Published in final edited form as:

Med Decis Making. 2018 February ; 38(2): 225–234. doi:10.1177/0272989X17732994.

Simple decision-analytic functions of the AUC for ruling out a risk prediction model and an added predictor

Stuart G. Baker, ScD

Division of Cancer Prevention, National Cancer Institute, Bethesda USA

Abstract

When using risk prediction models, an important consideration is weighing performance against the cost (monetary and harms) of ascertaining predictors. The minimum test tradeoff (MTT) for ruling out a model is the minimum number of all-predictor ascertainment per correct prediction to yield a positive overall expected utility. An approximation to this MTT is $1/[P(H(AUC_{model}))]$, where $H(AUC) = AUC - \{1/2(1-AUC)\}^{1/2}$, AUC is the area under the Receiver Operating Characteristic (ROC) curve, and P is the probability of the predicted event in the target population. The MTT for ruling out an added predictor is the minimum number of added-predictor ascertainment per correct prediction to yield a positive overall expected utility. An approximation to this MTT is $1/[P\{(H(AUC_{Model:2}) - H(AUC_{Model:1}))\}]$, where Model 2 includes an added predictor relative to Model 1. This approximation requires the Tangent Condition that the true positive rate at the point on the ROC curve with a slope of 1 is larger for Model 2 than Model 1. These approximations are suitable for back-of-the-envelope calculations. For example, in a study predicting the risk of invasive breast cancer, Model 2 adds to the predictors in Model 1 a set of seven single nucleotide polymorphisms (SNPs). Based on the AUC's and the Tangent Condition, I computed an MTT of 7200, which says that 7200 sets of SNP's are needed for every correct prediction of breast cancer to yield a positive overall expected utility. If ascertaining the SNP's costs \$500, this MTT suggests that SNP ascertainment is not likely worthwhile for this risk prediction.

Because risk prediction models are often used in medical decision making, there is great interest in improving their performance by incorporating an added predictor (or set of predictors). Predictors could include a biomarker (such as a protein in the blood), a risk factor (such as age), or the result of an imaging study. An important question is whether the increased risk prediction performance from an added predictor is worth the cost (monetary cost and harms) of ascertaining the added predictor. Let AUC denote the area under the Receiver Operating Characteristic (ROC) curve. Let Model 1 denote the basic risk prediction model, and let Model 2 denote a risk prediction model with the same predictors as Model 1 plus an added predictor (or set of predictors). Standard metrics for the statistical evaluation of an added predictor, such as the odds ratio, the difference in the AUC between Model 1 and Model 2 [1], and the net reclassification improvement (NRI) [2, 3], do not address this question

As an example consider the prediction of the 5-year rate of invasive breast cancer in asymptomatic women [4]. Women at high risk could receive daily tamoxifen or raloxifene pills to lower the risk of breast cancer, although there are serious side effects including an

increased risk of uterine cancer and thromboembolic events [5]. Model 1 involves predictors obtained via questionnaire, such as age and family history of breast cancer. Model 2 adds to the predictors in Model 1 a set of seven single nucleotide polymorphisms (SNPs). The ascertainment of SNP's involves genotyping blood at a considerable monetary cost. A comparison of Model 2 versus Model 1 yielded the following results: the largest odds ratio among the seven SNP's was 1.35, the AUC increased from 0.56 to 0.59, and the net reclassification improvement based on three risk categories (<1.5%, 1.5% to 2%, and >2%) was 0.085. Based on these metrics, the investigators characterized the addition of SNP's to the risk prediction model as a "modest improvement" [4]. However, these statistical measures do not summarize the tradeoff between the improvement in risk prediction with SNP's and the cost ascertaining the SNP's.

Relative utility curves [6, 7] are a decision-analytic approach to evaluating risk prediction. They are related to decision curves [8] but with a different theoretical foundation. I discuss the use of relative utility curves to derive the minimum test tradeoff (MTT), which was previously called the summary test tradeoff [9]. MTT is useful for ruling out a risk prediction model or ruling out an added-predictor based on the performance of prediction and the cost of ascertaining predictors. The key contributions here are approximations to the MTT that only involve the AUC and the probability of the predicted event in the target population.

