Received 26 April 2016; | Revised 6 June 2016; | Accepted 6 June 2016

DOI: xxx/xxxx

RESEARCH ARTICLE

A simple framework to identify optimal risk thresholds for a single screen

Hormuzd A. Katki*[1] | Ionut Bebu[2]

[1]US National Cancer Institute, Division of Cancer Epidemiology and Genetics, MD, USA

[2]Biostatistics Center, George Washington University, MD, USA

Correspondence

*Hormuzd A. Katki, US National Cancer Institute, NIH, DHHS, Division of Cancer Epidemiology and Genetics, 9609 Medical Center Dr. Room 7E592 Rockville MD 20850. Email: katkih@mail.nih.gov

Present Address

US National Cancer Institute, NIH, DHHS, Division of Cancer Epidemiology and Genetics, 9609 Medical Center Dr. Room 7E592 Rockville MD 20850

Summary

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, most applications assume a costless screening biomarker for simplicity, and DCA cannot formally identify optimal risk thresholds. Although a full decision analysis can identify optimal risk thresholds, typically used methods are complex and it can be difficult to understand why they arrive at their conclusion. We develop a simple framework to 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 simple expressions 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 the risk threshold for using the BRCAPRO risk model to screen women for mutations in *BRCA1/2* that cause high risks of breast and ovarian cancers. Our simple framework readily identifies sensitivity of DCA to biomarker test costs that is hard to see by DCA alone. Importantly, most, and sometimes even all, of the optimal risk thresholds identified by DCA were infeasible based on their associated dollars per life-year gained threshold. Our approach to estimating optimal risk thresholds in simple and transparent manner, providing intuition about which quantities are critical, may serve as a bridge between DCA and a full decision analysis.

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 [1,2,3]. The cost-effective threshold requires specifying the costs, benefits,

and harms of the tests and subsequent interventions. However, statistical metrics that do not require costs, benefits, and harms, such as Youden's index, AUC, and Mean Risk Stratification[4], cannot identify cost-effective thresholds.

The most important recent advance is Decision Curve Analysis (DCA)[5]. DCA takes an important step towards formal decision analysis, but stops short of requiring specification of costs and effectiveness. DCA is often used to identify a range of optimal risk-thresholds for using a risk-model to decide who should get intervention. However, there are two limitations of DCA. First, most researchers in practice use DCA by assuming costless risk models. However, most risk models are not costless, typically requiring shared decision-making with a health professional. DCA can acount for screening test costs, but this requires specification of costs and effectiveness and thus negates the simplicity of DCA. More importantly, because DCA does not require costs and effectiveness, DCA considers risk-thresholds only as a sensitivity analysis and thus cannot formally identify optimal thresholds.

In contrast, a full decision analysis, specifying costs, benefits, and the monetary value of life, can determine cost-effective thresholds. Often these analyses require complex decision trees and microsimulation. It can be hard to glean intuition about which quantities and modeling assumptions are critical, even when probabilistic sensitivity analysis is conducted. We suggest that there is a role for methodology for estimating optimal risk thresholds in simple and transparent manner, providing intuition about which quantities are critical, as a bridge between DCA and a full decision analysis.

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 calculate the Incremental Net Benefit (INB) based on specifying 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 risk-threshold for the screening test that depends on only the cost of the definitive test and a parameter we call the net effectiveness of early intervention (NE). Because the dollar-value of life-years gained is often hard to specify, we also provide an expression for the dollar-value of life-years gained implicitly implied by an optimal risk-threshold.

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[6]. Current US and UK guidelines refer women for mutation testing only if they have a strong family history of breast or ovarian cancer[7], as quantified by a risk model calculating that their risk of carrying a *BRCA1/2* mutation exceeds 10%[8]. However, as mutation-testing costs plummet, prominent voices have called for testing *all* women[9], equivalent to a 0% risk-threshold. The Screen Project offers *BRCA1/2* testing to any Canadian[10]. In a landmark ruling, the US FDA authorized direct-to-consumer genetic testing for *BRCA1/2* mutations among Ashkenazi-Jews[11].

We recently showed that 80% of Ashkenazi-Jewish mutation-carriers could be identified by testing only 44% of Ashkenazi-Jewish women at highest risk for carrying mutations[12], as identified by the BRCAPRO risk model[13]. 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, DCA, and Mean Risk Stratification[4], 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. In this example, we show that DCA is sensitive to the standard assumption of a costless screening test, but this is hard to see by DCA. Our framework easily identifies this sensitivity. More importantly, many, or even sometimes all, the risk thresholds identified by DCA as being favorable for screening implied unfeasible values of dollars per life-year gained. Our results are comparable to those from more comprehensive microsimulation-based decision analyses, while retaining the benefits of simplicity, transparency, and ease of identifying the critical quantities that the decision analysis is sensitive to.

2 | A COST-EFFECTIVENESS FRAMEWORK FOR A SINGLE-TIME SCREEN

The goal of a single-time screen is to identify all cases of prevalent disease for early detection and treatment. Although disease could be absent at the time of screen and develop later, that requires longitudinal screening to detect, which is beyond the scope of this paper. Thus the goal of a single-screen is solely to mitigate the impact of disease that is currently present by detecting it early for immediate treatment, rather than detecting it later when treatment may be less effective. This framework includes both detecting disease early when it is curable (i.e. early detection) and detecting disease in a precursor state where its removal prevents future potentially incurable disease (i.e. secondary prevention).

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. The first outcome is a positive test, after which a 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.

The second outcome is that the test is negative, after which there is no further action. However, future symptoms will prompt definitive testing to diagnose disease that was present but missed by the screen. If disease were diagnosed in the future based on symptoms, then presumably treatment is less effective (or most costly) than it would have been had disease been found earlier. Four parameters define the average costs for a single screen:

First is c_m, the cost of screening marker evaluation. Note that 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.

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

Third is c_1, 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.

Fourth is c_2, the treatment cost for a disease 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 life-expectancy:

First is e_0, the life-expectancy for someone who never develops disease. This is assumed to be the same for those who correctly tested negative or falsely tested positive for the single screening test.

Second is e_1, the life-expectancy for someone who screens positive and the definitive test is also positive for disease. This is early-detection and is the benefit of screening.

Third is e_2, 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.

