

# On Cost-Effectiveness of Screening Using a Marker/Risk-model

Ionut Bebu and Hormuzd Katki

August 18, 2018

## 1 Introduction

Consider the problem of screening for cancer when a marker/risk-model is available. Each individual has a marker value or a risk score derived using a risk-model and the patient's characteristics. Individuals with high marker values or high risk scores are then tested for cancer. The problem is to identify an optimal cutoff value above which an individual needs to be tested for cancer. Notice that the lower the cutoff for the screening test is, the more likely a cancer case is detected (higher effectiveness), but also the higher the number of negative tests (higher costs).

## 2 Main

Katki ? proposed a new measure of evaluating the risk before and after incorporating the marker information<sup>1</sup>. The mean risk classification (MRS) measure they proposed measure is the average change in risk (posttest - pretest). Briefly, their approach is to quantify the value of marker  $M$  to predict the outcome  $D$  by  $P(D+|M+) - P(D+)$  if  $M+$ , and by  $P(D-|M-) - P(D-)$  if  $M-$ , which yields (on average)

$$MRS = [P(D+|M+) - P(D+)] \cdot P(M+) + [P(D-|M-) - P(D-)] \cdot P(M-) \quad (1)$$

$$= 2[P(D+, M+) - P(D+) \cdot P(M+)]. \quad (2)$$

Here  $M+$  denotes the event  $M \geq m_0$ , for some cutoff value  $m_0$ , while  $M-$  denotes the event  $M < m_0$ .

It follows from (??) that  $MRS$  is simply twice the covariance of  $D+$  and  $M+$ . Therefore  $MRS$  is indeed a measure of association between the marker and the outcome, but it is not immediately clear how relevant it is if the goal is to find an optimal cutoff value  $m_0$ .

Katki ? further show that

$$MRS = 2J \cdot \pi(1 - \pi) = 4(AUC - 0.5) \cdot \pi(1 - \pi), \quad (3)$$

where  $\pi = P(D+)$  and  $J$  is Youden's index. However, notice that  $\pi$  does not depend on the marker  $M$ , so that all three measures  $MRS$ ,  $AUC$  and  $J$  are equivalent when it comes to choosing the marker cutoff value  $m_0$ .

Alternatively, one can introduce a cost-effectiveness framework for this problem as follows. Assume the effectiveness score for a subject who does not develop cancer is 1, and let  $e_1$  denote the effectiveness score for a subject with a positive marker  $M$  which then tests positive for cancer; similarly, let  $e_2$  be the effectiveness score for subject who is either not screened using the marker  $M$  or is  $M-$ , but then develops cancer. Since the cancer cases identified using the marker  $M+$  are diagnosed earlier, it follows that  $e_1 > e_2$ , and both are less than 1.

Let  $c_m$  denote the cost of a marker evaluation,  $c_0$  denote the cost of standard test for cancer,  $c_1$  the cost of treatment for a cancer case diagnosed early (i.e., tested positive for  $M$ ), and  $c_2$  the treatment cost for a cancer case diagnosed late (i.e., either not screened with  $M$  or tested negative for  $M$ ).

Two scenarios are considered at individual level:

---

<sup>1</sup>See the online MRS webtool.

1. No screening for  $M$ . There are two possibilities:

- (a) The individual has cancer ( $D+$ ): cost is  $c_0 + c_2$  and effectiveness is  $e_2$
- (b) The individual does not have cancer ( $D-$ ): cost is 0, effectiveness is 1.

2. Screening for  $M$ . There are four possibilities:

- (a) The screening is positive and the individual has cancer ( $M+, D+$ ): cost is  $c_m + c_0 + c_1$  and effectiveness is  $e_1$ .
- (b) The screening is positive but the individual does not have cancer ( $M+, D-$ ): cost is  $c_m + c_0$  and effectiveness is 1.
- (c) The screening is negative but the individual has cancer ( $M-, D+$ ): cost is  $c_m + c_0 + c_2$  and effectiveness is  $e_2$ .
- (d) The screening is negative and the individual has not have cancer ( $M-, D-$ ): cost is  $c_m$  and effectiveness is 1.

Let  $C$  and  $E$  denote the cost and effectiveness without screening, and  $C_M$  and  $E_M$  with screening. One has

$$\begin{aligned}\mathcal{E}(C) &= (c_0 + c_2) \cdot P(D+) \\ \mathcal{E}(E) &= e_2 \cdot P(D+) + P(D-) \\ \mathcal{E}(C_M) &= c_m + [1 - P(D-, M-)] \cdot c_0 + P(D+, M-) \cdot c_2 + P(D+, M+) \cdot c_1 \\ \mathcal{E}(E_M) &= P(D-) + P(D+, M+) \cdot e_1 + P(D+, M-) \cdot e_2,\end{aligned}$$

so that the differences in cost and effectiveness between the two scenarios are

$$\begin{aligned}\Delta C &= c_m + c_0 \cdot P(M+) + (c_1 - c_2 - c_0) \cdot P(D+, M+) \\ \Delta E &= (e_1 - e_2) \cdot P(D+, M+).\end{aligned}$$

One then selects  $m_0$  based on both  $\Delta C$  and  $\Delta E$ . Notice that  $\Delta E$  is maximized when  $m_0 = -\infty$  (i.e., screen everybody).

### 3 Illustration

Let  $\tilde{M}$  such that the true risk of cancer is

$$P(D = 1 | \tilde{M} = \tilde{m}) = L(\alpha_0 + \alpha_1 \cdot \tilde{m}), \quad (4)$$

where  $L(x) = \exp(x)/(1 + \exp(x))$ .

However,  $\tilde{M}$  is not directly observed. Instead, one observes a marker  $M$ , where

$$M = \tilde{M} + \epsilon. \quad (5)$$

We further assume  $\epsilon \sim N(0, \sigma_\epsilon^2)$ , and  $M \sim N(\mu_M, \sigma_M^2)$ . One obtains

$$\begin{aligned}P(M+) &= 1 - \Phi(m_0 - \mu_M, 0, \sigma_M^2 + \sigma_\epsilon^2) \\ P(D+, M+) &= \int_{m_0}^{\infty} \int_{-\infty}^{\infty} L(\alpha_0 + \alpha_1 \cdot (m - \epsilon)) \cdot \phi(\epsilon, 0, \sigma_\epsilon^2) \cdot \phi(m, \mu_M, \sigma_M^2) d\epsilon dm\end{aligned}$$

while  $P(D+)$  is obtain by replacing  $m_0$  by  $\infty$  in the expression for  $P(D+, M+)$  above.

The approaches are illustrated using  $\alpha_0 = -6.67$ ,  $\alpha_1 = 1$ ,  $\mu_M = 1$ ,  $\sigma_M = 1$ ,  $c_m = 0.5$ ,  $c_0 = 0.01$ ,  $c_1 = 1$  and  $c_2 = 2$ .