ROC and Relative Utility Curves

The starting point is a risk prediction model fit to a training sample. Applying this risk prediction model to a validation sample yields a score for each person in the validation sample. The score should have the property that, in the absence of sampling variability, risk increases as score increases. A basic measure of the prediction performance of the score in the validation sample is the ROC curve. The validation sample should be representative of a target population in the sense that both have the same ROC curves. Let s denote a cutpoint of the score. Let $FPR(s) = pr(\text{score} > s \mid \text{no event})$ denote the false positive rate at cutpoint s . Let $TPR(s) = pr(\text{score} > s \mid \text{event})$ denote the true positive rate at cutpoint s . The ROC curve is a plot of $TPR(s)$ versus $FPR(s)$. Let $Risk(s) = pr(\text{event} \mid \text{score} = s)$ denote the risk at cutpoint s . Let P denote the probability of the predicted event in the target population. The slope of the ROC curve is a simple function [7, 9] of P and $Risk(s)$,

$$ROCSlope(s) = \{(1 - P)/P\} \{Risk(s)/(1 - Risk(s))\}. \quad (1)$$

See Appendix A for a derivation. Viewing P as a prior probability, $ROCSlope(s)$ as a likelihood ratio, and $Risk(s)$ as a posterior probability, equation (1) is equivalent to the odds ratio formulation of Bayes formula [10].

Net Benefits

Three relevant decision rules are treat-none, treat-all, and treat-from-prediction. Treat-from-prediction at cutpoint s means receiving treatment if the score is greater than s (defined as a positive result) and receiving no treatment if the score is less than or equal to s (defined as a

negative result). To evaluate the decision rules, it is necessary to consider benefits and costs. Let B denote the benefit of a true positive, and let C denote the cost (including monetary costs and harms) of a false positive. For the aforementioned example involving the prediction of breast cancer, B is the benefit of reduced breast cancer risk arising from daily pills in a true positive (a woman correctly predicted to develop breast cancer in the absence of the daily pills). For the same example, C is the cost of an increased risk of uterine cancers and blood clots arising from daily pills in a false positive (a woman incorrectly predicted to develop breast cancer in the absence of daily pills).

An expected utility is the sum of expected benefits minus expected costs (monetary costs and harms), where costs and benefits are in the same units. The motivation behind the following mathematical development is to evaluate decision rules without explicitly specifying the benefits and the costs. I define the net benefit of treat-from-prediction versus treat-none as the expected utility of treat-from-prediction minus the expected utility of treat-none, which simplifies to

$$NBPred_{TreatNone}(s) = P \text{TPR}(s)B - (1 - P) \text{FPR}(s)C. \quad (2)$$

In terms of population quantities, the net benefit in the decision curve literature [8] equals $NBPred_{TreatNone}(s)/B$. I define the net benefit of treat-from-prediction versus treat-all as the expected utility of treat-from-prediction minus the expected utility of treat-all, which simplifies to

$$NBPred_{TreatAll}(s) = (1 - P)\{1 - \text{FPR}(s)\}C - P\{1 - \text{TPR}(s)\}B. \quad (3)$$

Equations (2) and (3) follow from more basic equations involving utilities of false and true positives and false and true negatives [6, 7]. See Appendix B for a derivation.

The risk threshold T is the risk of the event at which a person would be indifferent between treatment and no treatment. If a person with score s has $Risk(s) > T$, the person would, by definition, prefer treatment. An important result in decision analysis [11] is $C/B = T/(1 - T)$. See Appendix C for a derivation. An ROC curve is concave if the slope decreases monotonically from left to right. Another important result in decision analysis [12, 13] is that, for a concave ROC curve, the maximum the net benefit of treat-from-prediction versus treat-none or treat-all at risk threshold T occurs at cutpoint s that solves

$$ROCSlope(s) = \{(1 - P)/P\}\{T/(1 - T)\} \text{ or, equivalently, } Risk(s) = T. \quad (4)$$

See Appendix D for a derivation. Let $s_{opt}(T)$ denote the optimal cutpoint defined as the value of s that solves equation (4). By definition, the maximum net benefit of treat-from-prediction versus treat-none and treat-all are, respectively,

$$\max NB\text{Pred}_{TreatNone}(T) = NB\text{Pred}_{TreatNone}(s_{opt(T)}), \quad (5)$$

$$\max NB\text{Pred}_{TreatAll}(T) = NB\text{Pred}_{TreatAll}(s_{opt(T)}). \quad (6)$$

Relative Utility

Relative utility is the *maximum* net benefit of treat-from-prediction divided by the net benefit of treat-from-perfect-prediction,

$$RU(T) = \begin{cases} \max NB\text{Pred}_{TreatAll}(T) / NB\text{PerfectPred}_{TreatAll}, & \text{for a treat-all default,} \\ \max NB\text{Pred}_{TreatNone}(T) / NB\text{PerfectPred}_{TreatNone}, & \text{for a treat-none default.} \end{cases} \quad (7)$$