2.2 | Calculating Net Benefit (NB)

Define disease as D and the marker for the screening test as M. The marker is dichotomized as positive ($M+$) or negative ($M-$) based on a underlying threshold m_0 (our goal will be to identify the cost-effective m_0). The components of the average cost and effectiveness are:

1. True Positive: The screening is positive and the individual has disease ($M+$, $D+$).
   Cost is $c_m + c_0 + c_1$ and effectiveness is e_1.
2. False Positive: The screening is positive but the individual does not have disease ($M+$, $D-$).
   Cost is $c_m + c_0$ and effectiveness is e_0.
3. False Negative: The screening is negative but the individual has disease ($M-$, $D+$).
   Cost is $c_m + c_0 + c_2$ and effectiveness is e_2.
4. True Negative: The screening is negative and the individual does not have disease ($M-$, $D-$).
   Cost is c_m and effectiveness is e_0.

Let C_M and E_M be the average cost and effectiveness for screening, respectively, so that:

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

where $\mathcal{E}(\mathcal{X})$ represents the expected value of random variable X. The Net Benefit (NB) is the difference of average effectiveness and cost for screening:

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

where λ is the "willingness to pay parameter" that measures the dollar-value of 1 year of life.

2.3 | Optimal cost-effective risk threshold that maximizes the Net Benefit of screening

Given a screening marker or risk model M, a screening test is declared either $M+ = \{M \geq m_0\}$ as positive or $M- = \{M < m_0\}$ as negative. Assume that M is unbiased, meaning $P(D+|M = m) = m$. We derive the optimal threshold m_0 that maximizes the Net Benefit of screening. Let f_M denote the density of M, so that

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

$$P(D-, M+) = \int_{m_0}^{\infty} P(D-|M = m) f_M(m)\, dm = \int_{m_0}^{\infty} (1-m) f_M(m)\, dm$$

$$P(D+, M-) = P(D+) - P(D+, M+) = P(D+) - \int_{m_0}^{\infty} m f_M(m)\, dm.$$

Plugging the above into the Net Benefit for screening NB_M yields

$$NB_M = \lambda e_0 P(D-) - c_m - c_0 \int_{m_0}^{\infty} (1-m) f_M(m)\, dm + (\lambda e_1 - c_0 - c_1) \int_{m_0}^{\infty} m f_M(m)\, dm + (\lambda e_2 - c_0 - c_2) \left(P(D+) - \int_{m_0}^{\infty} m f_M(m)\, dm \right).$$

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

$$c_0(1-m_0) f_M(m_0) - (\lambda e_1 - c_0 - c_1) m_0 f_M(m_0) + (\lambda e_2 - c_0 - c_2) m_0 f_M(m_0) = 0.$$

Solving for m_0 yields the optimal cost-effective risk threshold, denoted $R(\lambda)$, that maximizes the Net Benefit of screening:

$$m_0 = \frac{c_0}{e(\lambda) + c_0} \equiv R(\lambda), \tag{2}$$

where $e(\lambda) = \lambda(e_1 - e_2) - (c_1 - c_2)$ is the "net effectiveness of early intervention". The net effectiveness $e(\lambda)$ contrasts the dollar-value of the life-expectancy gain with dollar increase in treatment cost for early vs. late detection. If $e(\lambda) < 0$, then optimal threshold $R(\lambda)$ does not exist. We emphasize that $e(\lambda)$, and hence optimal threshold $R(\lambda)$, is a function of the willingness-to-pay parameter λ, which varies between people and populations. We will vary λ in our examples.

The optimal threshold $R(\lambda)$ requires specification of only the definitive test cost c_0 and the net effectiveness of early intervention $e(\lambda)$. Life-expectancy in the absence of disease (e_0) is not needed. Also the screening test cost c_m is not needed, because risk thresholds only matter after the screening test is conducted, and thus the screening test cost is a sunk cost. However, e_0 and screening test cost will matter to check whether optimal screening is cost-effective (Section 2.4).

Defining $M_R+ = \{M \geq R(\lambda)\}$ and $M_R- = \{M < R(\lambda)\}$, we denote the Net Benefit of screening with the optimal R as

$$NB_M(R) = \lambda e_0 P(D-) - c_m - c_0 P(D-, M_R+) + (\lambda e_1 - c_0 - c_1) P(D+, M_R+) + (\lambda e_2 - c_0 - c_2) P(D+, M_R-). \tag{3}$$

The notation M_R+ and M_R- clarifies that marker/model must be dichotomized at the optimal threshold $R(\lambda)$.

The WebAppendix derives the optimal threshold R from the classical expected utility approach[1].

2.3.1 | Willingness-to-pay λ implied by optimal risk-thresholds

Because the willingness-to-pay λ for 1-year of life varies between people and populations, is difficult to specify the optimal $R(\lambda)$. Thus some analyses, such as Decision Curve Analysis[5], vary $R(\lambda)$ as a sensitivity analysis, without regard to λ. However, there is a one-to-one relationship of $R(\lambda)$and λ (equation 2), and thus a given $R(\lambda)$ may not be feasible in practice its corresponding λ is not realisitic. For example, if a particular optimal $R(\lambda)$ required valuing 1 year of life at $\lambda = \$1$ million USD, then that optimal threshold is not feasible in practice because typically $\lambda \leq \$100,000$ USD. Solving for λ in equation 2 yields

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

An optimal risk threshold $R(\lambda)$, fixing costs and life-years gained, implies a unique willingness-to-pay $\lambda(R)$. In analyses where the optimal risk threshold is varied, as in our example or Decision Curve Analysis[5], one needs to also calculate the $\lambda(R)$ underlying the optimal risk threshold $R(\lambda)$ to see if $R(\lambda)$ is a feasible threshold.

2.4 | Incremental Net Benefit versus screening no one or testing everyone

In addition to screening with the optimal threshold R, two other policies are possible: test no one, or do definitive testing on everyone. To ensure that Net Benefit from screening with the optimal threshold R (equation 3) is the best policy, we must compare to the Net Benefits of the other two policies.

