Simple and intuitive optimal risk thresholds for a single screen: The next step after a Decision Curve Analysis

Hormuzd A. Katki[1] and Ionut Bebu[2]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD
[2]Biostatistics Center, George Washington University, Rockville, MD, USA

February 14, 2019

Abstract

Introduction: Decision Curve Analysis (DCA) is a modern popular approach to identify the range of risk thresholds at which screening with a biomarker is the best decision. However DCA does not directly account for costs and effectiveness of subsequent interventions, and most applications assume a costless screening biomarker for simplicity. We develop a simple framework that accounts for costs and effectiveness to identify optimal cost-effective thresholds for a single screen, and compare to findings from DCA.

Methods: We calculate the Incremental Net Benefit (INB) for a single-time screen as a function of costs (screening test, definitive confirmatory test, and subsequent interventions) and effectiveness (dollar value of life-years gained by undergoing interventions). We provide a simple expression for the optimal cost-effective risk-threshold and for the dollars per life-year gained threshold associated with an optimal risk-threshold. We apply our framework to the controversy over using the BRCAPRO risk model to screen women for mutations in *BRCA1/2* that cause high risks of breast and ovarian cancers. Mutations are rare in the general population, but more common among Ashkenazi-Jews.

Results: According to DCA, assuming costless BRCAPRO, optimal risk thresholds from 1%-30% support using BRCAPRO to screen among Ashkenazi-Jews. However, our framework shows that these optimal risk-thresholds are infeasible because they imply a negative dollars per life-year gained threshold. Furthermore, DCA is sensitive to a small cost of shared decision-making with BRCAPRO ($100), which is hard to detect by DCA but easily seen by INB. In the general population, optimal risk thresholds from 0.12% to 0.91% support screening. However, these optimal thresholds represent dollars per life-year gained thresholds of $16,000 to $306,000, suggesting that none or only some of the optimal thresholds are feasible in practice, where thresholds below $100,000 are typical.

Limitations: Our framework requires specifying costs of tests and interventions, and the value of life-year gained.

Conclusions: Our simple framework identifies sensitivity of DCA to biomarker test costs that is hard to see by DCA. Optimal risk thresholds identified by DCA may be infeasible based on their associated dollars per life-year gained threshold. Our simple approach is the next step after DCA to better identify optimal cost-effective risk-thresholds.

Keywords: AUC; *BRCA1*; *BRCA2*; cost effectiveness; decision analysis; Decision Curves; Diagnostic Testing; Incremental Net Benefit; Net Benefit; risk prediction ; ROC; Screening; Youden's index.

1 Introduction

In population disease screening, a cheap, non-invasive, but imperfect test is conducted on a population to determine who should be referred for tests that are definitive for the presence of disease. Because screening tests are imperfect indicators of the presence of disease, the goal is choose a screening test threshold that is cost-effective at controlling the rates of false classification.

There is much literature on identifying cost-effective thresholds, such as the classic paper of Pauker and Kassirer[25]. The cost-effective threshold depends on the costs, benefits, and harms of the tests and subsequent interventions. Subsequent mathematical developments conglomerate costs, benefits and harms into utilities that are even harder to specify[2,8]. However, statistical metrics that do not require costs, benefits, and harms, such as Youdens index, AUC, and Mean Risk Stratification, cannot identify cost-effective thresholds. The important recent advance of Net Benefit[34] in practice typically assumes costless screening tests, and considers risk-thresholds as a sensitivity analysis and thus also cannot identify cost-effective thresholds.

In contrast, a full decision analysis, specifying costs, benefits, and the monetary value of life, can determine cost-effective thresholds. However, usually these analyses require complex decision trees and microsimulation to evaluate the situation as comprehensively as possible. In such analyses, it can be hard to glean intuition about which quantities are most important, even when probabilistic sensitivity analysis is conducted. To complement such comprehensive decision analyses, there is a need for an approach to risk thresholds that is simple, transparent, provides a reasonable estimate of the range of realistic cost-effective thresholds, and provides simple intuition on which quantities are most important.

We propose a simple framework that identifies the cost-effective risk threshold for a single-time screen, such as screening for gestational diabetes, syphilis in pregnancy, or genetic mutations. We show that the Incremental Net Benefit (INB) depends on only 3 quantities: the costs of the screening test and definitive test, and a parameter we call the net effectiveness of early intervention (NE). We show that the cost-effective risk-threshold for the screening test depends on only the cost of the definitive test and the NE. Our framework also provides a simple criterion, based on the slope of the Lorenz curve, for judging when one screening test is more cost-effective than another.

We apply our framework to the controversy over the risk-threshold to identify who should get testing for mutations in the *BRCA1/2* genes, which cause high risks of breast and ovarian cancer[17]. Current US and UK guidelines refer women for mutation testing only if they have a strong family history of breast or ovarian cancer[22], as quantified by a risk model calculating that their risk of carrying a *BRCA1/2* mutation exceeds 10%[23]. However, as mutation-testing costs plummet, prominent voices have called for testing *all* women[14], equivalent to a 0% risk-threshold. The Screen Project offers *BRCA1/2* testing to any Canadian[32]. In a landmark ruling, the US FDA recently authorized direct-to-consumer genetic testing, for *BRCA1/2* mutations among Ashkenazi-Jews[33]. Population mutation screening is inevitable, but there are strong disagreements on how such programs should be conducted.

