50% in probability, a value of  $O_t > 1$  corresponds to more than 50% in probability and suggests (the action of) advising the patient to further testing. Conversely, a value of  $O_s < 1$  suggests the opposing action (excluding the presence of the disease). As pointed out by Sox (1990), p. 28, a screening test should be obtained only when the outcome of the test could alter the management of the patient. In this sense, any two tests' outcomes  $T = s$  and  $T = t$  are (minimally) useful at the prevalence level  $\pi$  when the event

$$\min\{O_s, O_t\} < 1 < \max\{O_s, O_t\},$$

holds true. For each test outcome pair  $s, t$  we find an interval of prevalence values for which the above equation holds. The union of all such intervals as  $s, t$  varies, is called the *proper region* of the screening test, and denoted by  $R$ . Let

$$U = \min_t\{\lambda_t\}, \quad V = \max_t\{\lambda_t\}.$$

It then follows that

$$R = \left\{ \pi; \frac{U}{U+1} < \pi < \frac{V}{V+1} \right\}. \quad (2.2)$$

There is a connection between the test's proper region  $R$  and the test's relative likelihood ratio

$$\Delta = \frac{V}{U}, \quad (2.3)$$

which expresses the relative weighting of evidence intrinsic to the screening device (see also Kass and Raftery (1995)). Given a test with proper region  $R$ , the log relative likelihood ratio

$$H = \log \Delta$$

equals the constant height  $H$  between the two assessment curves  $\log O_U(\pi)$  and  $\log O_V(\pi)$ . It is important to note that  $\Delta$  and  $H$  are constant functions of the prevalence  $\pi$ . From the definitions (2.2) and (2.3) it follows that any two tests' log relative likelihood ratios  $H_1$  and  $H_2$  are monotonically related with their proper regions  $R_1$  and  $R_2$  in the sense that

$$R_2 \subseteq R_1 \Rightarrow H_2 \leq H_1. \quad (2.4)$$

The converse of (2.4) is not true. In fact, take any two proper regions  $R_1$  and  $R_2$  such that neither  $R_1 \subseteq R_2$  or  $R_2 \subseteq R_1$ . Then, because either  $H_1 \leq H_2$  or  $H_2 \leq H_1$  (the real line together with  $\leq$  is a completely ordered set), the converse fails for all such regions  $R_1$  and  $R_2$ .

A screening test with proper region  $R_1$  is said to be *at least as useful as* another test with proper region  $R_2$  if  $R_2 \subset R_1$ . When the set of all screening tests is identified with the corresponding set of proper regions (2.2) we obtain an ordered set (however not a completely ordered one).

### 3. STATISTICAL ASSESSMENT

The data associated with a screening test with  $k$  outcomes conditioning on disease status ( $\Omega_D$ ) consist of the observed frequencies  $\mathbf{x}$  of test results among  $n$  individuals with the disease and the observed frequencies  $\mathbf{y}$  of test results among  $m$  individuals without the disease. The numbers  $n$  and  $m$  are selected or restricted by practical considerations only and do not have to correspond to the distribution of cases and controls in any selected group. The likelihood on the data  $\{\mathbf{x}, \mathbf{y}\}$  is the product of the corresponding multinomial conditional probabilities  $P(\mathbf{x} \mid \eta_1, \dots, \eta_k)$  and  $P(\mathbf{y} \mid \theta_1, \dots, \theta_k)$ .

**3.1. Binary-valued tests.** The statistical assessment of a binary test's usefulness at a given prevalence level  $\pi$  is the posterior probability of the event  $\pi \in R$ , where we understand that  $R$  is a random interval and  $\pi$  is fixed. It then follows that  $P[\pi \in R] = P[U < O < V] = 1 - P[V < O] - P[O < U]$ . From the definitions of  $U, V$  and  $O$  it follows that  $R$  includes the prevalence level  $\pi$  with probability

$$\gamma = 1 - P[\pi < \pi\eta + (1 - \pi)\theta < 1 - \pi] - P[1 - \pi < \pi\eta + (1 - \pi)\theta < \pi], \quad \eta, \theta, \pi \in (0, 1). \quad (3.1)$$

Given the data  $\{\mathbf{x}, \mathbf{y}\}$  and conjugate prior (independent) probability densities for  $\eta$  and  $\theta$ , the probability  $\gamma$  is assessed with respect to the posterior probability  $P$  for  $\eta, \theta$ , at each prevalence value  $\pi$ .

