Web Appendix for "Quantifying risk stratification provided by diagnostic tests and risk predictions: Application to population mutation screening" by Hormuzd A. Katki

1 | TOTAL GAIN IS A SPECIAL CASE OF MEAN RISK STRATIFICATION

Total Gain (TG) measures the explanatory power of a continuous covariate M in a binary regression model $P(D + | M = m) = \pi(m)$, where $\pi(m)$ is typically a logistic regression[1]. Denote overall disease prevalence as $P(D+) = \pi$. Choose cutpoint m^* such that M is cut at disease prevalence: $P(D + | M = m^*) = \pi(m^*) = \pi$. Then

$$TG \equiv 2 \left| \int_{m^*}^{\infty} (\pi(m) - \pi) \, dF(m) \right| = 2|P(D+, M \geq m^*) - P(D+)P(M \geq m^*)|.$$

Thus TG is $|MRS|$ for continuous M cut at disease prevalence (see Web Appendix Section 2). MRS is valid for discrete/continuous M and allows any cutpoint of M. Unlike MRS, TG is non-negative.

2 | MRS EQUALS TWICE THE DEVIATION OF ANY JOINT PROBABILITY FROM THE PRODUCT OF ITS MARGINALS

Denote $PPV = P(D + | M+)$, $cNPV = 1 - NPV = P(D + | M-)$, $P(D+) = \pi$ and $P(M+) = p$. Thus

$$MRS = \{PPV - P(D+)\}P(M+) + \{P(D+) - cNPV\}P(M-)$$
$$= \{P(D+, M+)) - P(D+)P(M+)\} + \{P(D+)P(M-) - P(D+, M-)\}.$$

First, substituting $P(D+, M-) = \pi - P(D+, M+)$ yields MRS as a deviation from $P(D+, M+)$:

$$MRS = \{P(D+, M+)) - p\pi\} + \pi\{1 - p\} - \{\pi - P(D+, M+)\}$$
$$= \{P(D+, M+)) - p\pi\} - p\pi + P(D+, M+)$$
$$= 2\{P(D+, M+) - p\pi\}.$$

Thus $P(D+, M+) = MRS/2 + p\pi$.

Second, substituting $P(D+, M+) = \pi - P(D+, M-)$ yields MRS as a deviation from $P(D+, M-)$:

$$MRS = 2\{P(D+, M+) - p\pi\}$$
$$= 2\{\pi - P(D+, M-) - p\pi\}$$
$$= 2\{\pi(1 - p) - P(D+, M-)\}.$$

Thus $P(D+, M-) = \pi(1 - p) - MRS/2$.

Third, substituting $P(D+, M-) = (1 - p) - P(D-, M-)$ yields MRS as a deviation from $P(D-, M-)$:

$$MRS = 2\{\pi(1 - p) - P(D+, M-)\}$$
$$= 2\{\pi(1 - p) - \{(1 - p) - P(D-, M-)\}\}$$
$$= 2\{(1 - p)\{\pi - 1\} + P(D-, M-)\}$$
$$= 2\{P(D-, M-) - (1 - \pi)(1 - p)\}.$$

Thus $P(D-, M-) = MRS/2 + (1 - p)(1 - \pi)$.

Fourth and final, substituting $P(D-, M-) = (1 - \pi) - P(D-, M+)$ yields MRS as a deviation from $P(D-, M+)$:

$$MRS = 2\{P(D-, M-) - (1 - \pi)(1 - p)\}$$
$$= 2\{(1 - \pi) - P(D-, M+) - (1 - \pi)(1 - p)\}$$
$$= 2\{(1 - \pi)(1 - (1 - p)) - P(D-, M+)\}\}$$
$$= 2\{p(1 - \pi) - P(D-, M+)\}.$$

Thus $P(D-, M+) = p(1 - \pi) - MRS/2$.

TABLE 1 Simulation of MRS, Youden's index, their standard errors (se), and 95% confidence interval (CI) coverage. Sample size $n = 4589$ with outcome prevalence $\pi = 2.3\%$.

	0.78% threshold		10% threshold		30% threshold	
	Youden	MRS	Youden	MRS	Youden	MRS
true parameter	0.37587	0.016721	0.24422	0.010857	0.17878	0.007947
mean estimate	0.3757	0.016711	0.24406	0.010858	0.17861	0.007934
empirical se	0.03933	0.002354	0.04436	0.002217	0.03858	0.001862
estimated se	0.03885	0.002343	0.04409	0.002209	0.03819	0.00185
95% CI	94.037	94.767	94.442	94.25	93.929	94.231