The optimal  $m_0$  with respect to the ICER (i.e., lowest ICER) and the corresponding ICER are

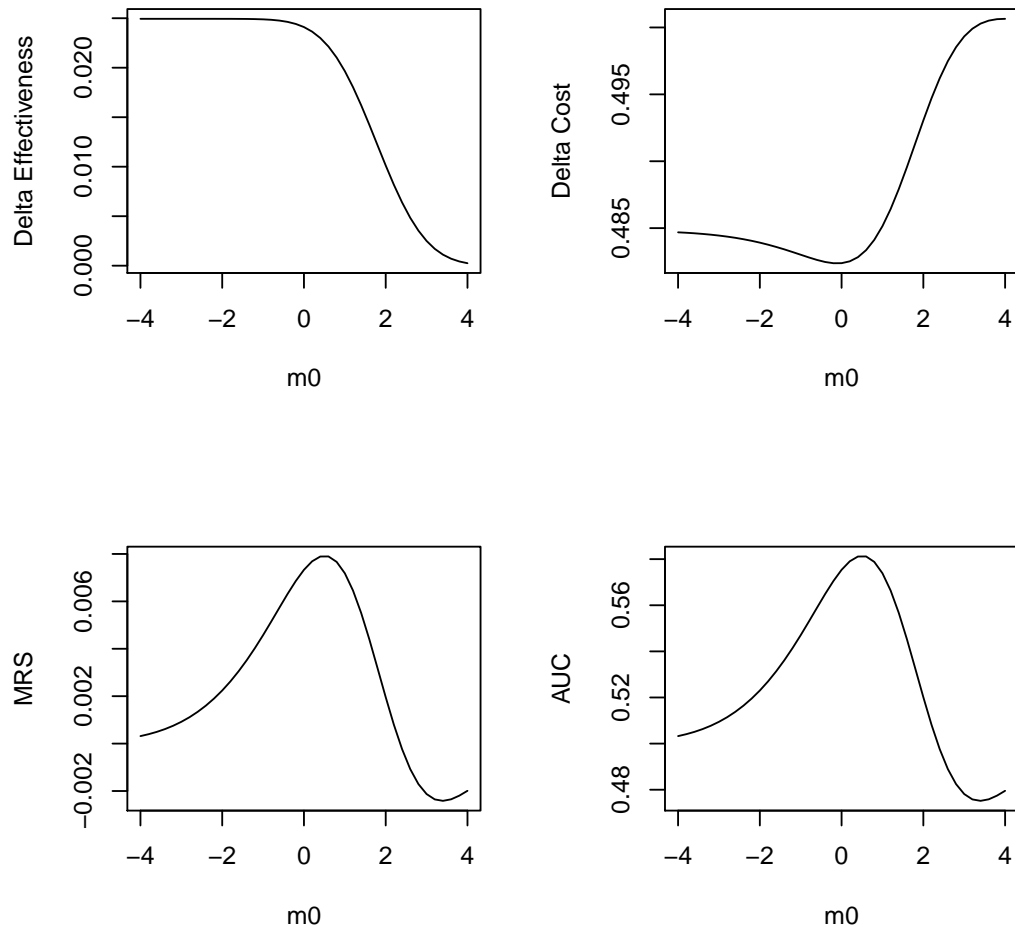


Figure 1: Effectiveness, cost, MRS and AUC as a function of the cutoff value  $m_0$ .

[1] -1.4

[1] 19.39465

The optimal  $m_0$  with respect to the AUC (i.e., largest AUC) and the corresponding ICER are

[1] 0.6

[1] 0.5811674

The optimal  $m_0$  with respect to the MRS (i.e., largest MRS) and the corresponding ICER are

[1] 0.6

[1] 0.00789477

As expected, MRS and AUC give identical results, but which are different than the ones for ICER. Moreover, the  $m_0$  value selected by MRS/AUC has an ICER value of

[1] 21.83748

or a

[1] 11.18642

percent (%) increase relative to the optimal ICER.

## 4 Discussion

Various approaches for calculating an optimal cutoff for a screening marker were considered. The MRS measure, which is twice the covariance between  $D+$  and  $M+$  is shown to be equivalent to AUC with respect to selecting this optimal cutoff  $m_0$ .

Note that the above calculations depend on the unobserved  $\alpha_0$ ,  $\alpha_1$  and  $\sigma_\epsilon$ . However, for various values of  $\sigma_\epsilon$ , one can estimate  $\alpha_0$  and  $\alpha_1$  based on the coefficients obtained by regressing  $D = 1$  on  $M$ , which can then be used to find the optimal cutoff value  $m_0$ .

## 5 Incremental Net Benefit

A large Incremental Cost-Effectiveness Ratio (ICER) can mask a small difference, and vice-versa. For this reason, the Incremental Net Benefit (INB) is better:

$$INB = \Delta E - \Delta C.$$

This requires redefining Effectiveness as the dollar value of each life-year gained.

### 5.1 Application to deciding who should undergo *BRCA1/2* genetic testing

We previously applied MRS to the controversy over who should get tested for mutations in the *BRCA1/2* genes, which cause high risks of breast and ovarian cancer (?). The mutations are rare in the general population ( $\approx 0.25\%$ ), but are 10 times more common among Ashkenazi-Jews (?). Currently, according to the UK National Institute for Health and Care Excellence (NICE) and the US Preventive Services Task Force, women are referred for mutation testing only if they have a strong family history of breast or ovarian cancer (?) as quantified by a risk model calculating that their risk of carrying a mutation exceeds 10% (?). However, as mutation-testing costs fall, prominent voices have called for testing *all* women (??). *BRCA1/2*

testing is already being offered to a large unselected Canadian population as a demonstration project (?). Testing all women would strain clinical resources by testing millions of women, 99.75% of whom will test negative. At US \$500-\$1,000 per test, testing millions of women has clear commercial implications.

In contrast, we recently demonstrated that 80% of Ashkenazi-Jewish mutation-carriers could be identified by testing only 44% of Ashkenazi-Jewish women (?). This is achieved by a low mutation-risk threshold of 0.78%, far lower than the current 10% UK NICE and US recommendation. However, we could not formally justify any choice of risk-threshold. To better understand the implications of different choices of risk threshold, we previously considered multiple metrics, including AUC, Net Benefit, and MRS/NBI (?). We showed that risk thresholds between 0.78% to 5.0% maximized the MRS ( $\approx 1.7\%$ ) for Ashkenazi-Jewish women, and for the general population, thresholds between 0.07% to 0.56% maximized the MRS ( $\approx 0.18\%$ ) (?). Here we compare risk thresholds that maximize MRS to risk thresholds that maximize the INB.

We use data on 4,589 volunteers (102 *BRCA1/2* mutation carriers) from the Washington Ashkenazi Study (WAS) (?). We calculated each volunteer's risk of carrying a mutation, based on their self-reported family-history of breast/ovarian cancers, using the BRCAPRO risk model (?). Here  $M$  is the BRCAPRO risk score, and because BRCAPRO is a well-calibrated risk model (?),  $m_0 = R$ , i.e. the cutpoint  $m_0$  equals the risk threshold  $R$ . Disease  $D$  indicates the presence of a *BRCA1/2* mutation.