The required condition for a binary test  $R_1$  (we use the same notation to identify the test and its proper region) to be at least as useful as another binary-valued test  $R_2$  is that  $R_1 \subset R_2$ . More precisely, if a binary-valued test  $R_1$  has likelihood ratios  $\lambda_{11} \geq \lambda_{12}$  and test  $R_2$  has likelihood ratios  $\lambda_{21} \geq \lambda_{22}$ , then  $R_1$  is at least as useful as  $R_2$  with probability

$$P[R_2 \subset R_1] = \min\{P[\lambda_{12} < \lambda_{11}], P[\lambda_{22} > \lambda_{21}]\}. \quad (3.2)$$

In fact,

$$\begin{aligned} P[R_2 \subset R_1] &= P[\{\lambda_{12} < \lambda_{22}\} \cap \{\lambda_{21} < \lambda_{11}\}] = P[\{\theta > 1 - \eta\lambda_{22}\} \cap \{\theta > (1 - \eta)\lambda_{21}\}] \\ &= P[\theta > \max\{1 - \eta\lambda_{22}, (1 - \eta)\lambda_{21}\}] = \min\{P[\lambda_{12} < \lambda_{22}], P[\lambda_{21} < \lambda_{11}]\}. \end{aligned}$$

The above result extends to the comparison of several binary-valued tests against a standard one. Given a test  $R_1$  with ratios  $\lambda_{11} \geq \lambda_{12}$  and a family  $\{R_j\}$  of competing tests with ratios  $\lambda_{j1} \geq \lambda_{j2}$ ,  $j = 2, \dots, k$ , then  $R_1$  is at least as useful as  $\{R_j\}$  with probability

$$P\left[\left(\bigcup_{j>1} R_j\right) \subset R_1\right] = \min\{\min_j P[\lambda_{12} < \lambda_{j2}], \min_j P[\lambda_{j1} < \lambda_{11}]\}. \quad (3.3)$$

Given the data  $\{\mathbf{x}, \mathbf{y}\}_j$  from each screening test and conjugate prior probability densities for  $(\eta, \theta)_j$ , expressions (3.2) and (3.3) are assessed with respect to the product  $P = \prod_j P_j$  of independent posterior probabilities  $P_j$  for  $(\eta, \theta)_j$ .

**3.2. Many-valued test outcomes.** Let  $\eta \sim D(n_1, \dots, n_k)$  and  $\theta \sim D(m_1, \dots, m_k)$  indicate posterior Dirichlet distributions of the test sensitivity and specificity, respectively. The question of interest is the assessment of the test's proper region  $R$ , as in (2.2), which is determined by  $U = \min\{\lambda_1, \dots, \lambda_k\}$  and  $V = \max\{\lambda_1, \dots, \lambda_k\}$ . The following result (see the Appendix for an outline of the proof), describes the probability with which the proper regions are determined.

**Proposition 3.1.** If  $\theta$  has a Dirichlet distribution  $D(m_1, \dots, m_k)$ , independent of  $\eta$  with distribution  $D(n_1, \dots, n_k)$  then

$$P[U = \lambda_t] = \frac{n_t}{m_t} \int_0^\infty \left[ \prod_{j \neq t} \left(1 - F_{2m_j, 2n_j}\left(\frac{n_j}{m_j} y\right)\right) \right] f_{2m_t, 2n_t}\left(\frac{n_t}{m_t} y\right) dy,$$

$$P[V = \lambda_t] = \frac{n_t}{m_t} \int_0^\infty \left[ \prod_{j \neq t} F_{2m_j, 2n_j}\left(\frac{n_j}{m_j} y\right) \right] f_{2m_t, 2n_t}\left(\frac{n_t}{m_t} y\right) dy,$$

where  $F_{2m_j, 2n_j}$  indicates the cumulative F distribution with degrees of freedom  $2m_j$  and  $2n_j$  and  $f_{2m_t, 2n_t}$  indicates the corresponding density function.

