Published in final edited form as:

Stat Med. 2012 October 15; 31(23): 2687–2696. doi:10.1002/sim.5443.

Using multiple risk models with preventive interventions

Mitchell H. Gail^{*,†}

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA.

Abstract

An ideal preventive intervention would have negligible side effects and could be applied to the entire population, thus achieving maximal preventive impact. Unfortunately, many interventions have adverse effects as well as beneficial effects. For example, tamoxifen reduces the risk of breast cancer by about 50% and the risk of hip fracture by 45%, but increases the risk of stroke by about 60%; other serious adverse effects include endometrial cancer and pulmonary embolus. Hence, tamoxifen should only be given to the subset of the population with high enough risks of breast cancer and hip fracture such that the preventive benefits outweigh the risks. Recommendations for preventive use of tamoxifen have been based primarily on breast cancer risk. Age- and race-specific rates were considered for other health outcomes, but not risk models. In this paper, I investigate the extent to which modeling not only the risk of breast cancer, but also the risk of stroke, can improve the decision to take tamoxifen. These calculations also give insight into the relative benefits of improving the discriminatory accuracy of such risk models versus improving the preventive effectiveness or reducing the adverse risks of the intervention. Depending on the discriminatory accuracies of the risk models, there may be considerable advantage to modeling the risks of more than one health outcome.

Keywords

absolute risk models; breast cancer; disease prevention; modeling multiple risks; risk-based prevention strategy; risk versus benefit

1. Introduction

Rose[1] distinguished between the ‘population strategy’ of disease prevention and the ‘high-risk strategy.’ An ideal preventive intervention would have negligible side effects and could be applied to the entire population (‘population strategy’), thus achieving maximal preventive impact. Unfortunately, many preventive interventions have adverse effects as well as beneficial preventive effects. For example, tamoxifen reduces the risk of developing breast cancer by about 50% and the risk of hip fracture by 45%, but increases the risk of stroke by about 60%; other life-threatening adverse effects include endometrial cancer and pulmonary embolism[2]. Hence, tamoxifen should only be given to the subset of the population with high enough risks of breast cancer and hip fracture such that the preventive benefits outweigh the risks[3]. This is one motivation for a ‘high-risk strategy’, namely to achieve a net expected benefit for those given the intervention. Recommendations for the preventive use of tamoxifen have been based primarily on breast cancer risk. In particular, the Food and Drug Administration drug label only recommends that tamoxifen be considered if the 5-year breast cancer risk is at least 1.67%. Recommendations for specific age and racial groups[3] took age- and race-specific rates of the various health outcomes

^{*}Correspondence to: Mitchell H. Gail, National Cancer Institute, EPS-8032, Rockville, MD, 20892-7244, USA.. [†]gailm@mail.nih.gov.

into account, but risk models based on individual characteristics other than age and race were not used to model the risks of non-breast cancer outcomes.

In this paper, I investigate whether modeling the risks of more than one health outcome can reduce the expected losses from an intervention that has favorable and unfavorable effects. To simplify the analysis, I consider only two outcomes, breast cancer and stroke, and investigate the extent to which modeling both the risks of breast cancer and of stroke can reduce the expected number of life-threatening events, namely the sum of breast cancers and strokes. The calculations define conditions under which modeling both the risk of breast cancer and of stroke is much more effective than modeling the risk of breast cancer alone, and they give insight into the relative benefits of improving the discriminatory accuracy of the risk models versus improving the preventive effectiveness of the intervention or reducing the adverse risk from the intervention. The relative risks describing tamoxifen intervention effects were obtained from the Breast Cancer Prevention Trial[2]. However, rather than estimate cause-specific absolute risk models from such data, I consider a wide range of hypothetical risk models with varying degrees of discriminatory accuracy and characterized by their joint distribution of breast cancer and stroke risks. The range of discriminatory accuracies investigated covers that found in currently available models as well as models with lower and higher discriminatory accuracies. This approach allows one to determine the circumstances under which modeling more than one risk can reduce expected losses appreciably, even if the currently available models do not have sufficient discriminatory accuracy to capture the full potential of joint risk modeling. Although the numerical examples are motivated by the problem of allocating tamoxifen to women whose risks of breast cancer and of stroke are influenced by the drug, the methods apply to any intervention that has a preventive benefit for one health outcome and a deleterious effect on another health outcome, and the ideas generalize to multiple outcomes.