The net benefits of treat-from-perfect-prediction versus treat-none and treat-all are obtained by substituting $TPR(s)=1$ and $FPR(s)=0$ into the equations (2) and (3) to yield $NB\text{PerfectPred}_{TreatNone} = PB$ and $NB\text{PerfectPred}_{TreatAll}(s) = (1-P)C$, respectively. Because the net benefit of treat-from-perfect prediction is less than or equal to the net benefit of treat-from-prediction and because net benefits are non-negative, $0 \leq RU(T) \leq 1$. The relative utility does not include the cost of ascertaining predictors. The treat-none default occurs when the expected utility of treat-none is larger than the expected utility of treat-all, corresponding to $T \geq P$. The treat-all default occurs in the complementary situation corresponding to $T < P$. Extensive simplification of equation (7) yields

$$RU(T) = \begin{cases} \{1 - FPR(s_{opt(T)})\} - \{1 - TPR(s_{opt(T)})\} / \text{ROCSlope}(s_{opt(T)}), & \text{if } T < P, \\ TPR(s_{opt(T)}) - \text{ROCSlope}(s_{opt(T)}) FPR(s_{opt(T)}), & \text{if } T \geq P. \end{cases} \quad (8)$$

Relative Utility Curves

A relative utility curve is a plot of $RU(T)$ versus T . The maximum of the relative utility curve occurs at $T = P$, with the relative utility curve monotonically decreasing on both sides of $T = P$. See Appendix E for a proof. Substituting $T = P$ into equation (4) gives the important result that the maximum of the relative utility curve corresponds to the point on the ROC curve where $\text{ROCSlope}(s_{opt(P)}) = 1$. From equation (8), the maximum relative utility is

$$RU(P) = TPR(s_{opt(P)}) - FPR(s_{opt(P)}). \quad (9)$$

Construction of a relative utility curve requires only a concave ROC curve and specification of P . Concavity of the ROC curve is needed to ensure that relative utility is based on the *maximum* net benefit of treat-from-prediction in equation (7). If the original ROC curve is

not concave, it is necessary to transform it to a concave ROC curve. To create a concave ROC curve from a non-concave ROC curve, I use a piecewise-linear concave envelope of the original ROC curve, which has a decision-analytic optimality [7].

To construct the relative utility curve, it is not necessary to start with T and then compute $RU(T)$. Instead, I use the following backdoor approach. Each point (with a corresponding FPR , TPR and $ROCSlope$) on the concave ROC curve corresponds to a point on the relative utility curve. Substituting FPR , TPR and $ROCSlope$ into equation (8), without explicit specification of $s_{opt}(T)$, yields RU . Substituting $ROCSlope$ and P into equation (1) yields $Risk$, which, according to equation (4) is the risk threshold T at the optimal cutpoint. These calculations yield one point (T, RU) on the relative utility curve. Repeating this pair of calculations for various points on the piecewise linear concave ROC curve yields the relative utility curve.

Calibration is the process of adjusting the score (typically when it is a predicted risk) from the training sample to obtain an unbiased estimate of the predicted risk in the target population, which can be compared to the risk threshold for the target population. With relative utility curves, calibration is automatic because $Risk$ in equation (1) applies to the target population via two quantities that apply to the target population, $ROCSlope$ and P . Thus, the risk threshold, which equals $Risk$ at the optimal cutpoint in equation (4), applies to the target population. The relative utility, which is a function of the population quantities FPR , TPR and $ROCSlope$, also applies to the target population.

Figure 1 presents ROC and relative utility curves for hypothetical data generated according to the procedure in Appendix F. Figure 2 presents these curves for the aforementioned study involving breast cancer risk prediction. The curves in Figure 2 are not smooth because the data came from a published table with only three categories of risk [4]. The treat-none and treat-all labels in Figures 1 and 2 refer to the regions of the risk threshold where those decisions yield the largest expected utilities.

Ruling Out a Risk Prediction Model

Let “all-predictor ascertainment” denote ascertaining all the predictors in the model. Let C_{model} denote the cost of all-predictor ascertainment, which includes the monetary costs of data collection as well as monetary costs and harms of any required tests. The quantity C_{model} is in the same units as B and C . The motivation for the test tradeoff for the model, denoted $TestTradeoff_{model}(T)$, is to draw conclusions about the relationship among C_{model} , B and C , without specifying these quantities.

Let $maxNBPred(T)$ denote either $maxNBPred_{TreatNone}(T)$ or $maxNBPred_{TreatAll}(T)$, depending on whether the default in the absence of risk prediction is treat-none or treat-all. I define the overall expected utility for the model as the maximum net benefit of treat-from-prediction minus the cost of all-predictor ascertainment, namely $maxNBPred_{model}(T) - C_{model}$. A risk prediction model is worthwhile if the overall expected utility is positive, which implies $TestTradeoff_{model}(T) < B/C_{model}$ where

$$TestTradeoff_{model}(T) = \begin{cases} (1 - T)/[T(1 - P) RU_{model}(T)], & \text{for a treat-all default,} \\ 1/[P RU_{model}(T)], & \text{for a treat-none default.} \end{cases} \quad (10)$$

See Appendix G for a derivation of equation (10). The ratio B/C_{model} is the number of all-predictor ascertainment per correct prediction. To see this, think of B as the price of an apple (benefit of a correct prediction) and C_{model} as the price of an orange (cost of all-predictor ascertainment). Then B/C_{model} is the number of oranges (all-predictor ascertainment) traded for each apple (correct prediction).