When  $k = 2$ , the distributions of  $U$  and  $V$  can be obtained from the following result, also derived in the Appendix:

**Proposition 3.2.** Let  $\eta \sim F_\eta$  and  $\theta \sim F_\theta$  indicate the (continuous) probability distributions of  $\eta$  and  $\theta$ . Then,  $0 < U < 1 < V$ , with probability one, and

$$P[V < v] = \frac{1}{v} \int_0^1 F_\theta(t) \left\{ f_\eta\left(\frac{v-t}{v}\right) - f_\eta\left(\frac{1-t}{v}\right) \right\} dt + F_\eta(1 - 1/v), \quad v > 1,$$

whereas

$$P[U < u] = 1 - \int_0^1 \{F_\theta(1 - ut) - F_\theta[u(1 - t)]\} f_\eta(t) dt, \quad u < 1.$$

Given the data  $\{\mathbf{x}, \mathbf{y}\}$  and conjugate prior (independent) probability densities for  $\eta$  and  $\theta$ , the probabilities described in Propositions 3.1 and 3.2 are assessed with respect to the corresponding probability Dirichlet densities.

## 4. APPLICATION TO CANCER SUSCEPTIBILITY GENOTYPES

Kawajiri et al. (1993) reported that lung cancer risk is increased in individuals who carry specific genotypes at CYP1A1 and GST. For example, individuals who carry both CYP1A1-Val/Va/ and the GST-null genotype were found to be at increased lung cancer risk compared with those who carry the CYP1A1 lle/lle and GST-positive genotype (estimated relative risk = 5.44). Those authors also report that risks of specific cancer subtypes (e.g., squamous cell cancers) are also increased in individuals that carry certain genotypes. The study included 358 healthy controls, 282 cigarette smoking associated lung cancer patients, 59 undifferentiated cell carcinoma patients, 128 adenocarcinoma patients and 95 squamous cell carcinoma patients. Table 1 shows the corresponding distribution of combined genotyping.

Consider first the CYP1A1 genotype alone. Note that individuals with the genotype Val/Val were at relatively high risk (estimated risk ratio of 2.89 relative to ILe/ILe) to lung cancer. To assess the usefulness of CYP1A1 to identify susceptible individuals, we first need to determine its proper region. CYP1A1 acts as a multiple valued test with  $k = 3$  possible outcomes. Starting with the corresponding likelihood ratios  $\lambda_t$  shown in Table 2, obtained directly from Table 1. For example, in the lung cancer group, the estimated value of  $\lambda_1$  is  $[(127 + 106)/358]/[(67 + 94)/282] = 1.140$ . The estimated extreme values (2.2) of  $\{\lambda_1, \lambda_2, \lambda_3\}$  are  $U = \lambda_3 = 0.391$  and  $V = \lambda_1 = 1.140$ . Transforming from the odds scale back to the probability scale we obtain the estimated enzyme's proper region for lung cancer, namely  $\hat{R} = (\hat{\pi}_1, \hat{\pi}_2) = (0.281, 0.532)$ . The estimate of the log relative likelihood ratio  $H_{u,v} = \log U/V$  is  $\hat{R} = 1.07$  and equals the estimated height between the parallel assessment extreme curves  $\max_t \log O_t(\pi)$  and  $\min_t \log O_t(\pi)$ . Note that  $e^{1.07} = 2.89$ , the observed risk ratio of Val/Val relative to ILe/ILe.