2. Methods

2.1 Absolute risks and risk distributions

I assume that there are two health outcomes, c_1 and c_2 , that might be influenced by an intervention. The absolute risk $R_1(\mathbf{X})$ is the probability that a person with risk factors $\mathbf{X}$ and free from c_1 at age a will develop c_1 in an interval of duration τ , namely $(a, a + \tau]$, in presence of the competing risks of death from other causes, including death from c_2 . The terms cumulative incidence and crude risk are sometime used to denote absolute risk, which is defined in more mathematical terms in[4]. Note, however, that $R_1(\mathbf{X})$ is not simply the pure cumulative risk from a standard survival analysis because it takes competing risks into account. Because the initial age and duration of the risk interval have a large impact on absolute risk, I also include them as risk factors in $\mathbf{X}$ for notational convenience in what follows. Absolute risk is reduced by competing risks; thus $R_1(\mathbf{X})$ may be less than 1.0, no matter how long the age interval. Similar comments apply to the absolute risk of c_2 , namely $R_2(\mathbf{X})$. Some of the components of $\mathbf{X}$ may affect only $R_1(\mathbf{X})$, others may affect only $R_2(\mathbf{X})$, and others may affect both. Because $\mathbf{X}$ has a distribution in the population, it induces a joint distribution[5] for $\{R_1(\mathbf{X}), R_2(\mathbf{X})\}$, namely $F(r_1, r_2) \approx P(R_1 \leq r_1, R_2 \leq r_2)$. The support of F is the Cartesian product $[0, r_1 \max] \times [0, r_2 \max]$, where $r_1 \max < 1$ and $r_2 \max < 1$. F has marginal distributions $F_1(r_1) \approx P(R_1 \leq r_1)$ and $F_2(r_2) \approx P(R_2 \leq r_2)$. If one is considering a population of individuals with the same initial age and duration of risk interval, the distribution F reflects variation in other components of $\mathbf{X}$ apart from initial age and duration of the risk interval. Otherwise, F reflects variation in age and risk interval duration as well as in other risk factors. The numerical studies for breast cancer and stroke below are based on the assumption that the population consists of 50 year old women who are followed for

ten years. For such women, variation in risk arises from factors apart from age and duration of follow-up.

2.2 Calibration and discriminatory accuracy

The model $R_1(\mathbf{X})$ is well calibrated if its expected value in subsets of individuals with the same value of $(\mathbf{X})$ equals the observed proportion of individuals with c_1 in those subsets [5]. I assume that $R_1(\mathbf{X})$ and $R_2(\mathbf{X})$ are well calibrated. A model has high discriminatory accuracy if its distribution of risk is widely dispersed. One common measure of discriminatory accuracy is the area under a plot of $P(R_1 > t | \text{case})$ versus $1 - F_1(t)$ as the cutpoint t varies. I call this quantity the AUC and note that it can be computed [5] from F_1 , provided the model is well calibrated. This quantity is the probability that a randomly selected case will have an estimated risk greater than that of a randomly selected member of the general population. It is therefore more appropriate for certain public health applications [6, 7] than the area under the receiver operating characteristic curve. Nonetheless, for rare diseases or for common diseases over short time intervals such as five years, these two quantities are numerically nearly equal [5].

2.3 Expected loss with various decision rules for allocating the intervention

I assume there is a loss $w_1 = 0$ whenever the event c_1 occurs and a loss $w_2 = 0$ whenever c_2 occurs. When $w_1 = w_2 = 1$, the expected loss represents the expected number of events c_1 and c_2 . In the numerical studies below, I assume $w_1 = w_2 = 1$, but the formulas are more general. This formulation only includes losses from adverse health outcomes, not monetary costs (see Section 4).

