Cost-Effectiveness of a Single Screen Using a Marker/Risk-model

Hormuzd Katki and Ionut Bebu

November 5, 2018

1   Introduction

Consider the problem of a 1-time screen for cancer or other outcome 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).

Katki **?** proposed a new measure of evaluating the risk before and after incorporating the marker information(See the online MRS webtool.). The mean risk stratification (MRS) is the average change in risk (postttest - pretest). Briefly, the 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)

$$
\begin{aligned}
MRS &= [P(D+|M+) - P(D+)] \cdot P(M+) + [P(D-|M-) - P(D-)] \cdot P(M-) \quad (1)\\
&= 2[P(D+,M+) - P(D+) \cdot P(M+)]. \quad (2)
\end{aligned}
$$

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 π 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.

2   Defining costs and effectiveness for a single-time screen

Alternatively, one can introduce a cost-effectiveness framework for this problem as follows. By effectiveness we mean the dollar value of life-expectancy:

1. Let the effectiveness (life-expectancy) for a subject who does not develop cancer be e_0

2. Let e_1 denote the effectiveness (life-expectancy) for a subject with a positive marker M who then tests positive for cancer

3. Let e_2 be the effectiveness (life-expectancy) 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, intuitively $e_1 \geq e_2$. Otherwise, screening causes harm.

For costs:

1. Let c_m denote the cost of marker evaluation

2. Let c_0 denote the cost of the definite test for cancer, such as biopsy, which is perfect but is so invasive and costly that practically it cannot be done on everyone

3. c_1 is the cost of treatment for a cancer case diagnosed early (i.e., tested positive for M)

4. c_2 the treatment cost for a cancer case diagnosed late (i.e., either not screened with M or tested negative for M)

Now we consider the difference in costs and effectiveness of never screening versus a 1-time screen:

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 e_0.

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 e_0.

    (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 e_0.

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

$$
\begin{aligned}
\mathcal{E}(C) &= (c_0 + c_2) \cdot P(D+) \\
\mathcal{E}(E) &= e_2 \cdot P(D+) + e_0 \cdot P(D-) \\
\mathcal{E}(C_M) &= c_m + [1 - P(D-, M-)] \cdot c_0 + P(D+, M-) \cdot c_2 + P(D+, M+) \cdot c_1 \\
\mathcal{E}(E_M) &= e_0 \cdot P(D-) + P(D+, M+) \cdot e_1 + P(D+, M-) \cdot e_2 ,
\end{aligned}
$$

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 ΔC and ΔE. Notice that ΔE is maximized when $m_0 = -\infty$ (i.e., screen everybody).

There are 3 interesting properties of costs and effectiveness:

1. Costs and effectiveness depend on the relationship of disease and marker only through the joint probability $P(D+, M+)$.

2. Only the dollar-value of life-years gained $e_1 - e_2$ requires specification.

3. Only the difference of early vs. late treatment costs $c_1 - c_2$ requires specification.

3   Incremental Net Benefit (INB)

Usually the ICER is calculated, but this has disadvantages (see Appendix). We prefer the INB:

$$
\begin{aligned}
INB &= \Delta E - \Delta C \\
&= [(e_1 - e_2) - (c_1 - c_2 - c_0)]P(D+, M+) - c_m - c_0 P(M+) \\
&= [e + c_0]P(D+, M+) - c_m - c_0 P(M+),
\end{aligned}
\tag{4}
$$

defining the net effectiveness of early intervention: $e = (e_1 - e_2) - (c_1 - c_2)$. The net effectiveness $e = (e_1 - e_2) - (c_1 - c_2)$ contrasts the dollar-value of the life-expectancy gain with dollar increase in treatment cost for early vs. late detection. We presume that $e_1 > e_2$, else screening is unethical. If treatment costs for early vs. late detection are the same ($c_1 = c_2$), then net effectiveness depends only only the value of the life-years gained.

INB trades off the net effectiveness for the only group that benefits from screening ($D+$ and $M+$) versus the cost of the definitive cancer test done on everyone who tests positive. Since the marker is applied to everyone, marker cost is a fixed constant.