If the risk ratio determines the genotype's usefulness to explain the inter individual difference in cancer susceptibility, its usefulness to identify susceptible individuals is determined primarily by the estimated enzyme's proper region  $\hat{R} = (0.281, 0.532)$  and only partially by  $H$ . As the lung cancer prevalence among prospective individuals falls below 0.28 or exceeds 0.532, the enzyme becomes less useful to identify susceptible individuals compared to when the lung cancer prevalence fell within the proper range.

To conclude the analysis, we evaluated the genotype's predictive values  $pv_t = O/(O + \lambda_t)$  at certain lung cancer prevalence levels. We selected  $\pi = 0.406$ , the prevalence value at the center of the proper region,  $\pi = 0.14$ , the value half-way to the left of the lower end of the observed proper region  $R$ , and  $\pi = 0.766$ , the prevalence value half-way to the right of the upper end of  $R$ . Since the outcome Val/Val ( $t=3$ ) was the *most positive* outcome we considered  $pv_3(o)$  in contrast to  $pv_1(o)$ , the *most negative* test outcome. At  $\pi = 0.14$ , the estimated  $pv_3$  is 0.29 whereas the estimated  $pv_1$  is 0.12: Both genotypes Val/Val or ILe/Ille suggest the absence of lung cancer. At  $\pi = 0.766$  the corresponding estimates are 0.89 and 0.74: Both genotypes Val/Val or ILe/Ille suggest the presence of lung cancer. However, at  $\pi = 0.406 \in \hat{R}$  we have the estimates 0.63 and 0.37, respectively: The genotype Val/Val suggests the presence of lung cancer whereas the opposing result ILe/Ille properly suggests the opposing action. By definition, this is true for each prevalence level within the test's proper region  $R$ .

The CYP1A1 genotype's usefulness in predicting the other types of cancer was assessed in a similar manner. The proper regions and log relative likelihood shown in Table 2 indicated a relatively equivalent usefulness among all types of cancer considered. The enzyme was slightly better in identifying undifferentiated cell carcinomas cases: the estimated proper region (0.258, 0.545) was the most inclusive of all regions and corresponded to the largest value  $\hat{H} = 1.23$  of log relative likelihood value  $H$  (associated estimated risk ratio of Val/Val relative to ILe/ILe:  $e^{1.23} = 3.42$ ).

We also assess the joint predictive effect of genotyping the two genotypes as a multi-valued test of dimension  $k = 6$ . Table 3 shows the likelihood ratios  $\lambda_t$ , extreme prevalence regions  $(\pi_1, \pi_2)$ , and corresponding log relative likelihood ratios  $H$ . The estimated lung cancer lower prevalence limit was 0.215, corresponding to the Val/Val and nulled GST1 combined outcomes ( $t=6$ ). Using Proposition 3.1, the statement for the lower limit was true with posterior probability 0.923. The complete set of posterior probabilities is shown in Table 4. Similarly, the estimated upper prevalence limit was 0.598, with posterior probability 0.871. This corresponded to the combined event ILe/ILe and positive GST1 ( $t=1$ ). When the prevalence of lung cancer rested outside the estimated proper region (0.215, 0.598) the combined CYP1A1-GST1 test was less useful. Note that the observed likelihood ratio ordering  $\lambda_6 < \lambda_5 < \lambda_2 < \lambda_3 < \lambda_4 < \lambda_1$  also estimates the profile of least positive to most positive test result in the combined scale of the two enzymes. Figure 1 shows the lower ( $t=1$ ) and upper ( $t=6$ ) assessment curves of combined CYP1A1 and GST1 for lung cancer.