The key to understanding $TestTradeoff_{model}(T)$ is that it is a lower bound for B/C_{model} in equation (10). Consequently, $TestTradeoff_{model}(T)$ is the minimum number of all-predictor ascertainment per correct prediction of the event *to yield a positive overall expected utility at risk threshold T* . The main limitation to applying $TestTradeoff_{model}(T)$ is specifying the risk threshold T . A standard approach to circumvent this limitation is to perform a sensitivity analysis for the test tradeoff over a plausible range of risk thresholds T [7, 14]. However, this sensitivity analysis also has limitations, namely, (i) specifying the plausible range of risk thresholds can be difficult because of insufficient knowledge of preferences, (ii) estimating the test-tradeoff at a risk threshold requires individual-level data that may be unavailable, and (iii) estimating test tradeoff can be unreliable for extreme values of the risk threshold due to small sample sizes

Minimum Test Tradeoff for the Model

Investigators can circumvent the limitations of the test tradeoff sensitivity analysis by considering a more limited form of inference, namely ruling out the risk prediction model. To rule out the risk prediction model, I use the minimum of $TestTradeoff_{model}(T)$ over all possible risk thresholds, which I denote MTT_{model} . The reason MTT_{model} leads to ruling out a risk prediction model is that an unacceptably high MTT_{model} implies an unacceptably high $TestTradeoff_{model}(T)$ for all values of T . To rigorously rule in a risk prediction model, one would need a maximum of the test tradeoff over a plausible range of risk thresholds, which is difficult to specify. Based on equations (9) and (10),

$$MTT_{model} = 1/\{P RU_{model}(P)\}. \quad (11)$$

MTT_{model} is the minimum number of all-predictor ascertainment per correct prediction of the event *to yield a positive overall expected utility*. To illustrate the use of MTT_{model} , I present the following example. Suppose that ascertaining all predictors involves performing sets of expensive genetic tests, the predicted event is not life-threatening, and $MTT_{model} = 1000$. In other words, a minimum of 1000 sets of expensive genetics tests would be needed for each correct prediction of an event that is not life-threatening in order to yield a positive overall expected utility. In this context $MTT_{model} = 1000$ would likely rule out the risk prediction model.

Approximation to the Minimum Test Tradeoff for the Model

Approximating MTT_{model} involves approximating the maximum relative utility, $RU_{model}(P)$ by

$$RU_{model(approx)}(P) = H(AUC) = AUC - \{(1 - AUC)/2\}^{1/2}. \quad (12)$$

I derived equation (12) using Figure 3, where point A is the point on the ROC curve with slope =1, f is the false positive rate at the point A, and $h = TPR - FPR$ at point A. Based on equation (9), h is the maximum relative utility. The ROC curve has an upper bound at line segment GB that is tangent to the ROC curve at point A, so has a slope of 1. The AUC of the upper bound equals $1 - \text{area of } DGB = 1 - \frac{1}{2}(1 - h)^2$. The ROC curve has a lower bound at line segments FA and AC. The AUC of the lower bound equals $\frac{1}{2} + \text{area of } FAE + \text{area of } ACE = \frac{1}{2} + \frac{1}{2}h(1 - f) + \frac{1}{2}hf = \frac{1}{2} + \frac{1}{2}h$. Separately solving for h in terms of AUC for the upper and lower bounds and taking the average gives $h = H(AUC)$ in equation (12). For a variety of ROC curves, the approximate maximum relative utility was close to its exact value (Figure 4). Substituting equation (12) into equation (11) yields the approximate MTT for the model,

$$MTT_{model(approx)} = 1/\{P H(AUC_{model})\}. \quad (13)$$

Ruling out an Added Predictor

The derivation of the MTT for ruling out an added predictor parallels the derivation of the MTT for ruling out a risk prediction model, but with an added condition. Let C_{added} denote the cost or harm of ascertaining the added predictor with Model 2 in comparison with Model 1. The motivation for the test tradeoff for the added predictor, denoted $TestTradeoff_{added}(T)$, is to draw conclusions about the value of adding a predictor without specifying C_{added} , B , or C . I define the overall expected utility for the added predictor as the maximum net benefit of treat-from-prediction is larger with Model 2 than Model 1, namely $maxNB_{Pred_{Model:2}}(T) - C_{added} - maxNB_{Pred_{Model:1}}(T)$. An added predictor is worthwhile if the overall expected utility is positive, which implies $TestTradeoff_{added}(T) < B/C_{added}$, where

$$TestTradeoff_{added}(T) = \begin{cases} (1 - T)/[T(1 - P)\{RU_{Model:1}(T) - RU_{Model:2}(T)\}], & \text{for treat-all default,} \\ 1/[P\{(RU_{Model:2}(T) - RU_{Model:1}(T))\}], & \text{for treat-none default.} \end{cases} \quad (14)$$

The ratio B/C_{added} is the number of added-predictor ascertainment per correct prediction. The key to understanding $TestTradeoff_{added}(T)$ is that it is a lower bound for B/C_{added} in equation (14). Therefore, $TestTradeoff_{added}(T)$ is the minimum number of added predictor ascertainment per correct prediction to yield a positive overall expected utility at risk threshold T .