## 5.2 Costs and Effectiveness of *BRCA1/2* genetic testing, preventive interventions, and cancer treatments

Let us redefine Effectiveness as the dollar value of the life-years gained by undergoing genetic testing. Let us presume that women without mutations have a normal lifespan of age 80. *BRCA1/2* mutation-carriers have 5.7 fewer years of expected life (?), so we set their life-expectancy to age 74.

If mutation-carriers get genetic testing, they will act to reduce their cancer risks. Most mutation-carriers undergo oophorectomy, the removal of ovaries. Some also choose to have mastectomy, the removal of breasts. These surgeries reduce cancer risk drastically, but not entirely; there is typically a 10% chance that cancer occurs in each organ anyway. These women also undergo frequent screening, and if cancer is found early, hopefully treatment is more effective. I don't know what their life-expectancy is, but let me set it as age 78.

I set the dollar-value of each life-year gained as \$100,000 per year. Thus Effectiveness is

$$e_1 - e_2 = (78 - 74) \times \$100,000 = \$400,000$$

For costs,  $c_m$  is the cost of using the BRCAPRO risk model. This is very low-cost, but you need to come to the doctor's office, your family history has to be elicited, it takes time. I will set it to the cost of a doctor's visit, say \$100.

$c_0$  is the cost of the BRCA genetic test. Prices are falling, it is now about \$1,000.

$c_1$  is the cost of interventions for BRCA mutation carrier found by screening. Generally they will undergo risk-reducing interventions, at least oophorectomy and perhaps also mastectomy, and then frequent lifetime cancer screening. But cancer can still occur, and those 25% (*should this be related to the 10% mentioned above?*) who still get cancer incur cancer treatment costs. For  $c_2$ , this is the cost of cancer treatment for mutation carriers who are not found until cancer is diagnosed. All of them will require full cancer treatment. I don't know how to specify  $c_1, c_2$ . Let me set them both equal to \$100,000 for now.

Thus we have for Costs:

$$\begin{aligned} c_m &= \$100 \\ c_0 &= \$1,000 \\ c_1 &= \$100,000 \\ c_2 &= \$100,000. \end{aligned}$$

### 5.3 MRS vs. INB for *BRCA1/2* genetic testing

Figure 2 (left panels) show the increase in effectiveness and costs due to genetic testing for both Ashkenazi-Jews and the general population, as a function of the BRCAPRO risk threshold to offer women genetic testing. When the threshold is low, most everyone is offered testing. The cost is highest here, and decreases down to only marker testing cost (i.e. BRCAPRO risk model cost) if no one is offered genetic testing. Effectiveness is also highest if everyone is tested for mutations, and goes down to zero if no one is. Note that while the cost scale (y-axes) are the same, the effectiveness scale (right-axes) is about 10 times higher for Ashkenazi-Jews. This reflects that mutations are 10 times more common for Ashkenazi-Jews.

(Parenthetically, note that in the limits of testing no one or testing everyone, I should subtract out marker cost  $c_m$  because the marker is unnecessary. The limits of the INB curve don't equal the correct limits. I should probably add two points on the figures for the INBs of testing everyone or no one.)

Figure 2 (right panels) compare the MRS to INB. MRS is zero if no one, or everyone, gets mutation testing, and always peaks at threshold equals prevalence (dotted line). INB for Ashkenazi-Jews is highest at a threshold of zero, meaning everyone should get testing. This happens because of the high value attached to each life-year gained, and Ashkenazi-Jews are much more likely to gain this benefit. For the general population, note that  $INB < 0$  for the lowest risk thresholds and for risk thresholds greater than 10% (recall that 10% is the current official threshold in the US and UK.). These are thresholds where net harm would be done to the population.

Importantly, for the general population, the INB curve is proportional to the MRS curve (i.e. the curves overlap, but note the very different scales on the left-axis and right-axis). Thus INB and MRS are maximized at the same risk-threshold and give the same answer. The next section shows that this happens because the costs and effectiveness balance each other in a special way.

### 5.4 Example: single CT lung-cancer screen

The eligible population for lung-cancer screening is you must be ages 55-80, smoked at least 30 pack-years, and cannot have quit for more than 15 years. We will assume this population is fixed, as all the data is based on this population anyway. The question is then, what is the cost-effectiveness of a single CT lung-cancer screen.

To estimate effectiveness, we need to know the the life-expectancy in this population, with and without a single screen.  $e_1$  is the number of life-expectancy for someone whose cancer is detected by CT and  $e_2$  is the life-expectancy if the cancer is not detected by CT (in dollar value). Without dollar value, the years-gained by Bill Black's NEJM paper is roughly  $e_1 - e_2 = 8.4792 - 6.8479 = 1.6313$ . That is for 3 screens, let's assume for 1 screen we can divide by 3, so the life-years gained is 0.54377. At \$100,000 value per life year, this is

$$e_1 - e_2 = 0.54377 \times \$100,000 = \$54,377$$

For costs,  $c_m$  is the cost of a CT screen. By Bill Black's NEJM paper, the screen is \$1,130 and work-up is \$835 for a total of  $c_m = \$1,965$ .

$c_0$  is the cost of the definitively diagnosing lung cancer. If this is a bronchoscopy with biopsy, Bill Black's paper gives lots of costs for different ways to do this, but this is about \$500.

$c_1$  is the cost of treating lung cancer found during screening and  $c_2$  is the cost of treating lung cancer when there is no screening. By Bill Black's paper, it looks like treatment cost under screening is  $c_1 = \$8,461$  and for no screening  $c_2 = \$7,835$ , so  $c_1 - c_2 = \$625$  (Note for Hormuzd: These treatment costs seem way too low.)