For testing no one, if the person has disease ($D+$), then cost is c_0+c_2 and effectiveness is e_2. If the person does not have disease ($D-$), then cost is 0 and effectiveness is e_0. The Net Benefit of testing no one is $NB_{none} = \lambda e_0 P(D-)+(\lambda e_2-c_0-c_2)P(D+)$. For testing everyone, if the person has disease ($D+$), then the cost is c_0+c_1 and effectiveness is e_1. If the person does not have disease ($D-$), then cost is c_0, effectiveness is e_0. The Net Benefit of testing everyone is $NB_{all} = (\lambda e_0-c_0)P(D-)+(\lambda e_1-c_0-c_1)P(D+)$.

To compare the 3 Net Benefits, we calculate the Incremental Net Benefit (INB) of screening vs. each of the other 2 policies:

$$INB_{none}(R) = NB_M(R) - NB_{none} \text{ and } INB_{all}(R) = NB_M(R) - NB_{all}.$$

If both $INB_{none}(R) > 0$ and $INB_{all}(R) > 0$, then screening at optimal threshold R is best. To summarize both conditions into a single condition, we define

$$INB_{Gain}(R) = min(INB_{none}(R), INB_{all}(R)), \tag{5}$$

and require that $INB_{Gain}(R) > 0$ for screening at threshold R to be the best policy. Note that the threshold R that maximizes $NB_M(R)$ (equation 2) also maximizes $INB_{Gain}(R)$ because neither NB_{none} nor NB_{all} are functions of R. Although optimal threshold R is not a function of life-expectancy e_0 or screening test cost c_m, note that $NB_M(R)$ requires c_m, and both NB_{none} and NB_{all} require e_0, so that both c_m and e_0 matter to check if optimal screening is superior to testing no one or testing everyone.

In the examples, we will use $INB_{Gain}(R)$ to identify the range of optimal thresholds R such that screening is the best policy. Note that by varying optimal threshold $R(\lambda)$, but fixing costs and effectiveness, we are implicitly varying λ. We will also plot $INB_{Gain}(R)$ versus $\lambda(R)$ to understand if an optimal threshold implies a feasible willingness-to-pay $\lambda(R)$.

2.4.1 | Comparison of INB to Decision Curve Analysis

Decision Curve Analysis (DCA) calculates a "Net Benefit (NB)" to identify risk thresholds where using a screening marker to decide who should get treated is better than treating no one or testing everyone[5]. We now show that the NB from DCA equals the INB_{none} divided by net effectiveness $e(\lambda)$. Note that $INB_{none}(R) = NB_M(R) - NB_{none}$ is

$$\begin{aligned} INB_{none}(R) &= \{\lambda e_0 P(D-) - c_m - c_0 P(D-, M_R+) + (\lambda e_1 - c_0 - c_1)P(D+, M_R+) + (\lambda e_2 - c_0 - c_2)P(D+, M_R-)\} \\ &\quad -\{\lambda e_0 P(D-) + (\lambda e_2 - c_0 - c_2)P(D+)\} \\ &= e(\lambda)P(D+, M_R+) - c_0 P(D-, M_R+) - c_m. \end{aligned}$$

The $INB_{none}(R)$ divided by net effectiveness $e(\lambda)$ equals the Net Benefit from DCA, denoted $NB_{DCA}(R)$ as a function of R:

$$NB_{DCA}(R) = \pi \cdot Sens - \frac{R}{1-R}(1-Spec)(1-\pi) - \frac{c_m}{e(\lambda)},$$

where $c_0/e(\lambda) = R/(1-R)$, $\pi = P(D+)$, $Sens = P(M_R+|D+)$, and $Spec = P(M_R-|D-)$.[5] Note that the marker/model must be dichotomized at the optimal threshold R.

DCA also must compare the NB_{DCA} for screening versus the NB_{DCA} of the two policies of testing no one or testing everyone. To summarize both comparisons into a single condition akin to $INB_{Gain}(R)$ of equation **??**, we define $NB_{Gain}(R)$ as increase in $NB_{DCA}(R)$ versus the NB_{DCA} of the other 2 policies[4]. The NB_{DCA} for testing no one is $(NB_{none} - NB_{none})/e(\lambda) = 0$. The NB_{DCA} for testing everyone is $(NB_{all} - NB_{none})/e(\lambda)$. Thus we define

$$NB_{Gain}(R) = NB_{DCA}(R) - max(0, (NB_{all} - NB_{none})/e(\lambda)).$$

For a costless screening test $c_m = 0$, NB_{DCA} does not require specifying the net effectiveness $e(\lambda)$. In this case, NB_{DCA} is a function of only risk threshold R. This is a major advantage of NB_{DCA} in practice, allowing one to plot NB_{DCA} as a function of R to identify the range of risk thresholds supporting screening, without having to specify any cost or effectiveness parameters. The thresholds R where $NB_{Gain}(R) > 0$ are thresholds where the risk model is better than screening no one or testing everyone. In particular, $NB_{Gain}(R)$ is maximized at optimal risk threshold equals disease prevalence, $R = P(D+)$.[4] If the risk threshold is below prevalence, then if $NB_{Gain}(R) < 0$, then testing everyone is the favored policy. If the risk threshold is above prevalence, then if $NB_{Gain}(R) < 0$, then testing no one is the favored policy.

These are the same thresholds as where $INB_{Gain}(R) > 0$. However, using INB_{Gain} has 2 key advantages over the $NB_{Gain}(R)$ of DCA. The first advantage of $INB_{Gain}(R) > 0$ is that one can examine if an optimal risk threshold $R(\lambda)$ is feasible by calculating the λ underlying it (equation 4). Second, screening tests are not costless. Even most risk models are not costless, requiring at least a patient's time to fill out inputs and to comprehend outputs, and usually requires a visit with a health professional. Although it is straightforward to specify c_m, note that this then requires also specifying net effectiveness $e(\lambda)$ which requires costs and effectiveness. Thus accounting for screening test cost in DCA also requires all the other costs and effectiveness, negating the key advantage of DCA as not requiring 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, 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%[8]. We focus on the BRCAPRO[13] risk model.

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[12]. 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)[14]. 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[13]. Here M is the BRCAPRO risk score, and because BRCAPRO is a well-calibrated risk model[12], $m_0 = R$, 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[12] (see WebAppendix).

3.2 | Costs and effectiveness parameters 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[15]. We use costs used by the Long and Ganz decision analysis from 2015[16] to compare results. 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 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. We use $c_1 = \$28,393$ and average cancer treatment cost is $c_2 = \$187,110$ (see WebAppendix).