To clarify that $INB < 0$ if $e < 0$, we express INB as a function of PPV

$$
INB = P(M+)[e \cdot PPV - c_0(1 - PPV)] - c_m.
$$

4   Optimal cutoff threshold that maximizes INB

The optimal threshold that maximizes INB is easy to calculate. Let M be an unbiased risk-model:

$$
P(D+|M = m) = m.
\tag{5}
$$

Let f_M denote the density of M, and denote the cutoff threshold m_0, so that

$$
\begin{aligned}
P(M+) &= \int_{m_0}^{\infty} f_M(m)\, dm \\
P(D+, M+) &= \int_{m_0}^{\infty} P(D+|M = m) \cdot f_M(m)\, dm \\
&= \int_{m_0}^{\infty} m \cdot f_M(m)\, dm.
\end{aligned}
\tag{6}
$$

INB can be writen as

$$
INB = (e + c_0) \cdot \int_{m_0}^{\infty} m \cdot f_M(m)\, dm - c_0 \cdot \int_{m_0}^{\infty} f_M(m)\, dm - c_m.
\tag{7}
$$

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

$$
-(e + c_0) \cdot m_0 \cdot f_M(m_0) + c_0 \cdot f_M(m_0) = 0,
$$

Denoting m_0 as π_{INB} where INB is maximized:

$$
m_0 = \frac{c_0}{e + c_0} = \pi_{INB}
\tag{8}
$$

This π_{INB} is the optimal risk threshold to maximize INB.

Note that π_{INB} does not require anything about test performance. It only requires the utilities: definitive 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 the optimal risk threshold π_{INB} doesn't require specifying marker cost c_m, because considering risk thresholds requires having the marker test already conducted, and thus it is a sunk cost.

The optimal threshold can also be derived from the classical risk-threshold framework (see Appendix).

4.1 Optimal screening versus screening no one and versus testing everyone

To be sure that π_{INB} maximizes INB, you have check the INB that this provides versus the boundaries: INB for screening no one or everyone. The INB for screening no one is 0 by definition and the INB for testing everyone is $INB = (e + c_0)P(D+) - c_0$. These boundary cases could be calculated by the INB equation **??**, but you have to set $c_m = 0$ because for the bounday cases, screening is unnecessary. Because the optimal threshold π_{INB} also doesn't require c_m, then one could set $c_m = 0$ in the INB equation **??** and you would know which option (screen no one, test everyone, only screen those with risk greater than π_{INB}) is best. The value of knowing c_m is only to know if INB for optimal screening is positive or negative.

Note that the net effectiveness of early intervention must achieve some minimal value for screening to be better than screening no one. However, if treatments are too good (i.e. treatment is very effective and very cheap), then everyone should be tested. Thus for screening to be better than screening no one or testing everyone, the net effectiveness must be bounded (see Appendix):

$$\frac{1}{PPV}\left[\frac{c_m}{P(M+)} + c_0(1 - PPV)\right] < e < \frac{1}{1 - NPV}\left[c_0 \cdot NPV - \frac{c_m}{P(M-)}\right].$$

This might provides a sensitivity analysis for newly discovered markers, where net effectiveness e is not known, but marker cost c_m, definitive test cost c_0, and marker PPV/NPV/positivity are known.

5 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 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 that life-years gained is $1.6313/3 = 0.54377$. At \$100,000 value per life year, this is

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

(Notes for Hormuzd: This $e_1 - e_2$ metric of years-gained among those with cancer, we had decided was the most meaningful in Li's life-gained paper. This may only be life-years gained only within the trial, Li can calculate what this for a life-time, so it should be much bigger (although Li's paper estimates only 1.53 years-gained per lung cancer detected for USPSTF-eligibles).)

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 definitivly 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.)