Comparing the extreme proper regions for the GST1 alone and the combined CYP1A1-GST1 genotypes lead to the following conclusions: for lung cancer, the combined test is more useful than the single test, because the proper region of the combined test includes the proper region of the single test. Note that the estimated lower prevalence limit for the combined test was 0.21. The corresponding log likelihood ratio was  $\hat{H} = 1.69$ . In fact, the uniform improvement held for all the other types of cancer considered in the sense that the maximal proper region of the combined test includes the single-enzyme proper region. The most improvement was obtained for the squamous cell type in which the estimated lower prevalence limit dropped from 0.27 to 0.17. The log likelihood ratio was  $\hat{H} = 2.07$ , which corresponds to the estimated odds ratio of  $e^{2.07} = 7.9$  reported by the authors. Table 4 shows that the lower prevalence limits of the maximal regions was estimated with probability at least 0.9 for all the types of cancer but the adenocarcinoma, in which the largest probability was only 0.54. The upper limits were less certain. The best estimate was 0.87 for lung cancer, followed by 0.74 for squamous cell. A possible interpretation is the fact that the upper limits are typically closer to 0.5 and hence subject to larger variability.

## 5. DISCUSSION

In the present study, we have examined the usefulness of genotypes at cytochrome P450IA1 (CYP1A1) and glutathione-S-transferase  $\mu$  (GST) to determine susceptibility to develop lung and other cancers. Our results indicate that CYP1A1 is unlikely to provide a useful index of lung cancer susceptibility by itself. The joint genotypes of CYP1A1 and GST may provide a more useful measure of lung cancer susceptibility than the CYP1A1 genotype alone. However, these genotypes CYP1A1 and GST may be only useful as screening tools for lung cancer susceptibility in high-prevalence groups, and not in groups with cancer prevalence less than 20%.

These results are consistent with our knowledge about the effects of these genes on carcinogen metabolism and cancer susceptibility. In the monooxygenation of environmental carcinogens by the

cytochrome P450, substrates are converted into biologically active forms by the phase I detoxification enzymes. The substrates for cytochrome P450 include carcinogenic compounds found in the diet or work place. The genotoxic electrophilic compounds created by P450 enzyme activity are converted to excretable hydrophilic compounds by enzymes such as GST, which accomplishes this conversion by conjugating a glutathione moiety to an electrophilic substrate. Therefore, genotypes associated with increased cytochrome P450 enzyme activity may increase cancer susceptibility, and decreased GST enzyme activity may increase cancer susceptibility. Furthermore, the combination of two or more of these genes may interact to have an even more profound impact on cancer risk. In the present, we have introduced a statistical methodology that can be used to assess under which conditions a diagnostic or screening test may be useful. This test may have a variety of applications in clinical epidemiology, screening strategies, and disease prevention. While we have applied this methodology to specific genetic effects in cancer, this approach should be equally applicable to the study of other epidemiologic risk factors in a wide variety of diseases [e.g., Rebbeck, Viana, Jordan, Weber and Rogatko (1996)].

We have shown that when a screening test with outcomes  $T = 1, \dots, t$  is assessed relative to the presence ( $D = 1$ ) or absence ( $D = 0$ ) of a disease or physiologic condition of interest, its usefulness can be expressed in terms of the test's proper region  $R$ , determined in the odds scale by the extreme points of the set

$$\{\lambda_1, \dots, \lambda_k\}.$$

Only when the prevalence of the disease is within  $R$  the test becomes properly useful, in the sense that the predictive probabilities of disease are expected to properly alter the management of the patient.

We pointed to the fact that the assessment of screening tests should include the assessment of the test's proper region. In fact, let

$$M(\pi) = \begin{bmatrix} \eta\pi & (1-\eta)\pi \\ (1-\pi)(1-\theta) & \theta(1-\pi) \end{bmatrix}, \quad \eta, \theta, \pi \in (0, 1),$$

indicate the parametric joint structure of a disease binary outcome  $D$  and a test binary outcome  $T$ . Let also

$$g(\pi) = \eta\pi - [\eta\pi + (1-\eta)(1-\pi)]\pi = \pi(1-\pi)(\eta + \theta - 1),$$

and  $f(\pi) = \det M(\pi) = \pi(1-\pi)(1-\eta)(1-\theta)(\Delta - 1)$ . Then, the following are equivalent: [a]  $D$  and  $T$  are independent, [b]  $\eta + \theta = 1$ , [c]  $\pi_1 = \pi_2$ , [d]  $g(\pi) \equiv 0$ . Moreover, the following are also equivalent: [e]  $\Delta = 1$ , [f]  $H = 0$ , [g]  $f(\pi) \equiv 0$ . However, [a]  $\rightarrow$  [e], and not conversely. This shows, along with the argument following (2.4), that proper region ( $R$ ) assumptions are weaker assumptions in comparison with relative risk ( $H$ ) assumptions.