We recently showed that 80% of Ashkenazi-Jewish mutation-carriers could be identified by testing only 44% of Ashkenazi-Jewish women[3] at highest risk for carrying mutations, as identified by the BRCAPRO[24] risk model. This is achieved by a low mutation-risk threshold of 0.78%, far lower than the current 10% US and UK guideline. We previously examined multiple statistical metrics, such as AUC, Net Benefit, and Mean Risk Stratification, to examine properties of different risk thresholds. However, no statistical metric can formally identify optimal cost-effective risk-thresholds. In this paper, we input costs and benefits into our simple framework to identify optimal risk thresholds to screen women for referral for *BRCA1/2* genetic testing. Using our simple framework, our findings are comparable to more comprehensive, but more complex and less transparent, microsimulation

approaches to this issue[1,19]. In this example, we show that Net Benefit is sensitive the standard assumption of a costless screening test and the ranges of risk thresholds identified as being favorable for screening can imply unrealistic values for dollars per life-year gained.

2 A cost-effectiveness framework for a single-time screen

2.1 Defining average costs and effectiveness for a single-time screen

In screening, asymptomatic people are given the single screening test. Two outcomes are possible:

1. If the test is positive, the person is referred for definitive disease testing. If disease is diagnosed, it is then treated. Because the screening test found the disease early (i.e. in the absence of symptoms), presumably treatment is more effective than for disease found later on the basis of symptoms.

2. If the test is negative, there is no further action. However, future symptoms will prompt definitive testing to diagnose disease. If disease were diagnosed in the future, presumably treatment is less effective than it would have been had disease been found earlier.

Four parameters define the average costs for a single screen:

1. Let c_m denote the cost of screening marker evaluation. The screening test is imperfect, else it would be the definitive test. The main goal of our paper is to identify the cost-effective threshold for deciding screening test positivity.

2. Let c_0 denote the cost of the definitive test for disease, e.g. biopsy or imaging. We presume this test is perfect but is so invasive and costly that practically it cannot be the screening test.

3. Let c_1 be the cost of treatment for a disease case diagnosed early. This occurs for those testing positive for disease on both the screening test and definitive test.

4. Let c_2 be the treatment cost for a cancer case diagnosed late. This occurs for those with negative screening test (or not screened at all), but in the future develop symptoms and disease is diagnosed by the definitive test.

Three parameters define the average effectiveness of a single screen, defining effectiveness as the value of the gain in life-expectancy:

1. Let e_0 be the life-expectancy for someone who never develops disease

2. Let e_1 be the life-expectancy for someone with a positive marker M who then tests positive for disease. This is early-detection and is the benefit of screening.

3. Let e_2 be the life-expectancy for someone who is either not screened or tests screen-negative, but then develops disease. This is late detection and is a screening failure. We presume that $e_1 \geq e_2$, else screening causes obvious harm.

Finally, the effectiveness value must be on the same scale as costs, and we call d the dollar-value of 1 life-year. In our examples of population mutation-screening and CT lung-cancer screening, we derive the 4 cost and effectiveness parameters e_1 and e_2 (we show that e_0 does not require specification). We propose sensitivity analyses for costs and effectiveness and, in the Discussion, we consider practical weaknesses of our approach to identifying costs and effectiveness.

2.2 Calculating Incremental Net Benefit (INB)

The INB is the difference in average costs and effectiveness of a 1-time screen versus never screening. Define disease as D and the marker for the screening test as M.

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

   (a) The individual has disease $(D+)$: cost is $c_0 + c_2$ and effectiveness is $d \cdot e_2$

   (b) The individual does not have disease $(D-)$: cost is 0, effectiveness is $d \cdot e_0$.

2. Screening for M. The marker is dichotomized as positive $(M+)$ and negative $(M-)$ based on a underlying threshold m_0 (our goal is to identify the cost-effective m_0). There are four possibilities:

   (a) True Positive: The screening is positive and the individual has disease $(M+, D+)$. Cost is $c_m + c_0 + c_1$ and effectiveness is $d \cdot e_1$.

   (b) False Positive: The screening is positive but the individual does not have disease $(M+, D-)$. Cost is $c_m + c_0$ and effectiveness is $d \cdot e_0$.

   (c) False Negative: The screening is negative but the individual has disease $(M-, D+)$. Cost is $c_m + c_0 + c_2$ and effectiveness is $d \cdot e_2$.

   (d) True Negative: The screening is negative and the individual has not have disease $(M-, D-)$. Cost is c_m and effectiveness is $d \cdot e_0$.

Let C and E be the average cost and effectiveness w/o screening, and C_M and E_M with screening:

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

The differences in cost and effectiveness between screening and never screening are:

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

The Incremental Net Benefit is

$$
\begin{aligned}
INB &= \Delta E - \Delta C \\
&= [d(e_1 - e_2) - (c_1 - c_2 - c_0)]P(D+, M+) - c_m - c_0 P(M+) \\
&= [e + c_0]P(D+, M+) - c_m - c_0 P(M+),
\end{aligned}
\tag{1}
$$

where $e = d(e_1 - e_2) - (c_1 - c_2)$ is the "net effectiveness of early intervention". The net effectiveness contrasts the dollar-value of the life-expectancy gain with dollar increase in treatment cost for early vs. late detection. If net effectiveness $e < 0$, screening is never cost-effective (See Appendix).

The INB reveals that only certain aspects of costs and effectiveness require specification. First, only the dollar-value of life-years gained $d(e_1 - e_2)$ requires specification. Life-expectancy in the absence of disease (e_0) is not needed. Second, only the difference of early vs. late treatment costs $c_1 - c_2$ requires specification. Importantly, if treatment costs were similar regardless of early versus late detection, or if treatment costs are far below the dollar value of life-years gained, then net effectiveness approximately depends only only the value of life-years gained.