3.3 | *BRCA1/2* mutation screening for Ashkenazi Jews

Figure 1 (left panel) shows that the Net Benefit of using BRCAPRO increases as the risk threshold decreases. For a willingness to pay of $\lambda = \$51,207$ per life-year gained, the optimal risk threshold is $c_0/(e(\lambda)+c_0) = 249/(414751+249) = 0.06\%$. However, testing all has the highest Net Benefit, and thus would be the preferred policy for $\lambda = \$51,207$. The only part of the curve of Figure 1 (left panel) that matters is the Net Benefit at the optimal risk threshold, because other thresholds are suboptimal for $\lambda = \$51,207$. Thus, for each λ, we calculate its associated optimal threshold (equation 4) and the Net Benefit only at the optimal threshold. We compare the optimal Net Benefit for screening at each optimal threshold versus the Net Benefits for testing no one or testing everyone (Figure 2). A key point is that as threshold R decreases, its associated willingness to pay λ increases. Although each threshold in the table is optimal for a given willingness to pay, many of these λ will be infeasible for anyone. For example, at the 0.02% threshold, the $\lambda = \$217,206$ is outside the typical range of willingness to play ($10,000 to $100,000), and is probably not feasible for anyone. Furthermore, note that thresholds above 0.16% imply *negative* willingness to pay, meaning

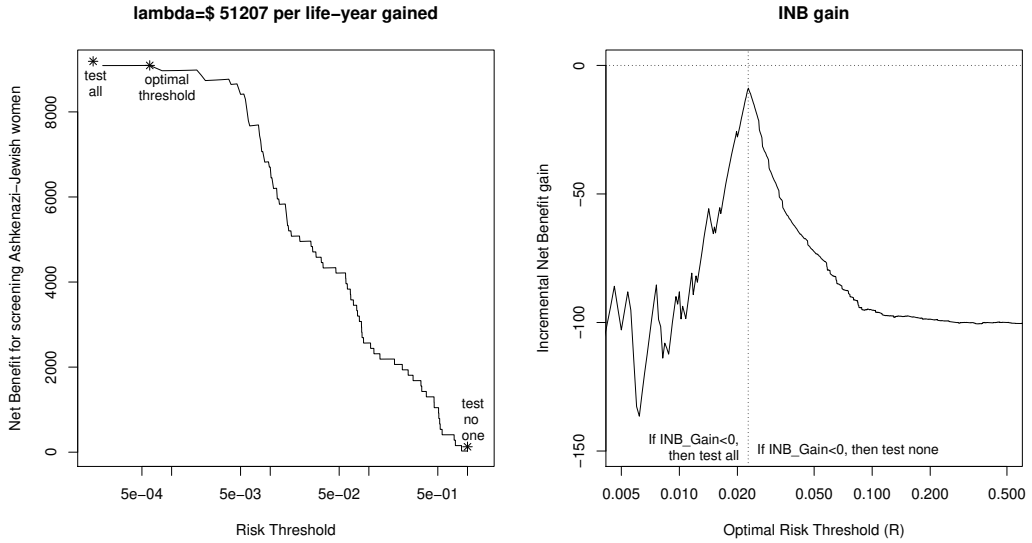

FIGURE 1 Left panel is Net Benefit for screening vs. risk threshold for using the BRCAPRO risk model to screen for which Ashkenazi-Jewish women should get *BRCA1/2* genetic testing. The optimal risk threshold is marked, as are the Net Benefits for testing no one or testing everyone. The right panel plots $INB_{Gain}(R)$ vs. optimal risk threshold. The vertical line is where the optimal risk threshold equals disease prevalence. To the left of the vertical line, if the INB gain is negative, then the testing everyone is the preferred policy. To the right of the dotted line, if the INB gain is negative, then the testing no one is the preferred policy. For optimal risk thresholds where INB gain is positive, screening is the preferred policy.

Threshold	lambda	NB_M(R)	NB_none	NB_all	INB_gain(R)
0.0002	217206.7	17365663	17337699	17365763	-99.54
0.0004	92706.7	7411327	7397518	7411425	-98.55
0.0006	51206.7	4093217	4084124	4093313	-96.10
0.0008	30456.7	2434068	2427427	2434257	-188.92
0.0010	18006.7	1438656	1433409	1438823	-166.80
0.0012	9706.7	775049	770730	775200	-151.37
0.0014	3778.1	301046	297389	301184	-138.67
0.0016	-668.3	-54456	-57618	-54328	-128.54
0.0018	-4126.6	-330957	-333734	-330837	-119.40

FIGURE 2 Optimal thresholds, their associated willingness to pay λ, and the Net Benefits of screening $NB_M(R)$, testing no one NB_{none}, and testing everyone NB_{all}, and the INB gain of screening versus the better of testing no one or testing everyone $INB_{Gain}(R)$.

that life-gained is devalued. These thresholds are immoral and cannot be considered. Thus the only optimal thresholds that can be considered are 0.04% to 0.12%.

The Net Benefit for screening at the optimal threshold $NB_M(R)$ increases as threshold R decreases. However, the INB_{Gain} is negative, meaning that either testing no one or testing everyone is the best policy versus each optimal threshold. Specifically, the Net Benefit for testing everyone NB_{all} is highest at each optimal threshold and is thus the best policy. Figure 1 (right panel) plots the $INB_{Gain}(R)$. See Figure 1 of the WebAppendix to see INB_{none} and INB_{all} separately