Suppose an intervention decreases R_1 to $rr_1(r_1, r_2)R_1$ with $rr_1(r_1, r_2) \leq 1$. The notation $rr_1(r_1, r_2)$ is used to indicate that the protective relative risk may vary depending on the values $R_1 = r_1$ and $R_2 = r_2$. Likewise, the intervention increases R_2 to $rr_2(r_1, r_2)R_2$ with $rr_2(r_1, r_2) \geq 1$. I assume that risks from having c_1 and c_2 are additive. In the absence of intervention, a person's expected loss is

$$L_{nor} = E(w_1 R_1 + w_2 R_2) = w_1 E(R_1) + w_2 E(R_2) \equiv w_1 \mu_1 + w_2 \mu_2, \quad (1)$$

where $\mu_1 = E(R_1)$ and $\mu_2 = E(R_2)$ can be computed from the marginal distributions of $F(r_1, r_2)$. If everyone receives the intervention, the expected loss is

$$L_{all} = E\{w_1 R_1 rr_1(R_1, R_2) + w_2 R_2 rr_2(R_1, R_2)\}. \quad (2)$$

To minimize the expected loss, one should only give the intervention to a person with risks (r_1, r_2) if $w_1 r_2 + w_2 r_2 \leq w_1 r_1 rr_1(r_1, r_2) + w_2 r_2 rr_2(r_1, r_2)$, or, equivalently, if

$$w_1 r_1 \{1 - rr_1(r_1, r_2)\} + w_2 r_2 \{1 - rr_2(r_1, r_2)\} \geq 0. \quad (3)$$

Define $S = \{(r_1, r_2) : \text{such that equation (3) is satisfied}\}$. The optimal strategy is only to give the intervention if the person has risks in the set S . Under this strategy, the expected loss is

$$Loss = \int_S [w_1 r_1 rr_1(r_1, r_2) + w_2 r_2 rr_2(r_1, r_2)] dF(r_1, r_2) + \int_{S^c} (w_1 r_1 + w_2 r_2) dF(r_1, r_2). \quad (4)$$

If (r_1, r_2) is discrete, then one can calculate (4) readily. If (r_1, r_2) is continuous, integration over the region S can be more difficult. The reduction in loss from using the strategy S is

$$U = \min(L_{nor}, L_{all}) - Loss. \quad (5)$$

The fractional reduction in loss from using the optimal strategy S is

$$FRL = U / \min(Lnorx, Lallrx). \quad (6)$$

Suppose one can pick out all the people destined to develop c_1 perfectly. Suppose the intervention is given only to those people, thereby reducing their risk of c_1 to $rr_1(1, r_2)$. The impact of intervention on the risk of c_2 depends on the joint distribution of risk F . If R_1 is from a perfectly discriminating model, it puts mass $1 - \mu_1$ at 0 and mass μ_1 at 1. Let $\mu_{21} = E(R_2 | R_1 = 1)$ and $\mu_{20} = E(R_2 | R_1 = 0)$. Note that $\mu_2 = \mu_1\mu_{21} + (1 - \mu_1)\mu_{20}$. Thus, the expected loss is

$$\begin{aligned} Lperf &\equiv w_1\mu_1 E\{rr_1(R_1, R_2) | R_1=1\} \\ &\quad + \mu_1 w_2 E\{R_2 rr_2(R_1, R_2) | R_1=1\} \\ &\quad + (1 - \mu_1) w_2 \mu_{20} \\ &= w_1\mu_1 E\{rr_1(R_1, R_2) | R_1=1\} \\ &\quad + w_2\mu_2 + \mu_1 w_2 [E\{R_2 rr_2(R_1, R_2) | R_1=1\} - \mu_{21}]. \end{aligned} \quad (7)$$

Note that

$$Lperf \geq w_1\mu_1 E\{rr_1(R_1, R_2) | R_1=1\} + w_2\mu_2 \equiv \min \text{ loss}. \quad (8)$$

The relative reduction in expected loss[8, 9] is

$$RRL = \{\min(Lnorx, Lallrx) - Loss\} / \{\min(Lnorx, Lallrx) - Lperf\}. \quad (9)$$