4

2.3 Optimal cost-effective risk threshold that maximizes INB

The optimal risk threshold can be derived from the classical expected utility approach[25]. 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}. These utilities are the difference of effectiveness and cost for each of the 4 outcomes in Section 2.1:

$$
\begin{aligned}
U_{TP} &= d \cdot e_1 - (c_m + c_0 + c_1) \\
U_{FN} &= d \cdot e_2 - (c_m + c_0 + c_2) \\
U_{TN} &= d \cdot e_0 - (c_m) \\
U_{FP} &= d \cdot e_0 - (c_m + c_0)
\end{aligned}
$$

The marker/model is dichotomized at cutpoint m_0 that defines a risk threshold R: $P(D+|M = m_0) = R$. The R that maximizes expected utility is determined by benefit $(B = U_{TP} - U_{FN})$ and cost $(C = U_{TN} - U_{FP})$, and is known to be $R = C/(B+C)$[25]. For a single screen, note that $B = e$ and $C = c_0$. Thus the optimal cost-effective risk-threshold is

$$
R = \frac{C}{B + C} = \frac{c_0}{e + c_0}. \tag{2}
$$

The WebAppendix derives this optimal threshold from first principles.

The value of this expression is that it does not require 4 abstract uncalculable utilities, but only 2 concrete calculable quantities: the definitive test cost c_0 and the net effectiveness of early intervention $e = d(e_1 - e_2) - (c_1 - c_2)$. In the following examples, we obtain c_0 and e from data to directly propose optimal thresholds for lung-cancer screening and population mutation-screening.

Also, neither screening test characteristics nor disease prevalence matter for calculating the optimal risk-threshold. This is unlike metrics such as MRS, Youden's index, AUC, Vickers's Net Benefit gain, which are all maximized at risk threshold equals disease prevalence (i.e $R = P(D+)$), regardless of costs and effectiveness.

Also, the optimal risk threshold R does not require the screening test cost c_m, because risk thresholds only matter after the screening test is conducted, and thus the screening test cost is a sunk cost. However, the screening test cost matters to check if screening is better than screening no one or testing everyone (see Section 2.4).

2.4 Gain in INB versus screening no one or testing everyone

If $INB(R) > 0$, then screening based on the marker is more cost-effective than screening no one. To also check if screening is also more cost-effective than just conducting the definitive test on everyone, we have to compare the INB versus the INB for testing everyone versus screening no one, called INB_{all}. For testing everyone, we have $P(D+, M+) = P(D+)$, $P(M+) = 1$, and $c_m = 0$ since we don't need the screening test. Plugging these into equation 2 yields $INB_{all} = (e + c_0)P(D+) - c_0$. The INB_{Gain} is the increase in INB beyond either screening no one or testing everyone:

$$
INB_{Gain}(R) = INB(R) - max(0, INB_{all}). \tag{3}
$$

Risk thesholds where $INB_{Gain}(R) > 0$ are thresholds where screening is better than screening no one or testing everyone.

2.4.1 Comparison of INB to Net Benefit (NB)

Net Benefit (NB) is a well-known and important approach to identify risk thresholds where using a test to decide who should get treated is better than treating no one or everyone[34]. NB equals the INB standardized by benefit $B = e$ (see Appendix). We define NB_{Gain} as the INB_{Gain} standardized by benefit $B = e$, as a function of optimal risk threshold R:

$$NB_{Gain}(R) = NB(R) - max(0, INB_{all}/e) = NB(R) - max(0, P(D+)/(1-R) - R).$$

The thresholds R where $NB_{Gain}(R) > 0$ are thresholds where the risk model is better than screening no one or testing everyone. These are the same thresholds as where $INB_{Gain}(R) > 0$. The advantage of NB over INB is that, for a costless test $c_m = 0$, one does not need to specify net effectiveness of early intervention e (see Appendix), so that only risk threshold R is required. All applications of NB we are aware of assume a costless test. However, every test in reality has some cost. We will compare INB to NB to show the value of explicitly specifying costs and effectiveness.

3 Example: Who should undergo *BRCA1/2* genetic testing?

3.1 Background

As detailed in the Introduction, mutations in the *BRCA1/2* genes cause high risk of breast and ovarian cancers. The mutations are rare in the general population ($\approx 0.26\%$), but much more common among Ashkenazi-Jews ($\approx 2.3\%$). Currently, women are asked to provide their family history of cancer (e.g. see Webappendix), and she is offered mutation-testing in the UK and US if a risk model calculates that her risk of carrying a mutation exceeds 10

However, as mutation-testing costs plummet, prominent voices have called for testing all women, which would strain clinical resources by testing millions of women, 99.75% of whom will test negative. Instead, a lower risk threshold, below 10%, might identify most all mutation-carriers, yet avoid unnecessary testing for most women. We recently showed that a low 0.78% risk-threshold would identify 80% of Ashkenazi-Jewish mutation-carriers yet test only 44% of Ashkenazi-Jewish women[3]. However, we could not formally justify any choice of risk-threshold. Here we examine cost-effectiveness of possible risk thresholds.

We used data on 4,589 volunteers (102 *BRCA1/2* mutation carriers) from the Washington Ashkenazi Study (WAS)[30]. 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[24]. Here M is the BRCAPRO risk score, and because BRCAPRO is a well-calibrated risk model[3], $m_0 = R$, i.e. the cutpoint m_0 equals the risk threshold R. Disease D indicates the presence of a *BRCA1/2* mutation. We use previously published methodology to use WAS data to also calculate BRCAPRO risk scores valid for the general population[3] (see WebAppendix).