Minimum Test Tradeoff for the Added Predictor

The minimum of $TestTradeoff_{added}(T)$ over all possible risk thresholds T is

$$MTT_{added} = 1/[P \max_T \{RU_{Model2}(T) - RU_{Model1}(T)\}]. \quad (15)$$

MTT_{added} is the minimum number of added-predictor ascertainment per correct prediction of the event *to yield a positive overall expected utility*. To illustrate the use of MTT_{added} , I present the following example. Suppose that ascertaining the added predictor involves an invasive test with a high-risk for mortality and morbidity, a predicted event of breast cancer incidence, and $MTT_{added} = 4000$. In other words, a minimum of 4000 high-risk invasive tests would be needed for each correct prediction of breast cancer to yield a positive overall expected utility. In this example, $MTT_{added} = 4000$ would likely rule out using the result of the invasive test as added predictor.

Approximation to the Minimum Test Tradeoff for the Added Predictor

The first step in approximating MTT_{added} in equation (15) is to substitute the difference in maximum relative utilities for the maximum difference in relative utilities,

$$MTT_{added(difmax)} = 1/[P\{RU_{Model2}(P) - RU_{Model1}(P)\}]. \quad (16)$$

I introduce a condition, which I call the Tangent Condition, to help indicate when $MTT_{added(difmax)}$ equals or approximates MTT_{added} . The Tangent Condition says that the point on the ROC curve with $ROCSlope=1$ has a larger TPR for Model 2 than Model 1. I justify the Tangent Condition based on a heuristic geometric argument involving Figures 5 and 6.

Figure 5 illustrates the Tangent Condition under the treat-none default ($T = P$), but a similar argument would apply to a treat-all default. Based on equation (8) for the treat-none default, namely $RU = TPR - ROCSlope \cdot FPR$, the tangent lines intersect the vertical axis at RU . Points A1 and A2 denote the points on the ROC curves with $ROCSlope=1$ for Models 1 and 2, respectively. Because $ROCSlope=1$ corresponds to the maximum relative utility, the tangent line for Point A1 intersects the vertical axis at the maximum relative utility for Model 1, denoted $maxRU1$. Similarly, the tangent line for Point A2 intersects the vertical axis at the maximum relative utility for Model 2, denoted $maxRU2$. Points B1 and B2 denote the points on the ROC curves with $ROCSlope=2$ for Models 1 and 2, respectively. $RU1b$ and $RU2b$ denote the relative utilities at which the tangent lines for Point B1 and B2 intersect the vertical axis. Geometrically, the Tangent Condition says that Point A2 is higher than Point A1. In many cases the Tangent Condition implies a steeper ROC curve between Points A and B for Model 2 than Model 1, which typically implies $maxRU2 - maxRU1$ is larger than $RU2b - RU1b$, as in Figure 5. See Appendix H for the exact conditions for $maxRU2 - maxRU1$ to be larger than $RU2b - RU1b$. The exact conditions are harder to check than the Tangent Condition, which is why the Tangent Condition is presented instead.

To more fully investigate the Tangent Condition, I considered eight panels of ROC and relative utility curves in Figure 6. In panels (A)–(D), the Tangent Condition holds, $RU(P)$ is the largest difference between the relative utility curves, and hence $MTT_{added(difmax)}$ equals MTT_{added} . In panels (E)–(F), the Tangent Condition holds, $RU(P)$ approximates the largest difference between the relative utility curves, and hence $MTT_{added(difmax)}$ approximates MTT_{added} . In panels (G) and (H), the Tangent Condition does not hold, $RU(P)$ is substantially smaller than the largest difference between the relative utility curves, and hence $MTT_{added(difmax)}$ poorly approximates MTT_{added} .