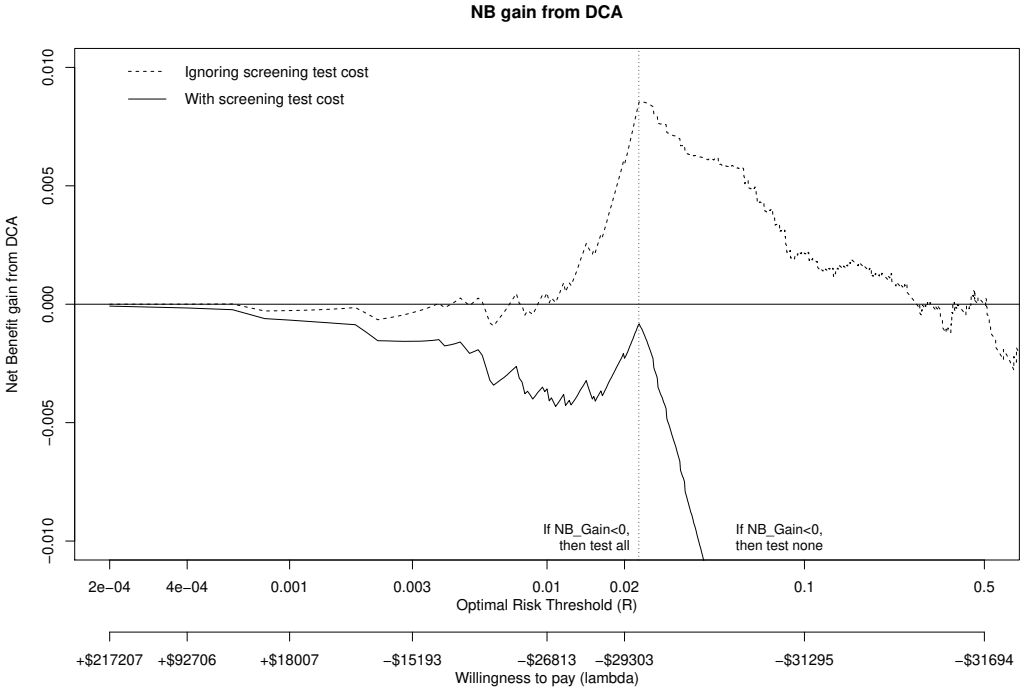

FIGURE 3 NB gain from Decision Curve Analysis (DCA) (both ignoring and with screening test cost) vs. BRCAPRO risk-threshold to screen for which Ashkenazi-Jews should get *BRCA1/2* genetic testing. The second x-axis plots the associated willingness to pay λ for each optimal risk threshold R. The vertical line is where the optimal risk threshold equals disease prevalence. To left of this line, if NB_{Gain} is negative then testing everyone is the favored policy; to the right, testing no one is favored.

The $INB_{Gain}(R)$ always peaks at $R = P(D+)$ disease prevalence, and thresholds where $INB_{Gain}(R) > 0$ favor screening as the policy. Since $INB_{Gain}(R)$ is negative at all optimal risk thresholds, then either testing no one or testing all are the preferred policies, depending on what the feasible optimal risk thresholds are. Since only thresholds from 0.04% to 0.12% are feasible, then the preferred policy is testing everyone.

3.4 | Effect of ignoring screening test cost

Figure 3 shows that the NB gain from Decision Curve Analysis (DCA), with BRCAPRO screening test cost $c_m = \$100$ (middle panel), is always negative. Thus screening is never the best choice. However, the NB gain estimated by usual costless test assumption 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 (Figure 1 , right panel) because its units are dollars. The INB asymptotes at -$100, clarifying that a $100 cheaper screening test would permit screening at a wide range of risk-thresholds.

Figure 3 (right panel) plots NB (for a costless BRCAPRO test) versus the threshold for dollars per life-year gained, using equation 4 to convert from optimal risk threshold to dollars per life-year gained. 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 doing genetic testing on all Ashkenazi-Jewish women is not merely cost-effective, but is actually *cost-saving*, versus screening with BRCAPRO at optimal thresholds. The reason for the negative dollars per life-year gained is that the difference in treatment costs for early- vs. late-detection $c_1 - c_2 = -\$158717$ is strongly negative. Thus our framework clarifies that the reason why testing all Ashkenazi Jews saves money versus screening with BRCAPRO is because treating cancer is far more expensive than treatment costs incurred by genetic testing.

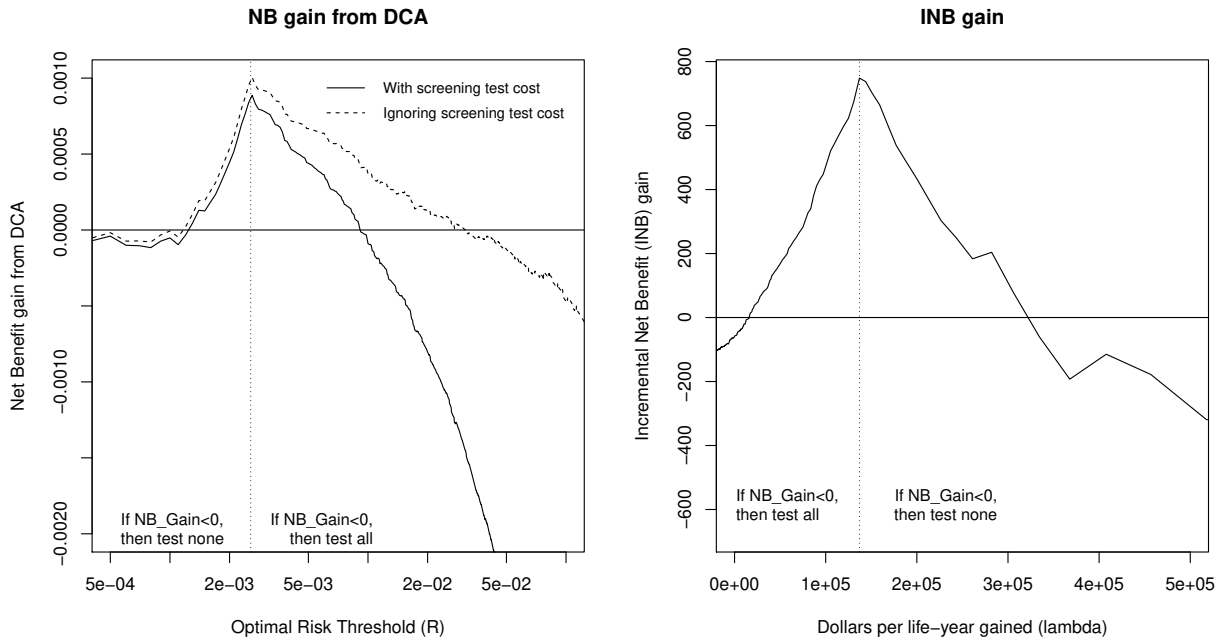

FIGURE 4 NB gain from Decision Curve Analysis (DCA) vs. risk-threshold (left panel) and INB gain vs. dollars per life-year gained (right panel) for using the BRCAPRO risk model to screen for which members of the general population should get *BRCA1/2* genetic testing.