Thus $e = (e_1 - e_2) - (c_1 - c_2) = \$54,377 - \$625 = \$53,752$. Note that e is insensitive to the costs, but is very sensitive to dollar valuation of life-years gained.

```
[1] "INB = $ 38.0408955333337 , minimum dollars per LYG to have positive INB = $ 153176.375019048"

[1] "cost-effective optimal risk-threshold =  0.00584818121564193"
```

```
[1] "net effectiveness of early intervention e = $ 84996.6666666667"

[1] "INB for NLST vs Lung-RADS is: 46.0323705556822 33.6407146286611 , difference= 12.391655927021

[1] "Lorenz.slope = 1.00421058581553 must be greater than 1.00527839337558 for Lung-RADS to have b

[1] "Lorenz.slope = 0.623787041345575 must be greater than 0.528379167229676 for NLST to have bett
```

The INB is $(e + c_0)P(D+, M+) - c_m - c_0 P(M+)$. CHECK ALL THAT FOLLOWS IT MIGHT BE WRONG The Church NEJM paper reports $P(M+) = 27.3\%$, $P(D+) = 1.1\%$, and I calculate $P(D+, M+) = sens * pi = 93.8\% * 1.093\% = 1.025\%$. The INB is $53752 * 0.01028 - 1965 - 500 * 0.273 = -\$1,549$. This is negative, which means screening should not be done (because INB for no screening is by definition zero, and INB for bronchoscoping everyone is $53752 * 0.011 - 500 = \$91.27$ which is, uh oh, positive.)

Note that the optimal risk-threshold for a CT screen is $\pi_{INB} = 500/(500+53,752) = 0.92\%$ 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.

However, we can change the dollar valuation of life-years gained to see what is required for screening to be cost-effective. Note that $INB > 0$ iff $(e + c_0)P(D+, M+) > c_m + c_0 P(M+) = \$2,102$ CHECK ALL THAT FOLLOWS IT MIGHT BE WRONG and thus $e > \$204,426$. To achieve this e, we require a factor of $204,551/54,377 = 3.76172$, a valuation of \$376,172 per life-year gained. At this e, the optimal threshold for declaring CT-positivity is $\pi_{INB} = 0.24\%$.

6 Example: 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.

6.1 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

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

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:

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

6.2 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 the 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
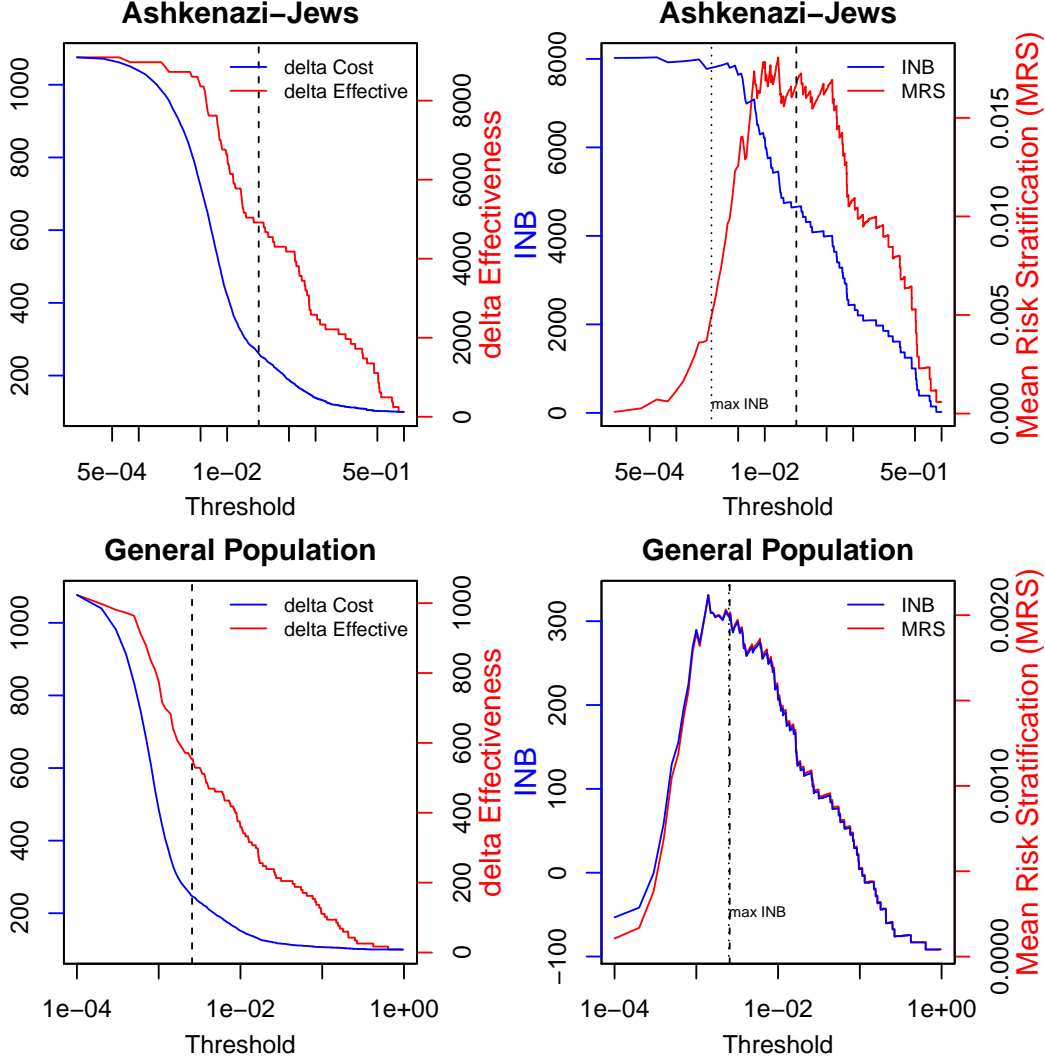
Figure 1: Effectiveness, cost, MRS and INB as a function of the cutoff Threshold values m_0. The marker is the BRCAPRO model score (cutoff at thresholds m_0), the confirmatory test is the BRCA genetic test itself. The outcome $D+$ is having a BRCA mutation. Costs c_1 and c_2 include risk-reducing interventions 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 π_{INB}

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.

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 maximixed 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.

7   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,

$$\left(\hat{P}(D+, M+), \hat{P}(M+)\right)' \sim N\left((P(D+, M+), P(M+))', \Sigma/n\right), \tag{9}$$

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

$$\sigma_{INB}^2 = A^2 \cdot \Sigma_{11}/n + B^2 \cdot \Sigma_{22}/n - 2 \cdot A \cdot B \cdot \Sigma_{12}/n. \tag{10}$$

The results are presented in Figure 2 for Ashkenazi-Jewish 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.