The next step to approximating equation (15) is to substitute equation (12) into equation (16) to yield,

$$MTT_{added(approx)} = 1/[P\{H(AUC_{Model:2}) - H(AUC_{Model:1})\}]. \quad (17)$$

Examples

Visual inspection of the ROC curve in Figure 1 shows that the Tangent Condition holds. For Figure 1 with $P=0.2$, $MTT_{added(approx)}=88$, which was close to the exact value of $MTT_{added}=101$. Visual inspection of the ROC curve in Figure 2 also suggested the Tangent Condition holds. For Figure 2 with $P=0.003$, $MTT_{added(approx)}=7200$, which was close to the exact value of $MTT_{added}=7500$. In other words, based on the approximation, it is necessary to measure 7200 sets of SNP's for every correct prediction of breast cancer to yield a positive overall expected utility. If ascertaining SNP's costs \$500, $MTT_{added(approx)}=7200$ suggests that SNP ascertainment is not worthwhile.

Discussion

If individual-level or count data are available, investigators can use software written in Mathematica to compute concave ROC curves, relative utility curves and the entire test tradeoff curve. See <https://prevention.cancer.gov/about-dcp/staff-search/stuart-g-baker-scd/evaluating-risk-prediction>. The main purpose of the approximation to the MTT is a back-of-the envelope calculation based on limited reported data.

Before computing the approximation to MTT (either for the model or added predictor), one should check that the ROC curve is concave or almost concave. Small deviations from ROC curve concavity should have little effect on conclusions, while large deviations from concavity could introduce substantial bias. If there are large deviations from concavity, the risk prediction model is poor (as risk does not monotonically increase with the score) and should not be used.

There is no guarantee that the Tangent Condition will always yield a good approximation for the MTT for the added predictor, but it was a reasonable condition in the examples in Figure 6. For perspective, one should realize that even with the entire test tradeoff or relative utility curve, the determination of MTT involves some judgment as to the whether the minimum of a test tradeoff that does not occur at the $T=P$ occurs at a plausible risk threshold. Also, one should appreciate the importance of correctly specifying the probability of the predicted

event in the target population (for either MTT). Particularly when the probability of a predicted event is small, as when the predicted event is cancer incidence, MTT can be sensitive to this probability and investigators may wish to perform a sensitivity analysis involving this probability.

Remarkably the approximation to the MTT involves the AUC, a quantity that does not contain information on costs and benefits. The intuition is that the MTT depends on the maximum relative utility, which corresponds to a slope of 1 on the ROC curve at a point which visually anchors the entire ROC curve. In summary, if published results from risk prediction analyses report only ROC curves and AUC's, the approximation to the MTT is useful for a back-of-the-envelope calculation for evaluating a risk prediction model or added risk predictor relative the cost of predictor ascertainment. This approximation to MTT is preferential to standard metrics which do not consider the tradeoffs between improved prediction and the cost of predictor ascertainment.

Acknowledgments

Financial support for this study was provided by The National Cancer Institute. The study sponsor had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

References

1. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. *Journal of the American Medical Association*. 2009; 302:2345–2352. [PubMed: 19952321]
2. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27:157–172. [PubMed: 17569110]
3. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology*. 2014; 25(1):114–21. [PubMed: 24240655]
4. Mealiffe ME, Stokowski RP, Rhee BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst*. 2010; 102(21):1618–1627. [PubMed: 20956782]
5. Vogel VG. The NSABP study of tamoxifen and raloxifene (STAR) trial. *Expert Rev Anticancer Ther*. 2009; 9:51–60. [PubMed: 19105706]
6. Baker SG, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. *J Royal Stat Soc Series A*. 2009; 172:729–748.
7. Baker SG, Van Calster B, Steyerberg EW. Evaluating a new marker for risk prediction using the test tradeoff: An update. *Int J Biostat*. 2012; 8:5.
8. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Dec Making*. 2006; 26(6):565–574.
9. Baker SG. The summary test tradeoff: a new measure of the value of an additional risk prediction marker. *Stat Med*. 2017; doi: 10.1002/sim.7132
10. Weinstein, MC., Feinberg, HV., Elstein, AS., et al. *Clinical Decision Analysis*. WB Saunders Company; Philadelphia: 1980.
11. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *New England Journal of Medicine*. 1980; 302:1109–1117. [PubMed: 7366635]
12. Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*. 1978; 8:283–298. [PubMed: 112681]

13. Gail MH, Pfeiffer RM. On criteria for evaluating models for absolute risk. *Biostatistics*. 2005; 6:227–239. [PubMed: 15772102]
14. Baker SG. Putting risk prediction in perspective: relative utility curves. *J Natl Cancer Inst*. 2009; 101:1538–1542. [PubMed: 19843888]
15. Pencina MJ, Steyerberg EW, D'Agostino RB Sr. Net reclassification index at event rate: properties and relationships. *Stat Med*. 2016; doi: 10.1002/sim.7041

Appendix A

This Appendix derives the relationship between *Risk* and *ROCslope*. Let $W(x) = \text{pr}(\text{score} = x)$. Then $FPR(s) = \int_s^\infty \{1 - Risk(x)\} W(x) dx / (1 - P)$ and $TPR(s) = \int_s^\infty Risk(x) W(x) dx / P$, implying $ROCslope(s) = dTPR / dFPR = (dTPR/ds) / (dFPR/ds) = \{Risk(s) / (1 - Risk(s))\} P / (1 - P)$.