This example demonstrates that optimal risk thresholds can imply infeasible thresholds of dollars per life-year gained. Note that dollars per life-year gained thresholds cannot be calculated by DCA and requires specifying costs and effectiveness.

Our conclusion agrees with that of microsimulation-based decision analyses that more comprehensively considered costs and effectiveness, but only examined screening at a single implicit risk-threshold [16,17]. The usefulness of our simple framework is that we examine all risk thresholds easily and it is easy to understand why a conclusion is reached.

3.5 | *BRCA1/2* mutation screening for the general population

In the general population, *BRCA1/2* mutations have only 0.26% prevalence (10 times lower than in Ashkenazi-Jews).

Figure 4 (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 4 (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 for screening to never be the preferred policy, which is unreasonably expensive for shared decision-making.

The dollars per life-year gained at peak INB is \$137,000, which is high even for the US. Figure 5 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 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.

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 5 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.

The more comprehensive analysis of Long and Ganz[16] 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[16]. Our approach reaches the same practical conclusion as the comprehensive decision-analysis, while also considering different risk thresholds. Our approach as allows for simple understanding of why cost-effectiveness is reduced in the general population versus in Ashkenazi-Jews: the cost of definitive genetic testing c_0 is much higher, and the mutations are 10 times rarer.

4 | SENSITIVITY ANALYSES

An important part of any decision analysis is assessing sensitivity to the input parameters. As noted in section 3.3, the INB is on the scale of the screening test cost c_m, so visually inspecting a plot like figure 4 (right panel) makes it easy to assess sensitivity to screening test cost. Below we consider sensitivity to other parameters.

4.1 | Different effectiveness and treatment costs affects dollars per life-year gained, but not optimal risk thresholds

The optimal risk threshold $R = c_0/(e+c_0)$ is a function of only definitive test cost c_0 and net effectiveness $e = \lambda(e_1-e_2)-(c_1-c_2)$. Thus, if R is fixed (as in DCA or plots of INB gain vs R), then also fixing c_0 means that e is determined. Thus if the life-years gained (e_1-e_2) or gain in treatment costs (c_1-c_2) changes, then the willingness to pay λ will adapt to ensure that e remains fixed. Thus optimal risk thresholds are entirely unaffected by changing the effectiveness or treatment costs, because λ varies to ensure that $e(\lambda)$, and hence R, remain fixed. However, the meaning of the risk threshold in terms of dollars per life-year gained will be altered. The associated dollars per life-year gained thresholds are readily calculated by plugging quantities into equation 4.

For example, if the life-years gained $(e_1 - e_2)$ increases to 6.5 from 5.0, the plot of INB gain or NB gain from DCA, versus risk threshold, such as figure 4 , is unaltered (not shown). The risk-thresholds for which BRCAPRO screening is viable in the general population remains $(0.13\%, 0.91\%)$ in spite of the increase in effectiveness. However, because life-years gained increased, equation 4 calculates that the dollars per life-year gained threshold implied by these thresholds now decreases to (\$12,000, \$240,000). Similarly, if the difference in treatment costs $(c_1 - c_2)$ is halved, the risk-thresholds supporting screening remain $(0.13\%, 0.91\%)$, but their associated dollars per life-year gained thresholds now increase to (\$32,000, \$322,000) because screening saves less money.

4.2 | Sensitivity to cost of the definitive test

In contrast, by equation 5, varying c_0 (fixing R and c_m) changes the toptimal risk thresholds R. This can be examined empirically by plugging in different c_0. For example, if the cost of *BRCA1/2* genetic testing in the general population was halved from \$2200 to \$1100, the range of risk thresholds supporting screening shrinks from $(0.13\%, 0.91\%)$ to $(0.14\%, 0.68\%)$ and their associated dollars per life-year gained thresholds improve from (\$16,000, \$306,000) to (\$3,000, \$125,000).

A pressing sensitivity analysis is the critical cost of definitive *BRCA1/2* genetic testing at which testing all women in the general population is better than any screening. This can be found by solving for c_0 in $INB_{Gain} = 0$, while fixing dollars per

life-year gained (i.e. fixing net effectiveness e):

$$c_0 = \frac{c_m + e(\lambda) \cdot P(D+, M-)}{P(D-, M-)}.$$

The critical cost, calculating $e(\lambda)$ by fixing the dollars per life-year gained threshold at $\lambda = \$0$, represents a minimal cost at which genetic testing for all women is the best decision, because λ is the lowest ethical willingness to pay. For $\lambda = 0$, the critical cost of *BRCA1/2* genetic testing is \$397. Thus, if the cost of *BRCA1/2* genetic testing in the general population fell to \$397, then testing all women is preferred versus screening, regardless of your threshold for dollars per life-year gained. Of course, any real willingness to pay threshold will be higher than 0, allowing higher critical definitive test costs. At willingness to pay thresholds of \$20,000, \$50,000, and \$100,000, the critical costs of *BRCA1/2* genetic testing are \$532, \$746, and \$1068, respectively. Although \$1000 *BRCA1/2* testing will be achieved soon, \$400 testing may take some time to achieve.

5 | DISCUSSION

We proposed a simple framework for identifying optimal risk-thresholds for a single-time screening test. Our framework provides a simple equation not only for the optimal risk-threshold, but more importantly, also for the dollars per life-year gained underlying an optimal risk-threshold. The dollars per life-year gained threshold is necessary to know which optimal risk-thresholds are feasible in practice. Because the dollars per life-year gained is a value that differs between people and societies, we plot the gain in Incremental Net Benefit (INB) versus the dollars per life-year gained threshold to identify the optimal risk-thresholds that support screening. Because the INB is on the scale of screening test cost, it easily identifies sensitivity to screening test costs. Although optimal thresholds are invariant to effectiveness and treatment costs, the dollars per life-year gained associated with the threshold can change substantially with effectiveness and treatment costs. Our framework relies only on simple closed-form expressions that are easy to query for sensitivity to costs and effectiveness to better understand the conclusions drawn by the INB.