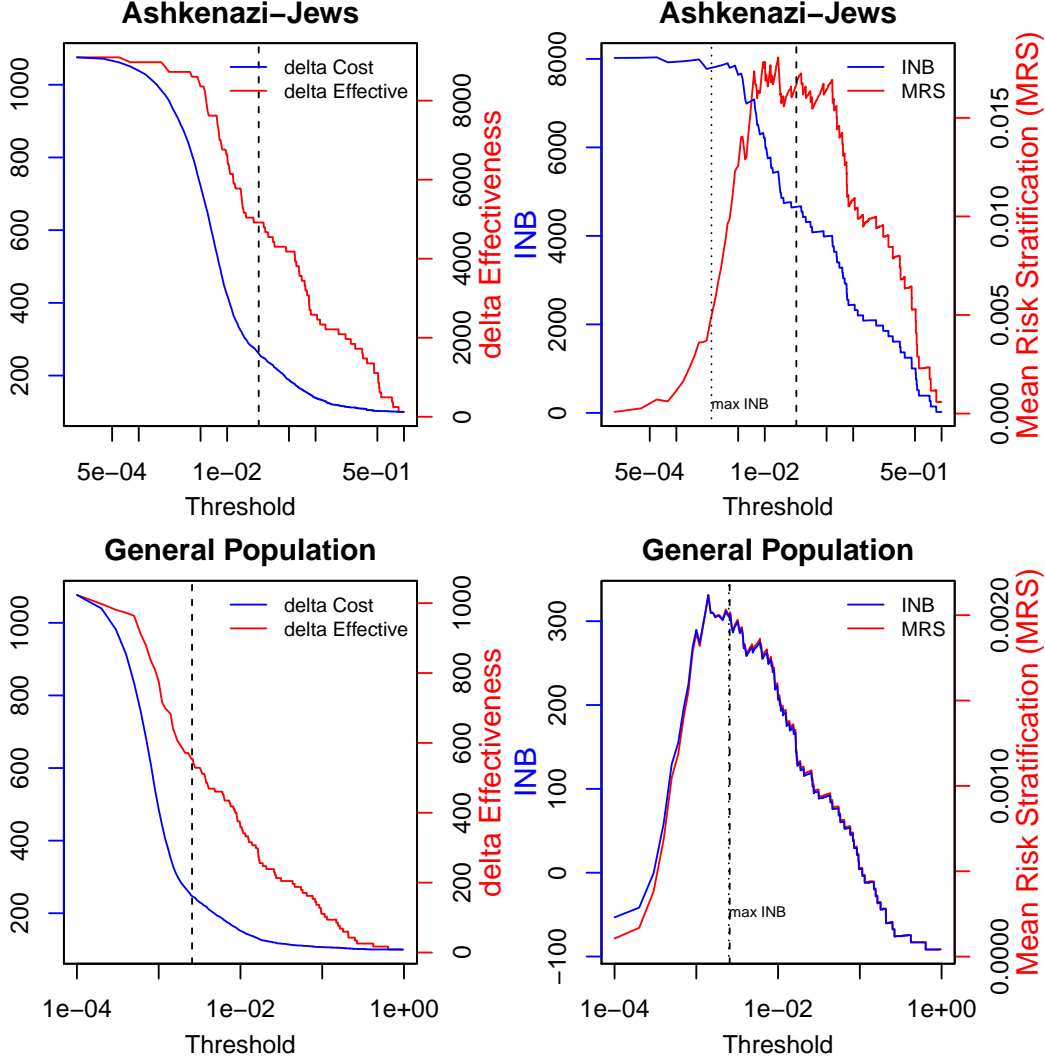


Figure 2: Effectiveness, cost, MRS and INB as a function of the cutoff Threshold values  $m_0$ . The marker is the BRCA PRO model score (cutoff at thresholds  $m_0$ ), the confirmatory test is the BRCA genetic test itself. The outcome  $D+$  is having a BRCA mutation. Costs  $c_1$  and  $c_2$  include risk-reducing interventions for the found BRCA mutation carriers, and cost of future cancer treatments. Effectiveness is the dollar value of the the life-years gained by early intervention. First row is for Ashkenazi-Jews, second row is for general population. Dashed line is where threshold equals mutation prevalence (2.3% for Ashkenazi-Jews, 0.26% for general population). Dotted line is the risk threshold that maximizes INB which is  $\pi_{INB}$

Thus we have for Costs:

$$\begin{aligned} c_m &= \$1,965 \\ c_0 &= \$500 \\ c_1 &= \$8,461 \\ c_2 &= \$7,835. \end{aligned}$$

Thus  $e = (e_1 - e_2) - (c_1 - c_2 - c_0) = \$54,377 - (\$625 - \$500) = \$54,227$ . The optimal risk-threshold for a CT screen is  $\pi_{INB} = c_0/(e + c_0) = 0.91\%$  lung-cancer risk. In comparison, by the Church NEJM paper, the 1st screen PPV was 3.8%, so the implicit NLST risk threshold was even lower than 3.8%. Perhaps it could be around 0.91%, since average risk is probably much bigger than the minimum risk threshold.

The INB is  $eP(D+, M+) - c_m - c_0P(M+)$ . The Church NEJM paper reports  $P(M+) = 27.3\%$ ,  $P(D+) = 1.1\%$ , and I calculate  $P(D+, M+) = 1.028\%$ . The INB is  $-\$1,543$ , which would mean screening should never be done.

Note that  $INB > 0$  iff  $eP(D+, M+) > c_m + c_0P(M+) = \$2,102$  and thus  $e > \$204,426$ . Given the costs as I understand them, there is no way a single CT screen can be cost-effective.

## 6 When is INB proportional to MRS, so they have the same maximum?

$$\begin{aligned} INB &= \Delta E - \Delta C \\ &= [(e_1 - e_2) - (c_1 - c_2 - c_0)]P(D+, M+) - c_m - c_0P(M+). \end{aligned}$$

Define  $p = P(M+)$  and  $\pi = P(D+)$ . Recall that  $P(D+, M+) = MRS/2 + p\pi$ . Define the net effectiveness of early intervention:  $e = (e_1 - e_2) - (c_1 - c_2)$ . Then

$$\begin{aligned} INB &= (e + c_0)[MRS/2 + p\pi] - c_m - c_0p \\ &= (e + c_0)MRS/2 - c_m + p[e\pi + c_0\pi - c_0] \\ &= (e + c_0)MRS/2 - c_m + p[e\pi - c_0(1 - \pi)]. \end{aligned}$$

Note that  $INB \propto MRS$  if  $e\pi - c_0(1 - \pi) = 0$ , which happens when

$$\frac{\pi}{1 - \pi} = \frac{c_0}{e} = \frac{c_0}{(e_1 - e_2) - (c_1 - c_2)}, \quad \text{or} \quad \pi_{INB} = \frac{c_0}{c_0 + e}.$$

INB and MRS will have the same maximum if the odds of disease equals the ratio of confirmatory test cost to the net effectiveness of early intervention. Let us call this critical prevalence  $\pi_{INB}$  to indicate that this is the disease prevalence where MRS and INB yield the same optimal threshold. In Figure 2,  $c_0/e = \$1,000/(\$400,000 - \$0) = 0.25\%$ , and thus the critical prevalence is  $\pi_{INB} = 0.25\%$ . For the general population,  $\pi = 0.26\%$ , which is very close to the critical prevalence.

There is no reason for the confirmatory test cost and net effectiveness to balance in a way that their ratio equals the prevalence odds. However, it is useful to calculate the critical prevalence and compare to the observed disease prevalence, to qualitatively understand which way the threshold for maximum MRS changes when costs and effectiveness are accounted for. For example, for Ashkenazi-Jews, the prevalence is much higher than the critical prevalence ( $2.5\% > 0.25\%$ ). Thus immediately you know that the optimal threshold will decrease from prevalence towards testing more women than MRS would indicate.



## 7 Optimal cutoff threshold to maximize Incremental Net Benefit (INB)

Recall that the optimal threshold to maximize MRS is the same as that for Youden's index, which is risk threshold equals prevalence (?), see Appendix.