## 6. APPENDIX

*Proof of Proposition 3.1.* Let  $(e_1, \dots, e_k)/\sum e_i$  indicate the representation of  $\theta \sim D(m_1, \dots, m_k)$  based on  $e_i \sim \text{Gamma}(\alpha, m_i)$ ,  $i = 1, \dots, k$ . Similarly, let  $(f_1, \dots, f_k)/\sum f_i$  indicate the representation of  $\eta \sim D(n_1, \dots, n_k)$  based on  $f_i \sim \text{Gamma}(\alpha, n_i)$ ,  $i = 1, \dots, k$ . Then the event  $U = \theta_t/\eta_t$  is equivalent to the event  $e_t/f_t = \min\{e_1/f_1, \dots, e_k/f_k\}$ , which, because  $(e_1, \dots, e_k, f_1, \dots, f_k)$  are

jointly independent, has probability

$$\int_0^\infty \prod_{j \neq t} (1 - H_j(q)) h_t(q) dq,$$

where  $H_j$  is the cumulative distribution of  $e_j/f_j$  and  $h_j$  its corresponding density. The result follows from the fact that  $H_j(q) = F_{2m_j, 2n_j}(n_j q/m_j)$ . The derivation of  $P[V = \theta_t/\eta_t]$  is similar.  $\square$

Note that in the fully symmetric case the above integral reduces to  $\int_0^1 u^{k-1} du = \frac{1}{k}$ , as expected.

*Proof of Proposition 3.2.* First we consider the distribution of  $V$ ;

$$P[V < v] = P\left[\frac{\theta}{1-\eta} < v, \frac{1-\theta}{\eta} < v\right] = P[1 - \eta v < \theta < v(1 - \eta)] = I_1 - I_2,$$

where

$$I_1 = \int [F_\theta[v(1 - \eta)] f_\eta(\eta) d\eta, \quad I_2 = \int [F_\theta(1 - v\eta)] f_\eta(\eta) d\eta.$$

However, because  $U < 1 < V$  with probability one, note that  $0 < F_\theta[v(1 - \eta)] < 1$  when  $0 < v(1 - \eta) < 1$ , or  $1 - 1/v < \eta < 1$ , so that

$$I_1 = \int_{1-1/v}^1 F_\theta[v(1 - \eta)] f_\eta(\eta) d\eta + \int_0^{1-1/v} f_\eta(\eta) d\eta,$$

which can be expressed, transforming  $t = v(1 - \eta)$ , as

$$I_1 = \frac{1}{v} \int_0^1 F_\theta(t) f_\eta\left(\frac{v-t}{v}\right) dt + F_\eta(1 - 1/v).$$

To evaluate  $I_2$  note that  $0 < F_\theta(1 - v\eta) < 1$  when  $0 < 1 - v\eta < 1$ , or  $0 < \eta < 1/v$ . Consequently,

$$I_2 = \int_0^{1/v} F_\theta(1 - v\eta) f_\eta(\eta) d\eta = \frac{1}{v} \int_0^1 F_\theta(t) f_\eta\left(\frac{1-t}{v}\right) dt,$$

transforming  $t = 1 - v\eta$ . Because  $F(v) = I_1 - I_2$ , the proof is complete. A similar argument can be used to obtain  $F_u$ .  $\square$

TABLE 1. Distribution of combined genotyping of CYP1A1 and GST1 enzymes in healthy controls and lung cancer (L); undifferentiated cell carcinoma (UCC); adenocarcinoma (A) and squamous cell carcinoma (SC) patients.