Our framework fills a niche as a bridge between Decision Curve Analysis (DCA) and a full-blown proper decision-analysis. DCA is the natural next step following an analysis of statistical properties such as classification ability and risk stratification, as it introduces the decision-analytic concept of optimal risk-thresholds without requiring specification of costs or effectiveness[4]. However, because costs and effectiveness are not specified, DCA cannot identify optimal risk-thresholds. In practice, DCA is usually conducted with a costless screening test assumption. This assumption may seem reasonable when the screening test is a risk model, but risk models usually require shared decision-making in practice. We showed that, DCA can account for screening test costs, but at the cost of requiring that all costs and effectiveness also be specified, which negates the simplicity of use of DCA in practice. Because, DCA is on the abstract scale of Net Benefit, it can be difficult to see whether the DCA is sensitive to screening test cost. Most importantly, while DCA can identify a range of thresholds that support screening, it cannot assess their associated dollars per life-year gained, which is necessary to know which thresholds are feasible. Our framework could be the next step after DCA to quickly and easily examine the importance of different quantities and structural assumptions, to help plan the final and comprehensive decision-analysis.

In contrast, a full decision-analysis requires specifying not only costs and effectiveness, but also all possible decisions for different people. This approach is the most comprehesive "final" answer for identifying optimal risk thresholds. However, comprehensive decision-analyses usually require complex methods, such as decision-trees or microsimulation. These methods, although powerful and comprehensive, can be hard to understand why they arrive at their conclusion. Probabilistic sensitivity analysis helps, but readers of such papers usually do not have access to the computational model to personally assess sensitivity to inputs and structural assumptions. We suggest that comparing answers from our approach to a comprehensive approach can help informs each other. If the two approaches agree, then either complexities don't matter or they happen to cancel each other's effects. If don't agree, either the our approach is missing something critical or the comprehensive approach may be speculatively modeling something incorrectly. Resolving the disagreement could prove insightful and advance the debate on what is cost-effective, and spur reseach into any quantities identified as being critical but poorly known.

In the *BRCA1/2* example, DCA, under the usual costless screening test assumption, identified a wide range of risk thresholds (1%-30%) supporting screening among Ashkenazi-Jews. But allowing a mere \$100 cost for shared decision-making wiped out all the thresholds and made clear the best decision is to do genetic testing for all Ashkenazi-Jewish women. The DCA uses the abstract scale of Net Benefit, so it is hard to see this sensitivity, but easy via the INB which is on the scale of the cost of the screening test. In the general population, risk thresholds from (0.13%-0.91%) support screening. However, not all of these will be feasible in practice because they represent dollars per life-year gained thresholds from \$16000 – \$306000; typical thresholds

in practice are below $100,000. This example shows that knowing the optimal thresholds does not suffice; the dollars per life-year gained associated with each threshold is necessary. We demonstrated that optimal thresholds are invariant to effectiveness and treatment costs, but this masks the fact that the meaning of the threshold, via its associated dollars per life-year gained threshold, indeed varies.

Any true decision-analysis requires costs and effectiveness. For the effectiveness measures, we chose life-years gained. However, any scale for which can be valued on a dollars scale could also be used, such as, quality-adjusted life-years gained. Costs can be difficult to identify and vary with place and time, but our framework limits to 4 key costs to identify. Because costs and effectiveness are variable, we conducted simple sensitivity analysis, but formal probabilistic sensitivity analysis could be imposed on our simple framework as well.

Although it may be tempting to jump to a comprehensive decision analysis as the "final" answer, it is important to understand all aspects of a problem to better understand the "final" answer. Adoption of biomarkers into screening follows a sequential process of developing a validated assay, assessing its association with disease, assessing its predictiveness, to assessing its ability to stratify disease risk[4]. These steps encompass the scientific aspects. The next steps require decision analysis, including DCA and comprehesinve decision analysis. Our approach to estimating optimal risk thresholds in simple and transparent manner, providing intuition about which quantities are critical, may serve as a bridge between DCA and a full decision analysis.

Stuart Baker: Think about the ICER.

References

1. Pauker S. G., Kassirer J. P.. The threshold approach to clinical decision making.. *N Engl J Med.* 1980;302(20):1109–1117.

2. Gail Mitchell H, Pfeiffer Ruth M. On criteria for evaluating models of absolute risk.. *Biostatistics (Oxford, England).* 2005;6:227–239.

3. Baker Stuart G, Cook Nancy R, Vickers Andrew, Kramer Barnett S. Using relative utility curves to evaluate risk prediction.. *Journal of the Royal Statistical Society. Series A, (Statistics in Society).* 2009;172:729–748.

4. Katki Hormuzd A.. Quantifying risk stratification provided by diagnostic tests and risk predictions: Comparison to AUC and Decision Curves. *Stat Med.* 2019;.

5. Vickers Andrew J, Elkin Elena B. Decision curve analysis: a novel method for evaluating prediction models.. *Med Decis Making.* 2006;26:565–574.

6. Kuchenbaecker Karoline B, Hopper John L, Barnes Daniel R, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers.. *JAMA.* 2017;317:2402–2416.

7. Moyer Virginia A. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: U.S. Preventive Services Task Force recommendation statement.. *Ann Int Med.* 2014;160:271–281.

8. NICE . *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.*. : National Institute for Health and Care Excellence Clinical Guidance; 2017.

9. King Mary-Claire, Levy-Lahad Ephrat, Lahad Amnon. Population-based screening for BRCA1 and BRCA2: 2014 Lasker Award.. *JAMA.* 2014;312:1091–1092.

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

11. Rabin Roni Caryn. F.D.A. Approves First Home Testing for 3 Breast Cancer Mutations, With Caveats. *New York Times.* March 6, 2018;.

12. Best Ana F, Tucker Margaret A, Frone Megan N, Greene Mark H, Peters June A, Katki Hormuzd A. A Pragmatic Testing-Eligibility Framework for Population Mutation Screening: The Example of *BRCA1/2*.. *Cancer Epidemiol Biomarkers Prev.* 2019;28:293–302.

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

14. Struewing J. P., Hartge P., Wacholder S., et al. The Risk of Cancer Associated With Specific Mutations of BRCA1 and BRCA2 Among Ashkenazi Jews. *N. Engl. J. Med.*. 1997;336:1401-1408.