3.2 Costs and effectiveness for *BRCA1/2* mutation screening

We set life-years gained $e_1 - e_2$ as 5.0 life-years, which is the estimate for removing both ovaries and breasts (followed by frequent cancer screening) in a 50-year old mutation-carrier who has completed childbearing. We use costs used by the Long and Ganz decision analysis from 2015[19] to compare results (In section ?? we update to current costs). We set c_m as the cost of using the BRCAPRO risk model, which is "free" but requires a doctor's visit to assess family history, costing roughly $c_m = \$100$. For the cost of BRCA genetic testing c_0, for Ashkenazi-Jews, the cost is now $c_0 = \$249$, but non-Ashkenazim require full gene sequencing which costs $c_0 = \$2200$. For treatment costs, c_1
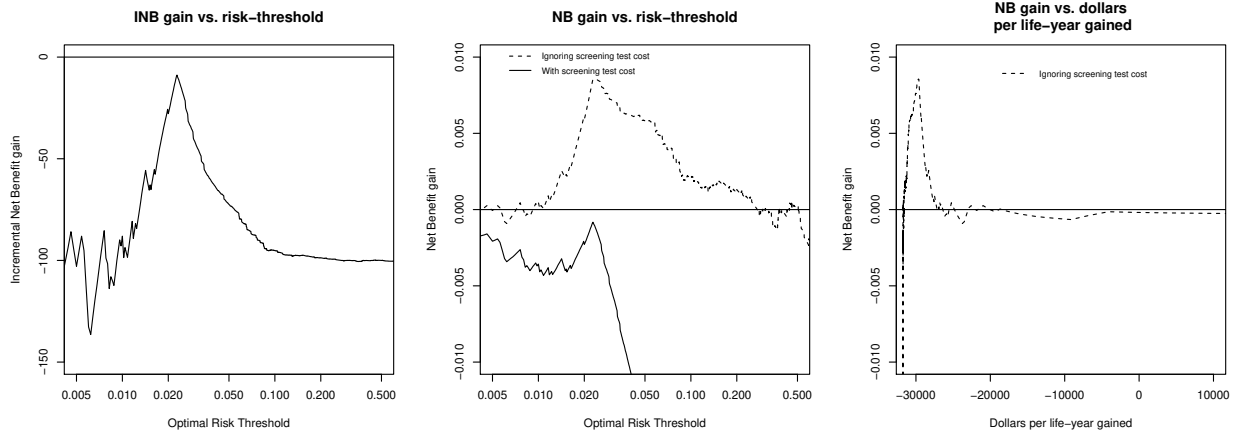
Figure 1: INB gain (left panel) and NB gain (middle panel) vs. BRCAPRO risk-threshold for using the BRCAPRO risk model to screen for which Ashkenazi-Jews should get *BRCA1/2* genetic testing. The right panel plots NB gain for a costless BRCAPRO screening test versus the dollars per life-year gained implied by each optimal risk-threshold.

is the cost of removing the breasts and ovaries and the average treatment cost for the rare breast or ovarian cancers that occur anyway. For c_2, this is the cost of cancer treatment for mutation carriers who are not found until cancer is diagnosed. By[19], $c_1 = \$28393$ and average cancer treatment cost is $c_2 = \$187110$ (see WebAppendix).

3.3 *BRCA1/2* mutation screening for Ashkenazi Jews

3.3.1 INB gain versus NB gain

Figure 1 shows that the INB gain (left panel) and NB gain with BRCAPRO screening test cost $c_m = \$100$ (middle panel) are always negative. Thus screening is never the best choice. However, the NB gain estimated by usual costless test assumption (middle panel) shows that screening is the best choice for a wide range of risk thresholds from 1% to 30%. Thus *BRCA1/2* mutation screening is sensitive to a small BRCAPRO screening test cost. It is hard to see this sensitivity from the costless NB gain because its units are abstract. In contrast, this sensitivity is easy to see from INB gain because its units are dollars. The INB asymptotes at -$100, clarifying that a screening test that was just $100 cheaper would permit screening at a wide range of risk-thresholds.

7

3.3.2 Dollars per life-year gained implied by optimal risk-thresholds

Fixing costs and life-years gained, an optimal risk threshold implies a threshold for dollars per life-year gained. Because

$$R = \frac{c_0}{e + c_0} = \frac{c_0}{d(e_1 - e_2) - (c_1 - c_2) + c_0},$$

where d is the dollars per gain in life-years $(e_1 - e_2)$, then

$$d = \frac{c_0(1-R)/R + (c_1 - c_2)}{e_1 - e_2}. \tag{4}$$

Figure 1 (right panel) plots NB (for a costless BRCAPRO test) versus the threshold for dollars per life-year gained. Surprisingly, the wide range of risk thresholds supporting screening for a costless BRCAPRO test implies that the threshold for dollars per life-year gained is *negative*. This means that screening with BRCAPRO at optimal thresholds actually costs more than doing genetic testing on all Ashkenazi-Jewish women, which agrees with other decision analyses[1,19]. This striking example demonstrates that optimal risk thresholds can imply unrealistic thresholds of dollars per life-year gained and demonstrates a weakness in any approach that examines only risk thresholds and not dollars per life-year gained thresholds. Note that dollars per life-year gained thresholds cannot be calculated by Net Benefit and requires specifying costs and effectiveness.

The reason for the negative dollars per life-year gained is that $c_1 - c_2 = -\$158717$ is strongly negative. This is not overcome by the cost increase due to genetic testing because c_0 is only \$249, the smallest optimal threshold is $R \approx 1\%$, and thus $c_0(1-R)/R \approx \$25000$. Thus we immediately see that the reason why testing all Ashkenazi Jews saves money versus screening with BRCAPRO is because treating cancer is far more expensive than all costs incurred by genetic testing.

Our simple framework easily allows for other calculations of interest. For example, the maximum cost of genetic testing to ensure that testing everyone saves money versus screening at an optimal risk threshold of 1% is approximately $(c_1 - c_2)R/(1 - R) = \$158717/99 \approx \$1600$.