Here I derive the optimal threshold that maximizes INB. Recall that

$$INB = (e + c_0)MRS/2 - c_m + p[e\pi - c_0(1 - \pi)] = (e + c_0)MRS/2 - c_m + p[\pi(e + c_0) - c_0].$$

INB increases as MRS increases, but the last term decreases INB as  $p = P(M+)$  increases. To find the optimal threshold, let's find a simpler function with the same maximum Dividing both sides by the constant  $e + c_0$  yields

$$\frac{INB}{e + c_0} = MRS/2 - \frac{c_m}{e + c_0} + p[\pi - \pi_{INB}].$$

Because  $MRS = 2J\pi(1 - \pi)$ , then dividing by the constant  $\pi(1 - \pi)$  yields

$$\frac{INB}{(e + c_0)\pi(1 - \pi)} = J - \frac{c_m}{(e + c_0)\pi(1 - \pi)} + p\frac{\pi - \pi_{INB}}{\pi(1 - \pi)}.$$

Explicitly writing INB,  $J$ , and  $p$  as functions of cutoff thresholds  $m_0$ ,

$$INB(m_0) \propto J(m_0) + p(m_0)c,$$

where  $c = \frac{\pi - \pi_{INB}}{\pi(1 - \pi)}$ . This is a simpler function to optimize. Following the Appendix, we optimize by differentiating  $J(m_0) + p(m_0)c$

$$J(m_0) + p(m_0)c = \int_{m_0}^{\infty} P(M = m|D+) - P(M = m|D-) + cP(M = m) dm$$

with respect to the cutpoint  $m_0$  using the Leibniz Integral Rule:

$$\frac{d(J(m_0) + p(m_0)c)}{dm_0} = P(M = m_0|D-) - P(M = m_0|D+) - cP(M = m_0).$$

Setting the derivative equal to zero, and using Bayes' rule:

$$cP(M = m_0) + \frac{P(D+|M = m_0)P(M = m_0)}{\pi} = \frac{P(D-|M = m_0)P(M = m_0)}{1 - \pi}.$$

Note that the risk threshold is defined as  $R(m_0) = P(D+|M = m_0)$ . Note that  $R(m_0) = m_0$  because  $M$  is a risk-model. Canceling  $P(M = m_0)$ ,

$$\begin{aligned} \frac{c\pi + R}{\pi} &= \frac{1 - R}{1 - \pi} \\ \pi(1 - R) &= (c\pi + R)(1 - \pi) \\ \pi(1 - R) &= c\pi(1 - \pi) + R(1 - \pi) \\ R &= \pi - c\pi(1 - \pi) = \pi - (\pi - \pi_{INB}) = \pi_{INB} \end{aligned}$$

Thus INB is maximized at risk threshold  $R = \pi_{INB} = c_0/(e + c_0)$ .

Note that  $\pi_{INB}$  only requires the utilities: genetic test cost  $c_0$  and the net effectiveness of early intervention  $e = (e_1 - e_2) - (c_1 - c_2)$ . None of the objective aspects of the problem matter, like disease prevalence, etc. This is unlike MRS, Youden's index, AUC, Vickers's Net Benefit gain, which are all maximized at risk threshold equals disease prevalence, regardless of your utilities.

Note that  $\pi_{INB}$  doesn't require the marker cost  $c_m$ . Thus you don't need to specify  $c_m$  to find the optimal risk threshold. Because  $c_m$  contributes to the INB level, we only need it to make sure that the INB is greater than the INBs for testing no one or testing everyone.

Figure 2, dotted line, is the optimal risk threshold cutoff for INB,  $R = \pi_{INB} = 0.25\%$ . For the general population, this coincides with disease prevalence, and thus  $INB \propto MRS$  so they are maximized at the same risk threshold. For Ashkenazi-Jews,  $R = \pi_{INB} = 0.25\%$  is to left of prevalence  $\pi = 2.5\%$ , which tests 85% of Ashkenazi-Jews to find 96% of BRCA mutation carriers. The INB at 0.25% is about \$7,800. INB for testing everyone is  $INB_{all} = \$8,119$ , which of course removes the \$100 cost of doing the marker, which is the BRCAPRO risk model. For Ashkenazi-Jews, INB is maximized where everyone gets genetic testing for BRCA mutations, regardless of BRCAPRO risk score.

If we halve the dollar value of each life-year gained from \$100,000 to \$50,000, then the optimal risk-threshold doubles to  $\pi_{INB} = 0.50\%$ . For the general-population, the  $INB > 0$  for risk threshlds from 0.17%-1.87% and the peak is indeed roughly at 0.50% (INB=\$33.94, barely positive). For Ashkenazi-Jews, the interior peak is at 0.50%, but testing everyone still has a slightly higher INB.

If we further halve the dollar value of each life-year gained to \$25,000, then the optimal risk-threshold doubles to  $\pi_{INB} = 0.99\%$ . For the general-population, all  $INB < 0$ , the least bad INB is for testing no one (this is a case where we can only minimize harm, there are no good options). For Ashkenazi-Jews,  $INB > 0$  for risk-thresholds of 0%-5%, and the peak is around 0.99% (INB=\$1,235), although numerically 0.54% has best INB=\$1,400.

Ionut, why don't you calculate INB and the optimal  $\pi_{INB}$  for your logistic regression example?

## 8 Conditions for a marker to have greater INB than another marker

If one marker has greater sensitivity but less test positivity (or greater specificity yet more test positivity), then it is obviously that the marker is better, no cost-effectiveness analysis is needed. However, there is usually a tradeoff, one test having better sensitivity but not specificity. The tradeoff should be evaluated based on INB.

For two markers  $\Delta INB = INB_1 - INB_2 > 0$  if

$$\begin{aligned} \Delta INB &= (e + c_0)[P(D+, M_1+) - P(D+, M_2+)] - c_0[P(M_1+) - P(M_2+)] \\ &= (e + c_0)[Sens_1 - Sens_2]\pi - c_0[p_1 - p_2] > 0 \end{aligned}$$

for disease prevalence  $P(D+) = \pi$  and test-positivities  $p_1, p_2$ . Presuming  $p_1 > p_2$ ,

$$\frac{Sens_1 - Sens_2}{p_1 - p_2} > \frac{c_0/[e + c_0]}{\pi} = \frac{\pi_{INB}}{\pi}, \quad (6)$$

A usable test requires that the ratio of the the sensitivity gain to the increase in marker positivity is less than a constant. This constant is the ratio of the optimal risk threshold to the disease prevalence, which is the optimal threshold under information metrics such as Youden's index, AUC, MRS, and Net Benefit, none of which specify costs and effectiveness. Thus the key is the ratio of optimal risk thresholds by considering costs/effectiveness versus ignoring them.

This inequality presumes that  $p_1 > p_2$ , but if the opposite were true, then if the sensitivity gain is also negative (usually true), then the negative signs cancel (and if the sensitivity gain were positive, then marker 1 is obviously the better marker).

Note that if  $\pi_{INB} = \pi$ , then increases in sensitivity and test-positivity are weighed equally. If  $\pi_{INB} < \pi$  which means we should be testing more people, then we will accept smaller sensitivity gain for larger test-positivity increase. But if  $\pi_{INB} > \pi$ , meaning we should be testing fewer people, then we require a greater sensitivity gain than a test-positivity increase.

The above inequality can be equivalently written in terms of specificity: the new marker can have less sensitivity, if the decrease in test positivity is sufficient (the numerator and denominator are both negative).