2.4 Computation of loss in special cases

Assume $rr_1(r_1, r_2) = rr_1$ and $rr_2(r_1, r_2) = rr_2$ are constants. Then the region S in the (r_1, r_2) plane is defined by a straight line $r_2 = w_1 r_1(1 - rr_1) / w_2(rr_2 - 1)$. Hence

$$\begin{aligned} Loss &= \int_0^1 \int_0^{w_1 r_1(1-rr_1)/w_2(rr_2-1)} (w_1 r_1 rr_1 + w_2 r_2 rr_2) dF(r_2 | r_1) dF(r_1) \\ &\quad + \int_0^1 \int_{w_1 r_1(1-rr_1)/w_2(rr_2-1)}^1 (w_1 r_1 + w_2 r_2) dF(r_2 | r_1) dF(r_1). \end{aligned} \quad (10)$$

If the model for c_1 is perfectly discriminating, the expected loss in (7) reduces to

$$Lperf = w_1\mu_1 rr_1 + w_2\mu_2 + \mu_1 w_2 \mu_{21} (rr_2 - 1), \quad (11)$$

and

$$\min \text{ loss} = w_1\mu_1 rr_1 + w_2\mu_2. \quad (12)$$

A further specialization arises when there is no risk model for c_2 . Then each individual is assigned the same average risk, μ_2 , and F_2 as all its mass at μ_2 . Expression (4) reduces to

$$Loss = \int_{\mu_2 w_2 (rr_2 - 1) / w_1 (1 - rr_1)}^1 (w_1 r_1 rr_1 + w_2 \mu_2 rr_2) dF_1(r_1) + \int_0^{\mu_2 w_2 (rr_2 - 1) / w_1 (1 - rr_1)} (w_1 r_1 + w_2 \mu_2) dF_1(r_1). \quad (13)$$

If the model for c_1 is perfectly discriminating, (13) reduces to

$$Lperf = \mu_1 (w_1 rr_1 + w_2 \mu_2 rr_2) + (1 - \mu_1) w_2 \mu_2 = \mu_1 w_1 rr_1 + \mu_2 w_2 + \mu_1 \mu_2 w_2 (rr_2 - 1), \quad (14)$$

which is only slightly larger than min loss in (12), because $\mu_1\mu_2$ is typically very small. The difference between L_{perf} and min loss is negligible for the purpose of calculating relative reduction in expected loss, RRL , and I use min loss in numerical studies.

Equation (13) is also useful for determining how much losses are increased by failing to model the risk of c_2 . The loss in equation (13) will exceed that from equation (10) if, in fact, the risk of c_2 varies in the population according to the joint risk model in equation (10).

If $w_1 = w_2 = 1$, the losses can be regarded as the sum of the probabilities of c_1 and of c_2 in the time period over which risk is projected. For a population of 100,000 women followed for ten years, the number of such events expected in that time period is 100,000 times that sum of probabilities, and this is the quantity plotted in many figures.

3. Numerical studies

3.1 Numerical methods for simulating absolute risks R_1 and R_2 and computing expected losses

In numerical studies, I assume that the marginal distributions $F_1(r_1) \approx P(R_1 \leq r_1)$ and $F_2(r_2) \approx P(R_2 \leq r_2)$ are rescaled beta distributions. That is, if V_1 has a beta distribution $G(v_1; \alpha_1, \beta_1)$ with mean $\alpha_1 / (\alpha_1 + \beta_1)$, then $R_1 = (r_1 \max)V_1$ has mean $(r_1 \max)\alpha_1 / (\alpha_1 + \beta_1)$, and $F_1(r_1) \approx P(R_1 \leq r_1) = G(r_1 / (r_1 \max); \alpha_1, \beta_1)$. Likewise, $F_2(r_2) \approx P(R_2 \leq r_2) = G(r_2 / (r_2 \max); \alpha_2, \beta_2)$.