3.4 *BRCA1/2* mutation screening for the general population

In the general population, *BRCA1/2* mutations have only 0.26% prevalence (10 times fewer than in Ashkenazi-Jews). Figure 2 (left panel) shows that accounting for just a \$100 screening BRCAPRO test cost greatly reduces the range of allowable risk thresholds for screening, from $(0.12\%, 3.1\%)$ to $(0.13\%, 0.91\%)$. Figure 2 (right panel) shows that screening with BRCAPRO is cost-effective at dollars per life-years gained thresholds ranging from $(\$16,000, \$306,000)$, which spans typical thresholds for Western countries. The peak INB of \$700 shows that a BRCAPRO screening that cost \$700 more (total $c_m = \$800$) would be required to wipe out any possibility of screening, which is unreasonably expensive for shared decision-making.

The more comprehensive analysis of Long and Ganz[19] suggests that testing everyone is cost-effective at \$920,000 per QALY. Most of this increase is due to quality adjustment, which roughly halved the value of life-years gained. If we halve the life-years gained due to screening (to 2.5), then testing everyone is cost-effective at \$612,000 per life-year gained, much closer to Long and Ganz[19]. The complementary value of our approach is that it is

The dollars per life-year gained at peak INB is \$137,000, which is high even for the US. Figure 3 shows the performance of screening along the range of allowable thresholds of dollars per life-year gained. As dollars per life-year gained increases, so does the fraction of mutation carriers identified (sensitivity; from 40% to 67%) but also the fraction of women referred for genetic testing (6.4% to
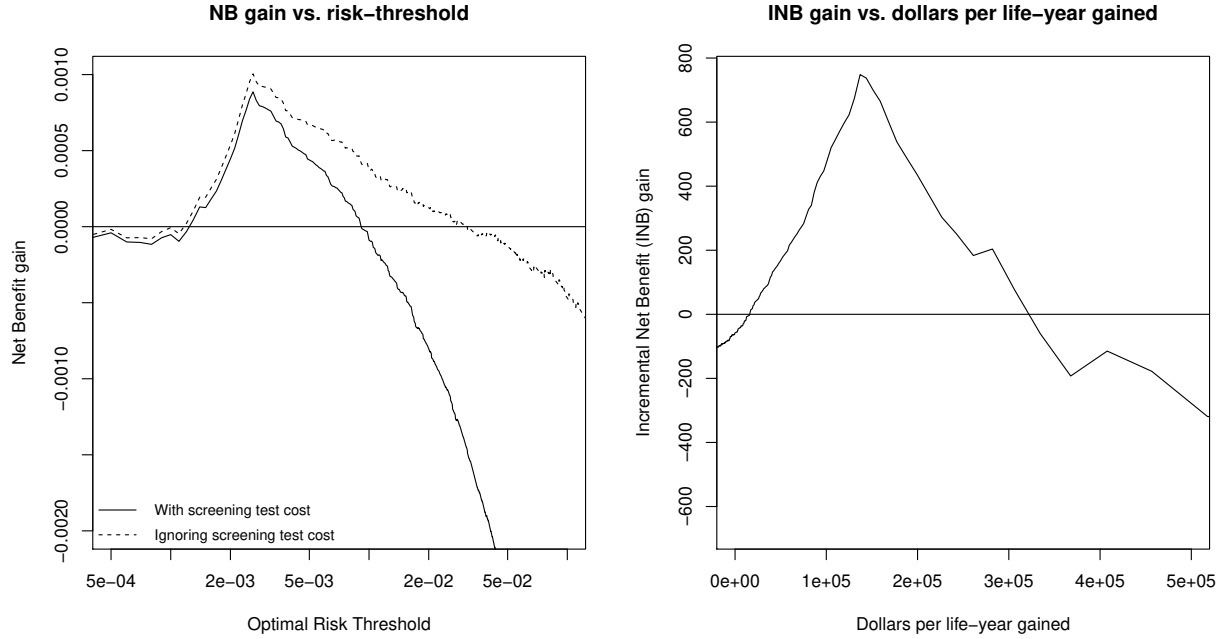
Figure 2: NB gain vs. risk-threshold (left panel) and INB gain vs. dollars per life-year gained (right panel) for using the BRCAPRO risk model to screen for which members of the general population should get *BRCA1/2* genetic testing.

Dollars/LYG	Threshold	INBgain	Positivity	PPV	cNPV	Sens	Spec	AUC
20197	0.0084	27.74	0.06377	0.016048	0.001640	0.4000	0.9371	0.6685
50835	0.0053	171.90	0.09364	0.012295	0.001553	0.4500	0.9073	0.6786
101150	0.0033	482.62	0.12929	0.010059	0.001445	0.5083	0.8717	0.6900
137047	0.0026	748.35	0.14741	0.009401	0.001375	0.5417	0.8536	0.6976
199395	0.0019	436.65	0.18377	0.008006	0.001332	0.5750	0.8172	0.6961
306278	0.0013	76.46	0.28016	0.006164	0.001155	0.6750	0.7209	0.6979

Figure 3: Peformance of using BRCAPRO to screen for *BRCA1/2* mutations at various dollars per life-year gained thresholds that favor screening in the general population.

28%). Note that sensitivities below 40% do not support screening (instead screen none: sensitivity is 0%) and neither do sensitivities above 67% (instead test everyone; sensitivity is 100%) The predictiveness of BRCAPRO increases from AUC of 0.669 to 0.698 at peak INB, then levels off.

A simple calculation also shows that the maximum cost of genetic testing for testing everyone to save money versus screening at an optimal risk threshold of 0.26% is approximately $158717/384 \approx$ $413. Although genetic sequencing costs are rapidly decreasing, it may be some time before $400 can be profitably achieved. Our simple framework easily enables such calculations.