In this situation, then the new marker has better specificity rather than sensitivity. To see this clearly, we write the above condition in terms of specificity by noting that  $P(D+, M+) = P(M+) - P(D-, M+) = p - cSpec \cdot (1 - \pi)$  (denoting  $cSpec = 1 - Spec$ ):

$$\begin{aligned}\Delta INB &= (e + c_0)[p_1 - cSpec_1 \cdot (1 - \pi) - p_2 + cSpec_2 \cdot (1 - \pi)] - c_0[p_1 - p_2] \\ &= e[p_1 - p_2] + (e + c_0)(1 - \pi)[cSpec_2 - cSpec_1] \\ &= e[(1 - p_2) - (1 - p_1)] + (e + c_0)(1 - \pi)[Spec_1 - Spec_2] > 0\end{aligned}$$

and presuming  $1 - p_1 > 1 - p_2$  (so that the sign flips in the inequality),

$$\frac{Spec_1 - Spec_2}{(1 - p_1) - (1 - p_2)} > \frac{e/[e + c_0]}{1 - \pi} = \frac{1 - \pi_{INB}}{1 - \pi}.$$

The two inequalities are equivalent, either can be used.

Note that these inequalities do not assume that the two markers are dichotomized at the optimal risk threshold of  $\pi_{INB}$ . In reality, the two markers need not be dichotomized at the optimal threshold, or even at the same threshold. For example, the Pap and HPV tests are implicitly categorized at different thresholds, neither of which is necessarily optimal. However, this leaves open the question that the comparison is unfair, much less suboptimal. If the two markers are dichotomized at the same risk-threshold, it's a most fair comparison, and if the risk threshold is also  $\pi_{INB}$ , then the comparison is definitive.

The equivalent condition in terms of MRS is:

$$\Delta INB = [MRS_1 - MRS_2](e + c_0)/2 - [p_1 - p_2][c_0 - \pi(e + c_0)] > 0$$

or

$$\frac{MRS_1 - MRS_2}{p_1 - p_2} > 2 \frac{c_0 - \pi(e + c_0)}{e + c_0} = 2(\pi_{INB} - \pi).$$

Note that this presumes that  $p_1 > p_2$ , reverse the inequality otherwise. It's interesting that the condition depends on the difference of optimal INB risk threshold versus optimal MRS threshold. This means that, if  $p_1 > p_2$ , marker 1 can have worse MRS yet better INB, but the optimal INB threshold must be less than prevalence (meaning, you should be testing more people anyway), and the MRS is not too much worse. Thus, a marker 1 with worse MRS for  $p_1 > p_2$  for  $\pi_{INB} > \pi$ , where you should be testing fewer people, can never be better. The equivalent expression via Youden's index J (via  $MRS = 2J\pi(1 - \pi)$ ) is

$$\frac{J_1 - J_2}{p_1 - p_2} > \frac{\pi_{INB} - \pi}{\pi(1 - \pi)}.$$

This is all interesting, but complex to explain, and provides no extra insight versus the conditions based on sensitivity and specificity.

## 9 Added

### 9.1 On Section 6, "Optimal cutoff threshold ..."

There are a lot of notations and changes in variables in Section 6, it seems difficult to follow. Here is a simplified version of your derivations.

Since  $M$  is an unbiased risk-model, one has

$$P(D+ | M = m) = m. \tag{7}$$

Let  $f_M$  denote the density of  $M$ , so that one obtains

$$\begin{aligned} P(M+) &= \int_{m_0}^{\infty} f_M(m) dm \\ P(D+, M+) &= \int_{m_0}^{\infty} P(D+ | M = m) \cdot f_M(m) dm \\ &= \int_{m_0}^{\infty} m \cdot f_M(m) dm. \end{aligned} \tag{8}$$

INB can be written as

$$INB = A \cdot P(D+, M+) - B \cdot P(M+) - C, \tag{9}$$

where  $A = (e_1 - e_2) - (c_1 - c_2 - c_0)$ ,  $B = c_0$  and  $C = c_m$ . Replacing (??) in (??) yields

$$INB = A \cdot \int_{m_0}^{\infty} m \cdot f_M(m) dm - B \cdot \int_{m_0}^{\infty} f_M(m) dm - C. \tag{10}$$

Taking derivative with respect to  $m_0$  and setting it to zero yields

$$-A \cdot m_0 \cdot f_M(m_0) + B \cdot f_M(m_0) = 0,$$

so that INB is maximized at  $m_0 = B/A$ .

## 9.2 Statistical Inference for INB

Consider a random sample of individuals with both marker ( $M$ ) values and disease status ( $D$ ) available, and let  $n$  denote the sample size. Then, given a cutoff marker value  $m_0$ , the quantities  $P(D+, M+)$  and  $P(M+)$  can be obtained as the sample proportions of the individuals with positive marker and disease present (denoted by  $\hat{P}(D+, M+)$ ), and positive marker (denoted by  $\hat{P}(M+)$ ), respectively. Moreover, the joint distribution of these sample proportions is asymptotically bivariate normal,

$$\left( \hat{P}(D+, M+), \hat{P}(M+) \right)' \sim N \left( (P(D+, M+), P(M+))', \Sigma/n \right), \tag{11}$$

where  $\Sigma_{11} = P(D+, M+)[1 - P(D+, M+)]$ ,  $\Sigma_{22} = P(M+)[1 - P(M+)]$ , and  $\Sigma_{12} = P(D+, M+) \cdot P(M+)$ . It follows that  $\widehat{INB}$ , the sample estimate of INB obtained by replacing  $P(D+, M+)$  and  $P(M+)$  by their sample estimates, is asymptotically normally distributed with mean INB and variance

$$\sigma_{\widehat{INB}}^2 = A^2 \cdot \Sigma_{11}/n + B^2 \cdot \Sigma_{22}/n - 2 \cdot A \cdot B \cdot \Sigma_{12}/n. \tag{12}$$

The results are presented in Figure 3 for Ashkenazi-Jewish population. *Hormuzd, INB in Figure 3 looks a little different than in your Figure 2, not sure whether I got the variables right... Also, I didn't know the sample size for the sample from the general population (I used  $n = 4589$  for AJ) so couldn't plot for the general population.*

Notice that a model-based estimator can be obtained as well by employing the joint model for  $M$  and  $D$  described in (??), and statistical inference can be based on the asymptotic normality of the MLE.

## 9.3 Sensitivity Analyses

### 9.3.1 With respect to input parameters

It is common in cost-effectiveness to see how sensitive the results are (i.e., INB herein) with respect to changes in input parameters. The input parameters are the cost (i.e.,  $c_0$ ,  $c_1$ ,  $c_2$  and  $c_m$ ) and the effectiveness (i.e.,  $e_1$  and  $e_2$ ) parameters. Since INB is linear in these parameters and  $P(D+, M+)$  and  $P(M+)$  are assumed fixed, it is straightforward to assess the sensitivity of the results (e.g., assuming these cost and effectiveness input parameters are normally distributed yields a normally distributed INB).