If components of $\mathbf{X}$ that affect R_1 are correlated with components of $\mathbf{X}$ that affect R_2 , these components will induce a dependence between R_1 and R_2 . To allow for such dependence in numerical studies, I first generate bivariate normal random variables (W_1, W_2) with means zero, variances 1, and correlation ρ . Applying the inverse standard normal distribution function Φ^{-1} and inverse beta distributions functions, I note that $G_{\alpha_1, \beta_1}^{-1}\{\Phi^{-1}(W_1)\}$ has marginal distribution $G(v_1; \alpha_1, \beta_1)$ and $G_{\alpha_2, \beta_2}^{-1}\{\Phi^{-1}(W_2)\}$ has a marginal distribution $G(v_2; \alpha_2, \beta_2)$. The rescaled variates $R_1 = (r_1 \max)G_{\alpha_1, \beta_1}^{-1}\{\Phi^{-1}(W_1)\}$ and $R_2 = (r_2 \max)G_{\alpha_2, \beta_2}^{-1}\{\Phi^{-1}(W_2)\}$ have the same means as before, but now they are correlated.

Although in principle the expected loss calculations in Section 2 could be carried out with these continuous distributions, the calculations can be difficult for boundaries in equation (3). Therefore, I approximate them by using a discrete distribution which has the same probabilities over 1,000,000 small rectangles as the continuous distribution $F(r_1, r_2)$. In particular I divide the interval $[0, 1]$ into 1000 equal segments for both V_1 and V_2 , thus creating 1,000,000 square sub-regions, each with its own probability. For example, in the first such square, $[0, 0.001] \times [0, 0.001]$, the corresponding probability mass would be placed at the midpoints $V_1 = 0.0005$ and $V_2 = 0.0005$. The quantities V_1 and V_2 are rescaled to produce $R_1 = (r_1 \max)V_1$ and $R_2 = (r_2 \max)V_2$.

3.2 Independent risks, R_1 and R_2 , with rescaled beta distributions

I considered a range of risk models and corresponding joint distributions of risk F , but all based on risk models for breast cancer and stroke[10, 11]. I assumed that R_1 and R_2 were independent rescaled beta variates with respective means (after rescaling) of μ_1 and μ_2 . To reflect mean ten-year invasive breast cancer (c_1) rates in white women aged 50-59, I set $\mu_1 = 10 \times 246.6 \times 10^{-5}$, and the rates for stroke (c_2) were set[3] to $10 \times 110 \times 10^{-5}$. The corresponding scale factors were $r_1 \max = 12.7\mu_1$ and $r_2 \max = 28.1\mu_2$. The value for $r_2 \max$ is the ratio of the highest risk for a 55 year old woman in the Framingham study to average risk[11], $0.844/0.03=28.1$. The largest relative risk for breast cancer for a white woman aged 55 was[10] 24.15, compared to baseline risk. Multiplying this quantity by one minus the attributable risk[10] makes it relative to average risk, and yields $24.15 \times 0.5264 =$

12.7; hence, $r_1 \text{ max} = 12.7\mu_1$. To cover a wide range of risk models with differing discriminatory accuracy but fixed mean absolute risks μ_1 and μ_2 , I considered a range of parameters α and β for the beta distributions. For R_1 , α_1 took the values {0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000}. The corresponding β_1 was chosen so that $\mu_1 = (r_1 \text{ max})\alpha_1 / (\alpha_1 + \beta_1)$. Values of α_2 and β_2 were determined analogously. Small values of α_1 correspond to large AUC. Respective AUC values were {0.958, 0.958, 0.956, 0.919, 0.740, 0.585, 0.527, 0.509} for R_1 and {0.976, 0.975, 0.973, 0.928, 0.745, 0.587, 0.528, 0.509} for R_2 . For Figure 5, I inserted additional AUC1 values 0.650 and 0.850 and additional AUC2 values 0.653 and 0.858.