t	CYP1A1	GST1	L	UCC	A	SC	controls
1	ILe/ILe	+	67	18	29	20	127
2	ILe/ILe	-	94	14	44	36	106
3	ILe/Val	+	46	8	25	13	55
4	ILe/Val	-	41	11	16	14	53
5	Val/Val	+	11	2	7	2	9
6	Val/Val	-	23	6	7	10	8
			282	59	128	95	358

TABLE 2. Likelihood ratios  $\lambda_t$ , extremal prevalence region  $\pi_1, \pi_2$  and log relative likelihood ratio H for CYP1A1 genotyping: lung cancer (L); undifferentiated cell carcinoma (UCC); adenocarcinoma (A); squamous cell (SC) relative to control subjects

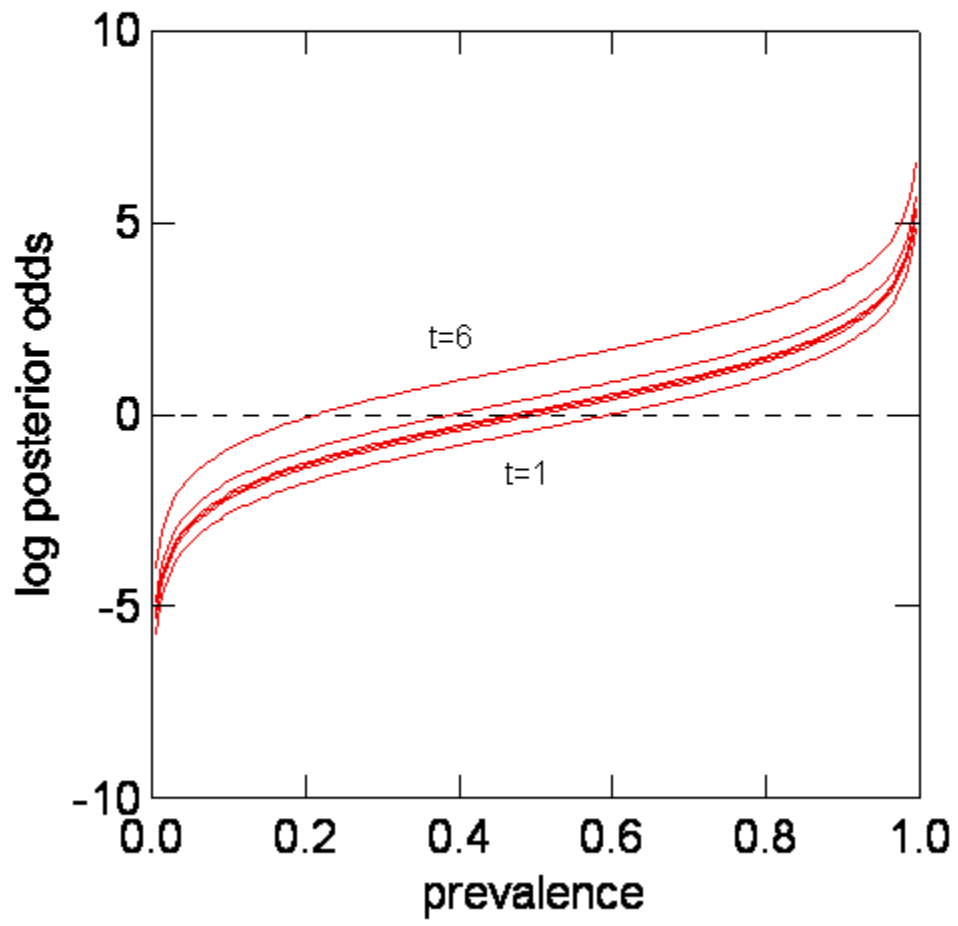
t	CYP1A1	L	UCC	A	SC
1	ILe/ILe	1.140	1.199	1.140	1.103
2	ILe/Val	.977	.934	.940	1.059
3	Val/Val	.391	.348	.431	.373
$\pi_1$		.281	.258	.301	.271
$\pi_2$		.532	.545	.532	.524
H		1.07	1.23	.56	.65