4   Simple condition for a screening test to have greater INB

IONUT: I DON'T THINK THIS FITS IN HERE. I THINK IT SHOULD BE ITS OWN PAPER.

Usually one screening test has better sensitivity but another has better specificity. Informational metrics like Youden's index, MRS, AUC, and NB do not explicitly consider costs and effectiveness and are all optimized when risk threshold equals disease prevalence. Ideally, the sensitivity vs. specificity tradeoff would be evaluated considering costs and effectiveness via the INB.

For two markers M_1 and M_2 with equal 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 $P(D+) = \pi$, sensitivities $Sens_i = P(M_i + |D+)$, and test-positivities $P(M_i+) = p_i$. Choosing test 1 so that $Sens_1 > Sens_2$, then presuming $p_1 > p_2$,

$$\frac{Sens_1 - Sens_2}{p_1 - p_2} \quad > \quad \frac{c_0/[e + c_0]}{\pi} = \frac{R}{\pi}. \tag{5}$$

The simple condition examines if the slope of the concentration curve is less than a constant. This constant is the ratio of the optimal risk threshold R 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 and effectiveness versus ignoring them. If $R = \pi$, then increases in sensitivity and test-positivity are weighed equally. If $R < \pi$ which means we should be testing more people, then we will accept smaller sensitivity gain for a larger test-positivity increase. But if $R > \pi$, meaning we should be testing fewer people, then we require a greater sensitivity gain than a test-positivity increase. The WebAppendix shows an equivalent condition based on specificity.

Note that if $p_1 < p_2$, then the inequality flips and the left side is negative while the right is positive: test 1 is always better if it has higher sensitivity yet lower test-positivity (equivalent to higher sensitivity and higher specificity), as long as the two screening test costs are equal. However, if the two screening test costs are unequal, $\Delta INB = INB_1 - INB_2 > 0$ if

$$[Sens_1 - Sens_2] \quad > \quad \frac{R}{\pi}\left([p_1 - p_2] + \frac{c_{m1} - c_{m2}}{c_0}\right), \tag{6}$$

where c_{mi} is the cost of screening test i. Compared to equation 5, this equation also includes the ratio of the increase in screening test cost to the definitive test cost. Unlike for equal test costs, if screening test 1 has both better sensitivity and smaller test positivity, but a big increase in screening test cost compared to definite test cost, then test 1 may be too expensive to be cost-effective.

4.1   Examples

[3] showed a concentration curve for the fraction of *BRCA1/2* mutation carriers identified (sensitivity) versus the fraction of women who test positive by BRCAPRO. They focus on two thresholds: a "Test 1" that identifies 90% of mutation carriers by testing 60% of women, and a "Test 2" that identifies 80% of mutation carriers by testing 44% of women. The AUC for Test 1 vs. Test 2 is 0.658 vs 0.688, respectively. We examine which threshold is more cost-effective.

(IONUT: the weakness of these examples is that 80% or 90% sensitivity is never preferred versus testing everyone. Maybe I need to modify).

Since these are two thresholds for the same test (BRCAPRO), the test costs are by definition equal. Given the effectiveness and cost for screening 50-year old women (section 3.2) and setting dollars per life-year gained at $d = \$50,000$, then $e = \$408,717$ (we fix this throughout this section). Thus $R = 2200/(408717 + 2200) = 0.54\%$. Since $BRCA1/2$ mutation prevalence is $\pi = 0.26\%$, the critical slope is $R/\pi = 2.1$. The slope of the concentration curve is $(0.9 - 0.8)/(0.6 - 0.44) = 0.625$. Since the slope is less than the critical slope, condition 5 is not satisified, and thus Test 2 is more cost-effective. By equation 5 the minimum dollars per life-year gained is $d = \$243,000$ for test 1 to be more cost-effective than test 2.

[3] also provide a simple modification to National Comprehensive Cancer Network (NCCN) guidelines so that 90% of mutation carriers are identified by testing 68% of women. These simple guidelines are statistically inferior to the BRCAPRO threshold at which 90% of carriers are identified by testing only 60% of women. However, since these simple guidelines do not require a BRCAPRO risk calculation, perhaps the screening test is $50 cheaper ($c_{m1} = \$50$ vs. $c_{m2} = \$100$). The condition 6 is not satisfied because the sensitivity gain is 0, but the right side is 0.12. Thus screening with the simple modification to NCCN guidelines is not cheap enough to be cost-effective versus screening with BRCAPRO. Even if the NCCN modification were costless, by equation 6, it would still not be more cost-effective than screening with BRCAPRO.

5 Sensitivity Analyses

5.1 Changing the effectiveness affects dollars per life-year gained, but not optimal risk thresholds

In the BRCAPRO example, we presumed screening woman at age 50. If we instead screen women at age 40 rather than 50, the life-years gained $(e_1 - e_2)$ increases to 6.5 from 5.0. However, the plot of INB or NB versus risk threshold, such as figure 2 (left panel), is the same for screening at ages 40 or 50. This is because, by equation 4, fixing both costs and risk threshold also fixes $d(e_1 - e_2)$. Thus if $(e_1 - e_2)$ increases, then dollars per life-year gained d must decrease so that $d(e_1 - e_2)$ remains fixed. The risk-thresholds for which BRCAPRO screening is viable in the general population remains $(0.13\%, 0.91\%)$. However, because life-years gained increased, thus dollars per life-year gained threshold implied by each threshold now decreases to ($12,000, $240,000). Thus changing the effectiveness does not change the optimal risk thresholds, but does alter their meaning in terms of dollars per life-year gained.

5.2 Cost of the definitive test

5.3 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).

IONUT: It would be important to see sensitivity wrt c_0: When is the definitive test so expensive we would never screen, and when is it so cheap that we would just test everyone? The mid-range is where screening could be supported at some risk threshold.