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-effectivenss 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.
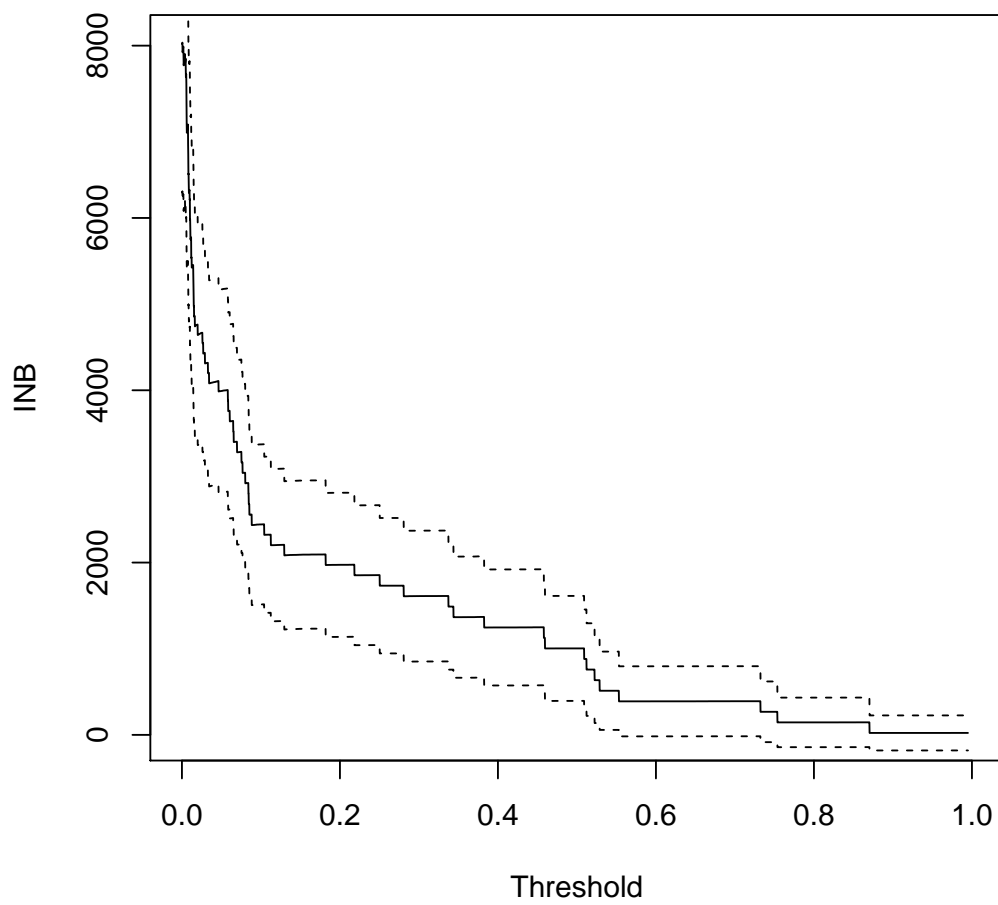
Figure 2: INB (solid line) with 95% confidence intervals (dashed lines) for Ashkenazi-Jews.

For two markers with the same marker cost, then $\Delta INB = INB_1 - INB_2 > 0$ 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 $p_1 > p_2$,

$$
\frac{Sens_1 - Sens_2}{p_1 - p_2} > \frac{c_0/[e + c_0]}{\pi} = \frac{\pi_{INB}}{\pi}, \tag{11}
$$

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 π_{INB}. In reality, the two markers need not be dichotmized 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 π_{INB}, then the comparison is definitive.

To account for unequal marker test costs, $\Delta INB = INB_1 - INB_2 > 0$ if

$$
\begin{aligned}
\Delta INB &= (e + c_0)[P(D+, M_1+) - P(D+, M_2+)] - c_0[P(M_1+) - P(M_2+)] - [c_{m1} - c_{m2}] \\
&= (e + c_0)[Sens_1 - Sens_2]\pi > c_0[p_1 - p_2] + [c_{m1} - c_{m2}] \\
&= [Sens_1 - Sens_2]\pi/\pi_{INB} > [p_1 - p_2] + [c_{m1} - c_{m2}]/c_0
\end{aligned}
$$

for disease prevalence $P(D+) = \pi$ and test-positivities p_1, p_2. The required condition is

$$
[Sens_1 - Sens_2] > \frac{\pi_{INB}}{\pi}\left([p_1 - p_2] + \frac{c_{m1} - c_{m2}}{c_0}\right) \tag{12}
$$

$$
. \tag{13}
$$

If marker test costs are equal, the right size is zero, and this reduces to the earlier simple condition. Here the right side is the ratio of the increase in marker test cost to the definitive test cost. For a marker 1 with better sensitivity and smaller test positivity, in the case of equal costs or a costless marker is a no-brainer better marker. But this is not necessarily true if the increase in marker cost compared to definite test cost is very high.

8.1 Sensitivity Analyses

8.1.1 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).

8.1.2 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.