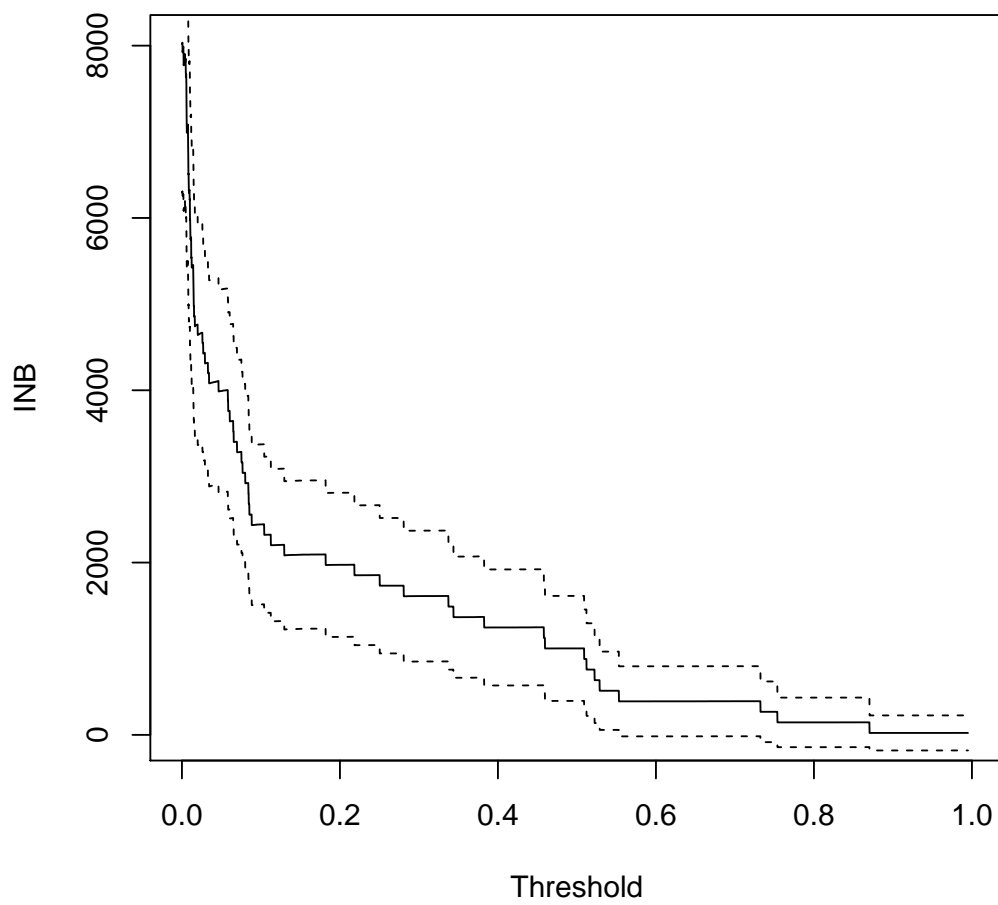


Figure 3: INB (solid line) with 95% confidence intervals (dashed lines) for Ashkenazi-Jews.

### 9.3.2 With respect to the risk-model

*Now that I look at it in this way, the model in Section 3 could be thought of as allowing for error in the risk-model (still unbiased, but with some random error), hence it is a sensitivity analysis. Could you please provide density plots for the risk marker  $M$  in AJ and the general population? Separately, we want to see whether we could employ some parametric model.*

## 10 Generalizing to usual screening with entry criteria and screening test

We have considering only 1 screening test applied to the entire population, with everyone screening positive getting the confirmatory test. But usually the whole population is not considered for screening a priori, only subsets based on age and sex, or ideally, subsets with sufficient risk of disease. Thus there are really two tests for even the simplest screening: the first test is the criteria for entry into screening, the second test is the usual biomarker screening test used to decide who should get the confirmatory test. We call the first test the "rule-out" test  $M_O$  as its main goal is to keep out those whom we are sure have no benefit from screening, and the second test the "rule-in" test  $M_I$  as its goal is to find those who might have disease today and should get the confirmatory test (Castle and Katki, Nat Rev Clin Oncol 2016). Thus we first see if a person is eligible for screening  $M_O+$ , then only those get the screening test  $M_I$ .

All the notation of Section 1 applies, except that now there are separate costs for the screening tests  $c_O$  and  $c_I$ . I also use the notation  $e_0$  for the effectiveness in those without disease, which is normal life-expectancy (this is denoted as "1" in Section 1).

We again calculate the increase in effectiveness and cost due to screening versus no screening. The "no screening" scenario remains the same:

1. The individual has cancer ( $D+$ ): cost is  $c_0 + c_2$  and effectiveness is  $e_2$
2. The individual does not have cancer ( $D-$ ): cost is 0, effectiveness is 1.

Now screening is a two-step "rule-out/rule-in" procedure:

1. We first apply marker  $M_O$ . Those who test  $M_O-$  are not eligible for screening, but the following outcomes occur for them:
  - (a) The individual has cancer ( $D+$ ): cost is  $c_O + c_0 + c_2$  and effectiveness is  $e_2$ . This has probability  $P(D+, M_O-)$ .
  - (b) The individual does not have cancer ( $D-$ ): cost is  $c_O$ , effectiveness is  $e_0$ . This has probability  $P(D-, M_O-)$ .
2. Those who are  $M_O+$  now get  $M_I$ . There are four possibilities:
  - (a) The screening is positive and the individual has cancer ( $M_I+, D+$ ): cost is  $c_O + c_I + c_0 + c_1$  and effectiveness is  $e_1$ . This has probability  $P(D+, M_I+, M_O+)$ .
  - (b) The screening is positive but the individual does not have cancer ( $M_I+, D-$ ): cost is  $c_O + c_I + c_0$  and effectiveness is  $e_0$ . This has probability  $P(D-, M_I+, M_O+)$ .
  - (c) The screening is negative but the individual has cancer ( $M_I-, D+$ ): cost is  $c_O + c_I + c_0 + c_2$  and effectiveness is  $e_2$ . This has probability  $P(D+, M_I-, M_O+)$ .
  - (d) The screening is negative and the individual does not have cancer ( $M_I-, D-$ ): cost is  $c_O + c_I$  and effectiveness is  $e_0$ . This has probability  $P(D-, M_I-, M_O+)$ .

Let  $C$  and  $E$  denote the cost and effectiveness without screening:

$$\begin{aligned}\mathcal{E}(C) &= (c_0 + c_2) \cdot P(D+) \\ \mathcal{E}(E) &= e_2 \cdot P(D+) + P(D-)\end{aligned}$$

With "rule-in/rule-out" screening, the costs and effectiveness are

$$\begin{aligned}\mathcal{E}(C_M) &= c_O + P(M_O+) \cdot c_I + [P(D+) + P(D-, M_I+, M_O+)] \cdot c_0 \\ &\quad + [P(D+) - P(D+, M_I+, M_O+)] \cdot c_2 + P(D+, M_I+, M_O+) \cdot c_1 \\ \mathcal{E}(E_M) &= P(D-) \cdot e_0 + P(D+, M_I+, M_O+) \cdot e_1 + [P(D+) - P(D+, M_I+, M_O+)] \cdot e_2.\end{aligned}$$

Comparing to section 1 for a single test, the effectiveness has the same form: you only get early detection effectiveness  $e_1$  if you are double-positive and have disease  $(D+, M_I+, M_O+)$ , and you get late detection effectiveness  $e_2$  for all other cases of disease  $P(D+) - P(D+, M_I+, M_O+)$ . Similarly, comparing to section 1 for a single test, the cost has the same form: everyone gets eligibility screen  $c_O$ , those who test positive  $M_O+$  get the rule-in screen  $c_I$ , the confirmatory test  $c_0$  is given to everyone with disease and the false-double-positives  $P(D+) + P(D-, M_I+, M_O+)$ , early-detection cost accrues only to those double-positive and with disease  $P(D+, M_I+, M_O+)$ , and late-detection cost accrues to everyone else with disease  $P(D+) - P(D+, M_I+, M_O+)$ .