IONUT: It would be important to see sensitivity wrt c_m: When would the screening test get too expensive and we screen no one (meaning either we screen no one or we definitively test everyone (i.e. marker cost doesn't matter))? I think this is just $c_m + max(INB)$, that you have raise the bar over the current peak INB.

5.4   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.

6   Discussion

The simple conditions for one test having more INB than another are useful when there isn't a risk calculator but you do have the concentration curve.

Value of using INB vs NB: NB has abstract scale, INB easily shows sensitivity to screening test cost, specifying costs/effectiveness allows us to understand the meaning of an optimal risk threshold in terms of dollars per LYG. Note that the meaning of an optimal risk threshold differs if we modify the LYG or any cost, but the optimal risk threshold stays the same. Using NB assuming a costless test could be awful - may well be sensitive to small costs. We need to know what a risk threshold means in terms of dollars per LYG - "optimal" risk thresholds at which screening is best might represent unrealistically low or high dollars per LYG.

Talk Prob Sens Analysis (PSA) as important, can be done for this simple approach as well.

Complementing microsim and decision trees. SImple analysis identifies key quantities that can be the focus of the more comprehensive analysis to either better identify or to do the PSA. Comparing the Aapproaches informs each other - if agree with simple approach, then complexities don't matter. If don't agree, either the simple approach is missing something critical or the complex approach may be speculatively modeling something incorrectly. Resolving the disagreement should prove insightful and advance the debate on what is cost-effective, and spur reseach into any quantities identified as being critical but poorly known.

QALYs would be preferable, easy extension of what we've done here.

SHould do sens analy on BRCA genetic test cost - when is this so cheap, we should just do it for everyone? This is the future of mutation screening, so this predicts when it will happen.

A   Appendix

A.1   Screening is never cost-effective if net effectiveness $e < 0$

To see that $INB < 0$ if $e < 0$, note INB is a function of $PPV = P(D+|M+)$:

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

A.2   Derivation of Net Benefit (NB) as INB standardized by benefit

Substituting $P(M+) = P(D+, M+) + P(D-, M+)$, an alternate expression for INB is

$$INB = [e + c_0]P(D+, M+) - c_m - c_0 P(M+)$$
$$= eP(D+, M+) - c_0 P(D-, M+) - c_m.$$

The INB standardized by benefit $B = e$ is

$$
\begin{aligned}
\frac{INB}{B} &= P(D+, M+) - \frac{R}{1-R}P(D-, M+) - \frac{c_m}{e} \\
&= \pi \cdot Sens - \frac{R}{1-R}(1 - Spec)(1 - \pi) - \frac{c_m}{e} = NB(R, c_m, e),
\end{aligned}
$$

where $c_0/e = R/(1 - R)$, $\pi = P(D+)$, $Sens = P(M+|D+)$, and $Spec = P(M-|D-)$. The last equation is the NB[34]. Note that for a costless screening test $c_m = 0$, the net effectiveness of early intervention e does not need to be specified. In this case, NB is a function of only risk threshold R, which is how NB is typically used.

References

[1] Antoniou, A., P. P. D. Pharoah, P. Smith, and D. Easton (2004, Oct). The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer 91*(8), 1580–90.

[2] Baker, S. G., N. R. Cook, A. Vickers, and B. S. Kramer (2009, October). Using relative utility curves to evaluate risk prediction. *Journal of the Royal Statistical Society. Series A, (Statistics in Society) 172*, 729–748.

[3] Best, A. F., M. A. Tucker, M. N. Frone, M. H. Greene, J. A. Peters, and H. A. Katki (2017). To test or not to test: Selection criteria for population-based BRCA1/2 mutation screening. *Submitted*.

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

[5] Cantor, S. B. and M. W. Kattan (2000). Determining the area under the ROC curve for a binary diagnostic test. *Med Decis Making 20*(4), 468–470.

[6] Cook, N. R. (2007, Feb). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation 115*(7), 928–935.

[7] Copas, J. (1999). The effectiveness of risk scores: The logit rank plot. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 48*(2), 165–183.

[8] Gail, M. H. and R. M. Pfeiffer (2005, April). On criteria for evaluating models of absolute risk. *Biostatistics (Oxford, England) 6*, 227–239.

[9] Greenhouse, S. W., J. Cornfield, and F. Homburger (1950, Nov). The Youden index: letters to the editor. *Cancer 3*(6), 1097–1101.

[10] Hanley, J. A. and B. J. McNeil (1982, April). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology 143*, 29–36.

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

[12] Huang, Y., M. Sullivan Pepe, and Z. Feng (2007). Evaluating the predictiveness of a continuous marker. *Biometrics 63*(4), 1181–1188.

[13] Katki, H. A. (2006, June). Effect of misreported family history on Mendelian mutation prediction models. *Biometrics 62*(2), 478–487.

[14] King, M.-C., E. Levy-Lahad, and A. Lahad (2014, September). Population-based screening for BRCA1 and BRCA2: 2014 Lasker Award. *JAMA 312*, 1091–1092.

[15] Kraemer, H. C. (1992). *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, CA: Sage Publications Inc.

[16] Kraemer, H. C. (2004, Jan). Reconsidering the odds ratio as a measure of 2x2 association in a population. *Stat Med 23*(2), 257–270.

[17] Kuchenbaecker, K. B., J. L. Hopper, D. R. Barnes, et al. (2017, June). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA 317*, 2402–2416.

[18] Lachin, J. M. (2000). *Biostatistical Methods: The Assessment of Relative Risks.* New York: Wiley-Interscience.

[19] Long, E. F. and Ganz P. A. (2015, Dec) Cost-effectiveness of Universal BRCA1/2 Screening: Evidence-Based Decision Making. *JAMA Oncol 1*, 1217-8.