Appendix B

This Appendix derives the equations for the expected utility of treat-from-prediction. Let $U_{TreatAll,Event}$ denote utility of treatment if the event would occur without treatment. Let $U_{NoTreat:Event}$ denote the utility of no treatment if the event would without treatment. Let $U_{NoTreat:NoEvent}$ denote the utility of no treatment if no event would occur without treatment. Let $U_{Treat:NoEvent}$ denote utility of treatment if no event would occur without treatment. The net benefit of treat-from-prediction versus treat-none is $NBPred_{TreatNone}(s) = (1 - P) FPR(s) U_{Treat:NoEvent} + (1 - P) \{1 - FPR(s)\} U_{NoTreat:NoEvent} + P \{1 - TPR(s)\} U_{NoTreat:Event} + P TPR(s) U_{Treat:Event} - \{(1 - P) U_{NoTreat:NoEvent} + P U_{NoTreat:Event}\}$. The net benefit of treat-from-prediction versus treat-all is $NBPred_{TreatAll}(s) = (1 - P) FPR(s) U_{Treat:NoEvent} + (1 - P) \{1 - FPR(s)\} U_{NoTreat:NoEvent} + P \{1 - TPR(s)\} U_{NoTreat:Event} + P TPR(s) U_{Treat:Event} - \{(1 - P) U_{Treat:NoEvent} + P U_{Treat:Event}\}$. Substituting $B = U_{Treat:Event} - U_{NoTreat:Event}$ and $C = U_{NoTreat:NoEvent} - U_{Treat:NoEvent}$ into the above equations yields the formulas for net benefits in the text.

Appendix C

This Appendix derives the formula for the risk threshold and its implications. Let r denote the risk of the predicted event. The net benefits of treat-none and treat-all are $NB_{TreatNone}(r) = (1 - r) U_{Treat:NoEvent} + r U_{Treat:Event}$ and $NB_{TreatAll}(r) = (1 - r) U_{Treat:NoEvent} + r U_{Treat:Event}$. Setting $NB_{TreatNone}(T) = NB_{TreatAll}(T)$ yields $T = C / (C + B)$. If $NB_{TreatNone}(P) > NB_{TreatAll}(P)$, then $T > P$, which is the treat-none default, otherwise $T < P$ and treat-all is the default.

Appendix D

This Appendix derives the optimal cutpoint of the score that maximizes the expected utility of treat-from-prediction, which is equivalent to maximizing the net benefit of treat-from-prediction versus treat-none. For a concave ROC curve, the maximum of $NBPred_{TreatNone}(s)$ occurs when $dNBPred_{TreatNone}(s) / ds = P dTPR(s) / ds B - (1 - P) dFPR(s) / ds C = 0$. This gives $ROCslope(s) \{dTPR(s) / ds\} / \{dFPR(s) / ds\} = \{(1 - P) / P\} (C / B)$.

Appendix E

This Appendix proves the maximum relative utility occurs at $T=P$ with a negative slope to the right and a positive slope to the left. The result comes from substituting $Risk$ for T , where the optimal cutpoint is implicit. For $T > P$, $RU = TPR - FPR \text{ ROC Slope}$, and $dRU/dRisk = (dTPR/dFPR) (dFPR/dRisk) - FPR (d\text{ROC Slope}/dFPR) (dFPR/dRisk) - \text{ROC Slope} (dFPR/dRisk) = - (d\text{ROC Slope}/dFPR) (dFPR/dRisk)$. Because the slope of the concave ROC curve decreases as FPR increases and FPR decreases as $Risk$ increases, this slope is negative. For $T < P$, $RU = (1 - FPR) - (1 - TPR) \text{ ROC Slope}^{-1}$, and $dRU/dRisk = -dFPR/dRisk + (dTPR/dRisk) dFPR/dTPR + \text{ROC Slope}^{-2} d\text{ROC Slope}/dRisk (1 - TPR) = (\text{ROC Slope})^{-2} d\text{ROC Slope}/dFPR (dFPR/dRisk) (1 - TPR)$. Because the slope of the concave ROC curve decreases as FPR increases and FPR decreases as $Risk$ increase, this slope is positive.