3 | RELATIONSHIP OF NET BENEFIT TO NET BENEFIT OF INFORMATION

Denote $P(D+) = \pi$ and $P(M+) = p$. Net Benefit (NB), ignoring test costs, is

$$NB = \pi Sens - \frac{R(1 - Spec)(1 - \pi)}{1 - R}$$
$$= P(D+, M+) - \frac{R}{1 - R} P(D-, M+)$$
$$= \frac{P(D+, M+) - pR}{1 - R},$$

by substituting $P(D-, M+) = P(M+) - P(D+, M+)$. Then

$$NB = \frac{PPV - R}{(1 - R)/p}.$$

Now starting from NBI, and substituting MRS equation (7) from the main paper

$$NBI = \frac{MRS/2}{1 - R}$$
$$= \frac{P(D+, M+) - P(D+)P(M+)}{1 - R}$$
$$= \frac{P(D+, M+) - p\pi}{1 - R}$$
$$= \frac{PPV - \pi}{(1 - R)/p}.$$

Thus

$$NB = NBI \times \frac{PPV - R}{PPV - \pi}.$$

Thus $NBI < NB$ if $R < \pi$, $NBI > NB$ if $R > \pi$, and $NBI = NB = J\pi$ at $R = \pi$ ($J = Sens + Spec - 1$ is Youden's index).

Note that NB and NBI are special cases of a more general framework for the net benefits for tests that rule-out or rule-in for interventions [2].

4 | VARIANCE OF MRS AND HYPOTHESIS TESTING

Asymptotic variances for MRS and Youden's index follow from applying the delta method to the quadrinomial variance matrix from a 2x2 table with n as the sample size and $a = P(D+, M+)$, $b = P(D+, M-)$, $c = P(D-, M+)$, $d = P(D-, M-)$. Each variance is $g^T V g$ where $g(a, b, c, d)$ is the usual gradient of the quantity. V is the usual quadrinomial variance matrix of the cell probabilities, for total sample size n:

$$V = 1/n \times \begin{bmatrix} a(1 - a) & -ab & -ac & -ad \\ -ba & b(1 - b) & -bc & -bd \\ -ca & -cb & c(1 - c) & -cd \\ -da & -db & -dc & d(1 - d) \end{bmatrix}$$

Recall that MRS is twice the cross-product difference of joint probabilities inside a 2x2 table: $MRS = 2(ad - bc)$ (see main paper Appendix). The variance of MRS is based on the gradient $g = 2[d, -c, -b, a]$. Calculating $g^T V g$ yields the variance

$$Var(MRS) = 4\{ad(a+d) + bc(b+c) - MRS^2\}/n.$$

This variance requires only the sample proportions of the joint probabilities. It does not require fixed or *a priori* known test positivity or disease prevalence. Parenthetically, the Cochran and Mantel-Haenszel tests are score-tests of MRS that use the variance as estimated under the null hypothesis. They cannot be used for confidence intervals which require the variance of MRS under no restrictions, which is what is derived above.

Table 1 examines the properties of the MRS variance and MRS confidence interval coverage by simulation. We did 1 million simulations for each of 3 quadrinomial 2x2 tables based on the 3 cutpoints we considered in the Washington Ashkenazi Study (WAS) : 0.78%, 10%, and 30%. The quadrinomials were based on sample size of 4589 (as in WAS), with expectations for the $[a, b, c, d]$ cell counts as:

1. $[84.72, 19.73, 1951.88, 2532.67]$ for 0.78%

2. $[29.63, 74.75, 177.70, 4306.92]$ for 10%

3. $[19.74, 84.62, 46.52, 4438.11]$ for 30%.

In all cases, MRS and its variance were estimated with little bias, and 95% confidence intervals performed nominally (Table 1).

To ensure proper MRS confidence intervals, note that $MRS \in [-0.5, 0.5]$. This is easy to see based on the MRS expression $MRS = 2J\pi(1 - \pi)$, where J is Youden's index and π is disease prevalence. The maximum/minimum MRS of ± 0.5 occurs when $\pi = 0.5$ and $J = \pm 1$. Thus $(0.5 + MRS) \in [0, 1]$ and a logit transformation of $(0.5 + MRS)$ will ensure that confidence intervals are proper. Applying the delta method yields

$$Var(logit(0.5 + MRS)) = ((0.5 + MRS)(0.5 - MRS))^{-2} \times Var(MRS).$$

This variance is used to calculate confidence intervals on the $logit(0.5 + MRS)$ scale. Then, convert back to the MRS scale by applying to each endpoint x the inverse function

$$\frac{e^x}{1 + e^x} - \frac{1}{2}.$$

4.1 | P-values for testing if two MRSs are equal

In general, testing if two independent MRSs differ can be based on the difference of the two MRSs, whose variance would be $Var(MRS_1) + Var(MRS_2)$. But if the MRSs are calculated within the same population, and hence same prevalence, the ratio of MRSs is a better statistic. This is because the nuisance parameter of prevalence cancels out, leaving the ratio of Youden's indices J_1, J_2:

$$\frac{MRS_1}{MRS_2} = \frac{2J_1\pi(1 - \pi)}{2J_2\pi(1 - \pi)} = \frac{J_1}{J_2}.$$

We will calculate the variance of the log of the ratio of two independent Youden's indices, based on a quadrinomial likelihood for 2x2 tables.