Appendix

Incremental Cost Effectiveness Ratio (ICER)

Cost-effectiveness is usually measured by the ICER:

$$
\begin{aligned}
ICER &= \frac{\Delta C}{\Delta E} \\
&= \frac{c_1 - c_2 - c_0}{e_1 - e_2} + \frac{c_0}{(e_1 - e_2) \cdot PPV} + \frac{c_m}{(e_1 - e_2)P(D+, M+)} \\
&= \frac{1}{e_1 - e_2}\left[(c_1 - c_2 - c_0) + \frac{1}{PPV}\left(c_0 + \frac{c_m}{P(M+)} \right) \right].
\end{aligned}
\tag{14}
$$

The advantage of ICER is that you only need to specify the ratio of the 3 cost functions (c_m, c_0, and $(c_1 - c_2 - c_0)$) to the life-years gained ($e_1 - e_2$). However, I don't know how to specify these ratios without specifying the costs and effectiveness separately. Furthermore, a large ICER can mask a small difference, and vice-versa. For these reasons, the Incremental Net Benefit (INB) is more useful.

Bounds on net effectiveness so that screening is better than screening no one or testing everyone

For screening to be cost-effective versus screening no one, then $INB > 0$, which happens only if

$$
e > \frac{c_m + c_0 P(M+)}{P(D+, M+)} - c_0 = \frac{1}{PPV}\left[\frac{c_m}{P(M+)} + c_0(1 - PPV) \right].
$$

The minimal net effectiveness is zero only if the marker cost c_m and definitive test cost c_0 are zero, else is a positive number. The minimal net effectiveness required increases with screening-positivity, but decreases with PPV. If the screening test were free so $c_m = 0$ (e.g the screening test were a risk-model based on

easily elicited risk factors), then only PPV matters. This expression holds for any threshold, including the optimal π_{INB}.

For screening to be better than testing everyone, we need $INB > (e + c_0)P(D+) - c_0$:

$$
\begin{aligned}
0 &< [e + c_0]P(D+, M+) - c_m - c_0 P(M+) - [(e + c_0)P(D+) - c_0] \\
0 &< -[e + c_0]P(D+, M-) - c_m + c_0 P(M-) \\
e &< \frac{c_0 P(M-) - c_m}{P(D+, M-)} - c_0 = \frac{1}{1 - NPV}\left[c_0 \cdot NPV - \frac{c_m}{P(M-)}\right].
\end{aligned}
$$

Note that this is an upper-bound on net effectiveness e. What this means is, if early treatment is too cheap and effective, then just test everyone.

Thus for screening to be better than screening no one or testing everyone, the net effectiveness must be bounded:

$$
\frac{1}{PPV}\left[\frac{c_m}{P(M+)} + c_0(1 - PPV)\right] < e < \frac{1}{1 - NPV}\left[c_0 \cdot NPV - \frac{c_m}{P(M-)}\right].
$$

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

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

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

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

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

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 π_{INB} to indicate that this is the disease prevalence where MRS and INB yield the same optimal threshold. In Figure 1, $c_0/e = \$1,000/(\$400,000 - \$0) = 0.25\%$, 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.

Deriving the optimal risk threshold from

The optimal risk threshold can be derived from the classical expected utility approach (see **?** for a comprehensive review). Calculating expected utility requires specifying the utility for the 4 possible outcomes: the utility of true positive prediction U_{TP}, the utility of true negative prediction U_{TN}, the utility of false positive prediction U_{FP}, and the utility of false negative prediction U_{FN}. The cost of marker M is U_{Test}.

The marker/model is dichotomized at risk threshold R: $P(D+|M = m_0) = R$. The R that maximizes expected utility is determined by the ratio of benefit ($B = U_{TP} - U_{FN}$) to costs ($C = U_{TN} - U_{FP}$), which

is $R = 1/(1 + B/C)$?. This is the basis for recent developments in risk prediction metrics, such as Net Benefit.

In current approaches, the utilities are left unspecified and all possible risk thresholds are considered as a sensitivity analysis.

Next Steps

1. Do this for IFNL4 genotype

2. Extend to screen-and-treat program for precursors, application to cervical screen-and-treat

3. Extend to also having entry criteria for screening and jointly optimize threshold for entry and threshold for screening test

4. Extend to allowing a triage test for the primary screen, jointly optimize both thresholds

5. Extend to multiple screens over time