Appendix F

This Appendix describes the construction of hypothetical data in Figure 1. Let $Y=0, 1$ denote the outcome. Let X_1 and X_2 denote predictors. For $Y=0$, $X_1 \sim \text{Normal}(0,1)$ and for $Y=1$, $X_1 \sim \text{Normal}(1,1)$. For $Y=0$, $X_2 \sim \text{Normal}(0,1)$ and for $Y=1$, $X_2 \sim \text{Normal}(.5, 1)$. We generated $n=10000$ sets of $\{X_1, X_2, Y\}$ for the training sample and fit two logistic regressions: Model 1: $\text{logit}\{\text{pr}(Y=1)\} = \alpha_0 + \alpha_1 X_1$ and Model 2: $\text{logit}\{\text{pr}(Y=1)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. The parameter estimates were $\alpha_{0(est)} = -0.877$, $\alpha_{1(est)} = -1.002$, $\beta_{0(est)} = -0.748$, $\beta_{1(est)} = -1.012$, and $\beta_{2(est)} = -0.509$. I generated $n=10000$ sets of $\{X_{1(test)}, X_{2(test)}, Y_{(test)}\}$ for the test sample and computed score = $\text{expit}\{\alpha_{0(est)} + \alpha_{1(est)} X_{1(Test)}\}$ for Model 1 and score = $\text{expit}\{\beta_{0(est)} + \beta_{1(est)} X_{1(Test)} + \beta_{2(est)} X_{2(Test)}\}$ for Model 2, where $\text{expit}(x) = \exp(x) / \{1 + \exp(x)\}$. For these data, $MTT = 101$ and $MTT_{approx} = 88$.

Appendix G

This Appendix derives the test tradeoff for the risk prediction model based on the condition that the overall expected utility is positive, so $\max_{NB} \text{Pred}_{model}(T) > C_{model}$. If $T < P$, this condition says $(1 - P) C_{RU}(T) > C_{model}$ so $1 / \{(1 - P) RU(T)\} < C / C_{model}$ so $(B/C) / \{(1 - P) RU(T)\} < B / C_{model}$ so $\{(1 - T) / T\} / \{(1 - P) RU(T)\} < B / C_{model}$. If $T > P$, this condition says $P B RU(T) > C_{model}$ so $1 / (P RU(T)) < B / C_{model}$.

Appendix H

This Appendix presents exact conditions for $MTT_{added(difmax)} = MTT_{added}$. Let (FPR_{A2}, TPR_{A2}) , (FPR_{A1}, TPR_{A1}) , (FPR_{B2}, TPR_{B2}) , (FPR_{B1}, TPR_{B1}) , denote the false and true positive rates at point A2, A1, B2, and B1, respectively. The first condition is that the slope between point A and B is larger for Model 2 than Model 1, $(TPR_{A2} - TPR_{B2}) / (FPR_{A2} - FPR_{B2}) > (TPR_{A1} - TPR_{B1}) / (FPR_{A1} - FPR_{B1})$. The second condition is that the change in false positive rate from point A to B is larger for Model 2 than Model 1, $FPR_{A2} - FPR_{B2} > (FPR_{A1} - FPR_{B1})$. These conditions imply $(TPR_{A2} - TPR_{B2}) - (FPR_{A2} - FPR_{B2}) > (TPR_{A1} - TPR_{B1}) - (FPR_{A1} - FPR_{B1})$. Rearranging terms gives $(TPR_{A2} - FPR_{A2}) - (TPR_{A1} - FPR_{A1}) > (TPR_{B2} - FPR_{B2}) - (TPR_{B1} - FPR_{B1})$. Because $FPR_{B2} > FPR_{B1}$, for ROC Slope

$> 1, (TPR_{A2} - FPR_{A2}) - (TPR_{A1} - FPR_{A1}) > (TPR_{B2} - ROCSlope FPR_{B2}) - (TPR_{B1} - ROCSlope FPR_{B1})$, which is the desired result that $maxRU2 - maxRU1 > RU_{B2} - RU_{B1}$. In Figure 6, the first two conditions hold in Panels (A)–(D), only the first condition holds in Panels (E) and (F), and neither condition holds in Panels (G) and (H).

Figure 1.
ROC and RU curves for hypothetical data discussed in Appendix F. Model 1 is the baseline model, and Model 2 includes an added predictor to those in Model 1.

Figure 2. ROC and RU curves for the prediction of invasive breast cancer based on published count data. Model 1 involves baseline risk factors, and Model 2 includes the predictors in Model 1 plus seven additional SNP's.

Figure 3. Diagram for the derivation of the approximation to the maximum relative utility. The slope of the ROC curve at point A is 1.

Figure 4.
Comparison of exact and approximate maximum relative utilities for various ROC curves.

Figure 5.

A visual example of the Tangent Condition (Point A2 higher than Point A1) implying that the difference in maximum relative utilities ($maxRU2 - maxRU1$) is the maximum difference in relative utilities.

Figure 6.

The Tangent Condition holds in all panels except (G) and (H). When it holds, the difference in maximum relative utilities (at $T=P$) equals or approximates the maximum difference in relative utilities.