Recall that Youden's index $J = Sens + Spec - 1$ can be written in terms of the joint probabilities in a 2x2 table:

$$J = \frac{a}{a + b} + \frac{d}{c + d} - 1.$$

To calculate the variance of a single Youden's index, the gradient is

$$g = \left[\frac{b}{(a+b)^2} \quad \frac{-1}{a(1+b/a)^2} \quad \frac{-1}{d(1+c/d)^2} \quad \frac{c}{(c+d)^2} \right]$$

The variance of a single Youden's index J is $V_J = g^T V g$. Table 1 shows that the variance for Youden's index is unbiased and 95% confidence intervals performed nominally.

For two independent Youden's indices, the variance of the log of their ratio is asymptotically

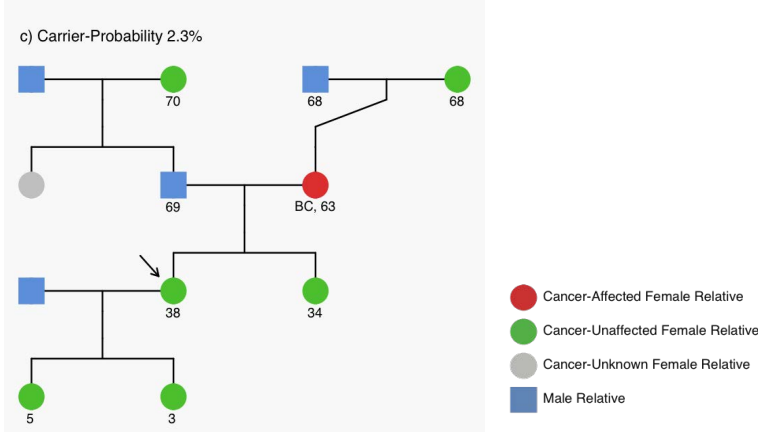$$V_{12} = \frac{V_{J_1}}{J_1^2} + \frac{V_{J_2}}{J_2^2},$$

FIGURE 1 Based on this pedigree, BRCAPRO calculates that the volunteer (arrow) has 2.3% chance of carrying a *BRCA1/2* mutation. Mother had breast cancer at age 63, no other cancers in the family.

and asymptotically $log(J_1/J_2)/\sqrt{V_{12}} \sim N(0,1)$.

The p-values comparing MRSs in sections 5.1 and 5.2 of the main paper are based on the ratio of Youden's indices. For Ashkenazi-Jews, comparing the MRSs at a 0.78% threshold vs. 10%, the p-value based on the difference of MRSs is 0.0703, but that based on ratio of Youden's indices is 0.0392 (as reported in the main paper section 5.1). For comparing the 0.78% threshold vs. 30%, the p-value based on the difference of MRSs is 0.0035, but that based on ratio of Youden's indices is 0.00054 (as reported in the main paper section 5.2). In each situation, the smaller p-values by using the ratio of Youden's indices reflects the gain in statistical power by removing the nuisance parameter, disease prevalence π.

5 | FIXING THE ROC CURVE TO ACCOUNT FOR DISEASE PREVALENCE: THE FREQUENCY-SCALED ROC CURVE AND ITS RELATIONSHIP OF AUC AND MRS

The ROC plots $P(M+|D+)$ vs. $P(M+|D-)$. The frequency-scaled ROC (fROC) plots $P(D+,M+)$ vs. $P(D+,M-)$[3]. Unlike the square ROC, fROC accounts for prevalence and is a π by $1-\pi$ rectangle. The diagonal is the uninformative fROC curve, the area under which is $\pi(1-\pi)/2$. The area under the fROC curve for a test, which is a single point with lines extending to the bottom-left and top-right corners, can be shown to be $P(D+,M+)(1-\pi)/2 + P(D-,M-)\pi/2$. The *ratio* of the area under the fROC to the chance area for random selection, equals the AUC:

$$\frac{area}{chance\ area} = \frac{P(D+,M+)(1-\pi)/2 + P(D-,M-)\pi/2}{\pi(1-\pi)/2} = \frac{Sens + Spec}{2} = AUC.$$

The *difference* between the area under the fROC to the chance area equals $MRS/4$:

$$area - chance\ area = P(D+,M+)(1-\pi)/2 + P(D-,M-)\pi/2 - \pi(1-\pi)/2 = MRS/4.$$

A high ratio (AUC) might conceal a small difference (MRS), which is apt to be the case for uncommon diseases.

6 | FIGURE 1: EXAMPLE CANCER FAMILY HISTORY AND FAMILY TREE

References

1. Bura, E and JL Gastwirth (2001). The binary regression quantile plot: Assessing the importance of predictors in binary regression visually. Biometrical Journal 43(1), 5–21.

2. Pennello J, Pantoja-Galicia N, Evans S (2016). Comparing diagnostic tests on benefit-risk. J Biopharm Stat 26, 1083–1097.

3. Hilden, J (1991). The area under the ROC curve and its competitors. Medical Decis Making 11, 95–101.