TABLE 3. Likelihood ratios  $\lambda_t$ , extremal prevalence region ( $\pi_1, \pi_2$ ) and log relative likelihood ratio H for combined genotyping of CYP1A1 and GST1: lung cancer (L); undifferentiated cell carcinoma (UCC); adenocarcinoma (A); squamous cell (SC) relative to control subjects

t	CYP1A1	GST1	L	UCC	A	SC
1	ILe/ILe	+	1.493	1.163	1.566	1.685
2	ILe/ILe	-	.888	1.248	.861	.781
3	ILe/Val	+	.942	1.133	.787	1.123
4	ILe/Val	-	1.018	.794	1.184	1.005
5	Val/Val	+	.644	.742	.460	1.194
6	Val/Val	-	.274	.220	.409	.212
$\pi_1$			.215	.180	.290	.174
$\pi_2$			.598	.555	.610	.627
H			1.69	1.73	1.34	2.07

TABLE 4. Posterior probabilities for the events  $U = \lambda_t$  and  $V = \lambda_t$ : lung cancer (L); undifferentiated cell carcinoma (UCC); adenocarcinoma (A); squamous cell (SC) relative to control subjects.

t	CYP1A1	GST1	L		UCC		A		SC	
			U	V	U	V	U	V	U	V
1	ILe/ILe	+	.000	.871	.001	.186	.000	.746	.000	.488
2	ILe/ILe	-	.000	.003	.000	.277	.010	.006	.004	.001
3	ILe/Val	+	.001	.025	.003	.243	.026	.009	.000	.088
4	ILe/Val	-	.000	.064	.015	.045	.001	.216	.001	.045
5	Val/Val	+	.074	.036	.070	.247	.421	.012	.014	.377
6	Val/Val	-	.923	.000	.909	.000	.540	.008	.979	.000



## REFERENCES

- Boys, R. and Dunsmore, I. (1987), 'Diagnostic and sampling models in screening', *Biometrika* **74**(2), 365–374.
- Dawid, A. P. (1976), 'Properties of diagnostic data distributions', *Biometrics* **32**, 647–658.
- Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.
- Kawajiri, K., Nakachi, K., Imai, K., Watanabe, J. and Hayashi, S. (1993), 'The CYP1A1 gene and cancer susceptibility', *Critical Reviews in Oncology and Hematology* **14**, 77–87.
- Marshall, R. (1988), 'Bayesian analysis of case-control studies', *Statistics in Medicine* **12**(7), 1223–1230.
- Rebbeck, T. R., Viana, M., Jordan, H. A., Weber, B. L. and Rogatko, A. (1996), 'Evaluation of susceptibility genes in disease risk assessment', *American Journal of Human Genetics* **59**(4), A28.
- Sox, H. C., ed. (1990), *Common Diagnostic Tests Use and Interpretation*, American College of Physicians, Philadelphia.
- Viana, M. A. G. and Farewell, V. (1990), 'A test for diagnostic utility', *Canadian Journal of Statistics* **18**(4), 289–295.
- Viana, M. A. G. and Farewell, V. (1994), 'Assessing the diagnostic utility of a test', *Biometrical Journal* **36**(2), 131,145.
- Viana, M. A. G. and Ramakrishnan, V. (1992), 'Bayesian estimates of predictive value and related parameters of a diagnostic test', *Canadian Journal of Statistics* **20**(3), 311–321.
- Zelen, M. and Parker, R. (1986), 'Case-control studies and Bayesian inference', *Statistics in Medicine* **5**, 261–269.

The University of Illinois at Chicago  
College of Medicine M/C 648  
1855 West Taylor Street, Chicago, Il 60612

Fox Chase Cancer Center  
Department of Biostatistics  
Philadelphia, PA 19111

The University of Pennsylvania Medical Center  
Department of Biostatistics and Epidemiology  
Philadelphia, PA 19104