The increase in effectiveness for screening versus not screening are:

$$\Delta E = (e_1 - e_2) \cdot P(D+, M_I+, M_O+).$$

This is the same as in section 1 for a single test, but now the probability must include  $M_O+$  to identify the only group that gets early-detection, which is the only group whose life can be extended by screening.

The increase in costs for screening versus not screening are:

$$\begin{aligned}\Delta C &= c_O + c_I \cdot P(M_O+) + c_0 \cdot P(D-, M_I+, M_O+) + c_1 \cdot P(D+, M_I+, M_O+) + c_2 \cdot P(D-, M_I+, M_O+) \\ &= c_O + c_I \cdot P(M_O+) + c_0 \cdot P(M_I+, M_O+) + (c_1 - c_2 - c_0) \cdot P(D+, M_I+, M_O+)\end{aligned}$$

The intuition here is the same as for section 1 with a single test. Screening has extra  $c_O$  for everyone, extra  $c_I$  for the screening-eligible ( $M_O+$ ), an extra confirmatory test for the false-double-positives ( $D-, M_I+, M_O+$ ), early-detection cost for those with early-detected disease ( $D+, M_I+, M_O+$ ), and extra late-detection cost for the false-double-positives ( $D-, M_I+, M_O+$ ). The final expression is the same as that in section 1 for a single test, except that we have to include the cost of the extra test among screening-eligibles and include  $M_O+$  in the joint probabilities for the terms involving the confirmatory test and treatment costs.

The  $INB = \Delta E - \Delta C$  is (denoting  $e = (e_1 - e_2) - (c_1 - c_2)$ )

$$INB = (e + c_0) \cdot P(D+, M_I+, M_O+) - c_O - c_I \cdot P(M_O+) - c_0 \cdot P(M_I+, M_O+).$$

This has the same form for INB for a single test, except it includes second test cost  $c_I$  and includes  $M_O$  in all joint probabilities.

## 10.1 Optimal cutoff thresholds for general 1-time screen

Here we try to derive expressions for the 2 thresholds required  $M_O < m_O$  and  $M_I < m_I$  to maximize the  $INB = \Delta E - \Delta C$ . I may well have done this wrong and hope you can see a simple way to a solution.

As in the 1 marker case, I presume we have an unbiased risk calculator, this time involving both markers:

$$P(D+ | M_I = m_I, M_O = m_O) = P(D+ | R(m_I, m_O) = r(m_I, m_O)) = r(m_I, m_O).$$

The annoying thing is that this time there is some function  $r()$  rather than just unbiased risk estimates  $m_I$  or  $m_O$ . Maybe  $r()$  is linear or multiplicative or something, but I think it has to be specified.

The INB in terms of integrals is

$$INB = (e + c_0) \int_{m_O}^{\infty} \int_{m_I}^{\infty} r(m, n) f(m, n) dm dn - c_O - c_I \int_{m_O}^{\infty} f(m) dm - c_0 \int_{m_O}^{\infty} \int_{m_I}^{\infty} f(m, n) dm dn$$

Taking partial derivatives with respect to  $m_I$  and  $m_O$ :

$$\begin{aligned} \frac{\partial INB}{\partial m_I} &= -(e + c_0) \int_{m_O}^{\infty} r(m_I, n) f(m_I, n) dn + c_0 \int_{m_O}^{\infty} f(m_I, n) dn \\ \frac{\partial INB}{\partial m_O} &= -(e + c_0) \int_{m_I}^{\infty} r(m, m_O) f(m, m_O) dm + c_I f(m_O) + c_0 \int_{m_I}^{\infty} f(m, m_O) dm. \end{aligned}$$

Jointly setting these equal to zero and solving looks hard. However, practical screening programs set the rule-in and rule-out thresholds separately rather than jointly. So let's focus on the problem at hand in lung-cancer screening, which is, given that the thresholds for deciding a positive CT is set (i.e say  $m_I$  is known), who should be eligible for screening (i.e. what is the optimal  $m_O$ )? That means we only have  $m_O$  to optimize:

$$\begin{aligned} \frac{\partial INB}{\partial m_O} &= -(e + c_0) f(m_O) \int_{m_I}^{\infty} r(m, m_O) f(m|m_O) dm + c_I f(m_O) + c_0 f(m_O) \int_{m_I}^{\infty} f(m|m_O) dm = 0 \\ 0 &= -(e + c_0) \int_{m_I}^{\infty} r(m, m_O) f(m|m_O) dm + c_I + c_0 \int_{m_I}^{\infty} f(m|m_O) dm. \end{aligned}$$

Writing this again in terms of probabilities

$$0 = -(e + c_0) P(D + |M_I +, M_O = m_O) P(M_I + | M_O = m_O) + c_I + c_0 P(M_I + | M_O = m_O).$$

I had thought that a second simplifying assumption would be based on the fact that the two thresholds are decided independently. But that doesn't mean that  $M_I \pm$  is independent of  $M_O \pm$ . In fact, telling me that you are eligible for screening ought to increase the chance that you will test positive on the screening test. So I'm left with this integral equation that I think requires specifying parametric forms for the probabilities to make further progress.

## Appendix

### MRS is maximized when dichotomizing at disease prevalence

Recall that  $MRS(m_0) = 2J(m_0)\pi(1 - \pi)$ , where  $J(m_0)$  is Youden's index calculated at cutpoint  $m_0$ . MRS is maximized as a function of cutpoint  $m_0$  when Youden's index  $J(m_0)$  is maximized, which occurs when dichotomizing at disease prevalence:  $P(D + |M = m_0) = P(D +) = \pi$ . To prove this we differentiate Youden's index as a function of  $m_0$

$$J(m_0) = \int_{m_0}^{\infty} P(M = m|D+) - P(M = m|D-) dm$$

with respect to the cutpoint  $m_0$  using the Leibniz Integral Rule:

$$\frac{dJ(m_0)}{dm_0} = P(M = m_0|D-) - P(M = m_0|D+).$$

Setting the derivative equal to zero, and using Bayes' rule:

$$\begin{aligned} \frac{P(D + |M = m_0)P(M = m_0)}{\pi} &= \frac{P(D - |M = m_0)P(M = m_0)}{1 - \pi} \\ P(D + |M = m_0) &= \pi \end{aligned}$$



Thus dichotomizing marker/model  $M$  at  $m_0$  at disease prevalence  $\pi$  maximizes Youden's index and thus MRS. Thus the "sweetspot" of risk-thresholds that maximize MRS will always include disease prevalence. At this cutpoint, MRS equals Total Gain (?).