15. Sigal Bronislava M, Munoz Diego F, Kurian Allison W, Plevritis Sylvia K. A simulation model to predict the impact of prophylactic surgery and screening on the life expectancy of BRCA1 and BRCA2 mutation carriers.. *Cancer epidemiology, biomarkers & prevention.* 2012;21:1066–1077.

16. Long Elisa F, Ganz Patricia A. Cost-effectiveness of Universal BRCA1/2 Screening: Evidence-Based Decision Making.. *JAMA oncology.* 2015;1:1217–1218.

17. Manchanda Ranjit, Legood Rosa, Burnell Matthew, et al. Cost-effectiveness of population screening for BRCA mutations in Ashkenazi jewish women compared with family history-based testing.. *Journal of the National Cancer Institute.* 2015;107:380.

WebAppendix for the Paper:

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

by Hormuzd A. Katki and Ionut Bebu

1 | DERIVIATION OF THE OPTIMAL COST-EFFECTIVE RISK THRESHOLD FROM CLASSICAL EXPECTED UTILITY THEORY

The optimal risk threshold can be derived from the classical expected utility approach[1]. 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:

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

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 $R = C/(B + C)$[1]. For a single screen, note that $B = e$ and $C = c_0$. Thus the optimal cost-effective risk-threshold is

$$R = \frac{C}{B + C} = \frac{c_0}{e + c_0}. \tag{6}$$

The value of this derivation is that it clarifies which "cost" is C (here, definitive test cost c_0) and what is the definition of "benefit" B (here, it is net effectiveness of early intervention e).

2 | INTERVENTION COSTS FOR WOMEN WITH *BRCA1/2* MUTATIONS

We obtain all costs from the supplement of Long and Ganz (2015). They report 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 $\$86013 + 8 \times 7547 + 63790 = \$210179$. 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 $\$124838 + 3 \times 13724 + 87218 = \$253,228$. 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 $c_1 = \$12286 + 7393 + 0.027 \times 210179 + 0.012 \times 253228 = \$28,393$. Because 17.1% of women with *BRCA1/2* mutations will not develop breast or ovarian cancer (53% will develop breast cancer and 29.9% will develop ovarian cancer), $c_2 = \$0.53 \times 210179 + 0.299 \times 253228 + 0.171 \times 0 = \$187,110$.

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. Although we assume the chance of developing both breast and ovarian cancer is negligible, this could be accounted for.

□

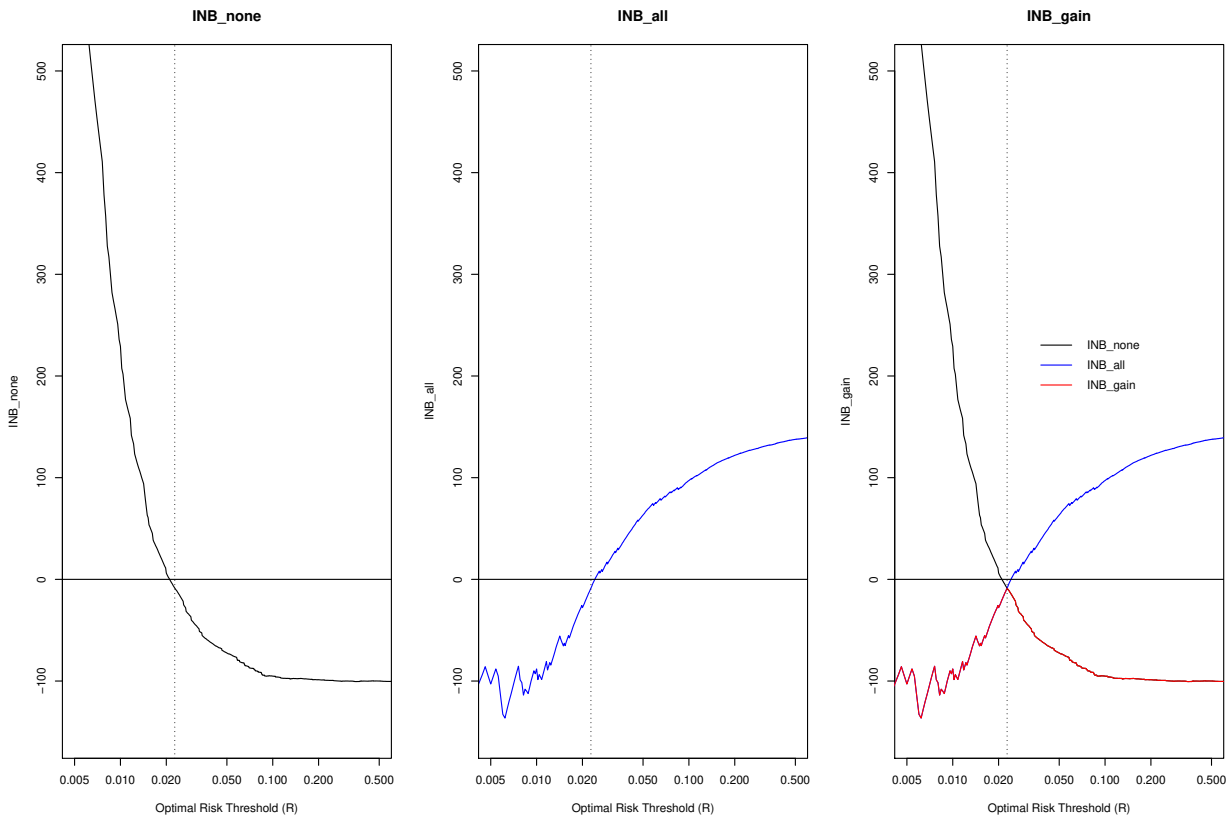

FIGURE 6 Left panel is Incremental Net Benefit of screening vs testing no one (INB_{none}). Thresholds where $INB_{none} > 0$ favor screening versus testing no one. Middle panel is Incremental Net Benefit of screening vs testing everyone (INB_{all}). Thresholds where $INB_{all} > 0$ favor screening versus testing everyone. Right panel is plots both INB_{none} and INB_{all}, highlighting in red the part of each that is $INB_{Gain} = min(INB_{none}, INB_{all})$. The vertical line is where the optimal risk threshold equals disease prevalence, which is where INB_{Gain} switches over from being equal to INB_{all} to being equal to INB_{none}. The INB_{Gain} equals that of Figure 1 (right panel) of the article.