[20] Mai, P. L., Chatterjee N, Hartge P, Tucker, M , Brody L , Struewing JP and Wacholder S (2009, July) Potential excess mortality in BRCA1/2 mutation carriers beyond breast, ovarian, prostate, and pancreatic cancers, and melanoma. *PloS one 4*, e4812.

[1] Manchanda, R., Legood R, Burnell M, McGuire A , et. al. (2015, July) Cost-effectiveness of population screening for BRCA mutations in Ashkenazi jewish women compared with family history-based testing. *J Natl Cancer Inst 107(1)*, 380.

[22] Moyer, V. A. (2014, February). Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: U.S. Preventive Services Task Force recommendation statement. *Ann Int Med 160*, 271–281.

[23] NICE (2017). Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer, recommendation 1.5.11. Technical report, National Institute for Health and Care Excellence Clinical Guidance https://www.nice.org.uk/guidance/cg164/chapter/Recommendations#genetic-testing.

[24] Parmigiani, G., D. Berry, and O. Aguilar (1998, Jan). Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet 62*(1), 145–158.

[25] Pauker, S. G. and J. P. Kassirer (1980, May). The threshold approach to clinical decision making. *N Engl J Med 302*(20), 1109–1117.

[26] Pencina, M. J., R. B. D'Agostino, R. B. D'Agostino, and R. S. Vasan (2008, Jan). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med 27*(2), 157–72; discussion 207–12.

[27] Pepe, M. S., H. Janes, G. Longton, W. Leisenring, and P. Newcomb (2004, May). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol 159*(9), 882–890.

[28] Rizk, C. (2017). Researchers debate merits of population-wide genetic testing at AACR. Technical report, GenomeWeb https://www.genomeweb.com/genetic-research/researchers-debate-merits-population-wide-genetic-testing-aacr.

[29] Staff Reporter (2017). Veritas Genetics to provide BRCA testing for Canadian hereditary cancer screening effort. Technical report, GenomeWeb https://www.genomeweb.com/clinical-sequencing/veritas-genetics-provide-brca-testing-canadian-hereditary-cancer-screening.

[30] Struewing, J. P., P. Hartge, S. Wacholder, S. M. Baker, M. Berlin, M. McAdams, M. M. Timmerman, L. C. Brody, and M. A. Tucker (1997). The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N. Engl. J. Med. 336*, 1401–1408.

[31] Sweeting, M. J. and S. G. Thompson (2012, Apr). Making predictions from complex longitudinal data, with application to planning monitoring intervals in a national screening programme. *J R Stat Soc Ser A Stat Soc 175*(2), 569–586.

[32] The Screen Project . http://thescreenproject.ca; 2017. Accessed January 29, 2018.

[33] Rabin RC. F.D.A. Approves First Home Testing for 3 Breast Cancer Mutations, With Caveats. *New York Times.* March 6, 2018.

[34] Vickers, A. J. and E. B. Elkin (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making 26*, 565–574.

[35] Wentzensen, N. and S. Wacholder (2013, Feb). From differences in means between cases and controls to risk stratification: a business plan for biomarker development. *Cancer Discov 3*(2), 148–157.

[36] Youden, W. J. (1950, Jan). Index for rating diagnostic tests. *Cancer 3*(1), 32–35.

<div align="center">

WebAppendix for the Paper:

"Simple and optimal cost-effective risk thresholds for a single screen"

by Hormuzd A. Katki and Ionut Bebu

</div>

A   Direct deriviation of the optimal cost-effective risk threshold

Let M be an unbiased risk-model:

$$P(D+|M=m)=m\,.$$

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

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

INB can be writen as

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

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:

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

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

B   Intervention costs for women with *BRCA1/2* mutations

The supplement of[19] reports that the cost for risk-reducing mastectomy (RRM) was \$12286 and for risk-reducing salpingo-oophorectomy (RRSO)was \$7393. The breast cancer treatment cost is the sum of the 1st and last year costs, plus 8 years of survival in between (i.e. assuming an average of 10 years of survival with breast cancer), which is \$86013 + 8 × 7547 + 63790 = \$210179. The ovarian cancer treatment cost is the sum of the 1st and last year costs, plus 3 years of survival in between (i.e. assuming an average of 5 years of survival with ovarian cancer), which is \$124838 + 3 × 13724 + 87218 = \$253228. For c_1, because women with RRM have a 2.7% chance of breast cancer, and women with RRSO have 1.2% chance of ovarian cancers, then $c_1 = \$12286 + 7393 + 0.027 \times 210179 + 0.012 \times 253228 = \$28393$. Because 17.1% of women with

<div align="center">i</div>

BRCA1/2 mutations will not develop breast or ovarian cancer (53% will develop breast cancer and 29.9% will develop ovarian cancer), $c_2 = \$0.53 \times 210179 + 0.299 \times 253228 + 0.171 \times 0 = \$187110$.

We do not need to include the cost of lifetime breast cancer screening because it cancels out in $c_1 - c_2$, i.e. women with RRM still undergo breast cancer screening because RRM does not totally eliminate breast cancer risk. Currently, there is no recommended method for ovarian cancer screening. We assume the chance of developing both breast and ovarian cancer is negligible.

C   Equivalent specificity condition for a test to have greater INB

The inequality condition for a test having greater INB was written in terms of specificity (equation 5). It can be equivalently written in terms of specificity: 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),

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

The inequalities based on sensitivity or specificity are equivalent, either can be used.

References

[1] Manchanda, R., Legood R, Burnell M, McGuire A , et. al. (2015, July) Cost-effectiveness of population screening for BRCA mutations in Ashkenazi jewish women compared with family history-based testing. *J Natl Cancer Inst 107(1)*, 380.