The expected loss times 10^5 is plotted against AUC1, the AUC for the breast cancer model, for the case that a highly discriminating stroke model (AUC2 = 0.973) is also available (Figure 1). Solid lines correspond to treatment decisions based on both risk models, and dashed lines to treatment decisions based on the breast cancer model alone, assuming that the women have an average ten-year stroke risk, $\mu_2 = 110 \times 10^{-4}$. The ordinate represents the expected number of invasive breast cancers plus strokes in 100,000 women over ten years. The relative risks $rr_1(r_1, r_2) = rr_1$ and $rr_2(r_1, r_2) = rr_2$ were assumed constant for all r_1 and r_2 . This assumption is supported by data[2] for rr_1 . The baseline effects of tamoxifen were $rr_1 = 0.5$ and $rr_2 = 1.6$, marked with open circles (o) in Figure 1. A 20% improvement in breast cancer prevention, $rr_1 = 0.4$ and $rr_2 = 1.6$, is marked with solid circles ($\cdot$), and a 20% reduction in risk of stroke from tamoxifen, $rr_1 = 0.5$ and $rr_2 = 1.28$, is marked with plus signs (+). Figure 1 thus allows one to compare the effects of improvements in the interventions and improvements from using not only a model for breast cancer risk but also a model for stroke risk.

Several points are apparent in Figure 1. First, using the stroke model in addition to the breast cancer risk model reduces losses substantially. For example, for AUC1=0.584, the baseline expected number of life-threatening events (i.e. 100,000 times expected loss) is reduced from 3011 without the stroke model to 2437 with the stroke model, which is a reduction of 574 (19%) life-threatening events (compare dashed line marked with o to solid line marked with o). Second, in the presence of a highly discriminating stroke model, one gets good performance over a wide range of discriminatory accuracies for the breast cancer model, as indicated by the horizontal solid lines. Third, with no stroke model, one assigns the same population average risk of stroke to every woman, which is equivalent to AUC2=0.5; in this case, losses decrease with increasing AUC1, but unrealistically high AUC1 values above 0.9 are needed to achieve the losses that can be achieved when using a highly discriminating stroke model. Fourth, if both the breast cancer and stroke models are used, reducing stroke risk by 20% only reduces the expected loss by $2437 - 2423 = 14$ (0.6%) (solid line marked with +) with AUC1=0.584, because one can avoid giving tamoxifen to those at highest stroke risk with a highly discriminating model. In contrast, in the absence of a stroke model, reducing the stroke risk by 20% reduces life-threatening events from 3011 to 2658 at AUC1=0.584, a reduction of 353 (12%) events (dashed line marked by +). Fifth, reducing breast cancer risk by 20% reduces expected events by $2437 - 2204 = 233$ (9.6%) with AUC1=0.584 with the stroke model (solid lines marked by $\cdot$) and by $3011 - 2757 = 254$ (8.4%) without the stroke model (dashed line marked by $\cdot$). Thus, introduction of a highly discriminating stroke model can lead to greater reductions in life-threatening events than 20% improvements in breast cancer prevention or 20% reductions in the risk of stroke, compared to tamoxifen.

When the model for stroke risk has poor discriminatory accuracy (AUC2 = 0.587), using both risk models hardly reduces expected life-threatening events, compared to using the breast cancer risk model alone and assuming that all women have the same stroke risk (Figure 2). For the baseline tamoxifen parameters (marked by o), with AUC1 = 0.585,

expected events decrease from 2989 to 2980(0.3%) by using both models. For the intervention with $rr_1=0.5$ and $rr_2=1.28$, marked by +, there are hardly any visible improvements from adding the stroke model (Figure 2), and the expected reduction in events at $AUC1 = 0.585$ is $2641-2641=0$ (0.0%). For $AUC1$ values less than 0.76, which is higher than that for currently available models for breast cancer risk[12, 13], decreasing the risk of stroke by 20% reduced life-threatening events more than improving the breast cancer chemopreventive effects of tamoxifen by 20% (Figure 2). Only if $AUC1$ increases to unrealistically high levels, would the improvement in preventive efficacy be preferable to the reduction in stroke risk, because if $AUC1$ is very high, one can restrict the intervention to a small segment of the population at highest breast cancer risk, thereby reducing the number of women at risk of stroke from the intervention.

With a highly discriminating model for stroke ($AUC2 = 0.973$), using both risk models reduces the expected losses (and number of life-threatening events) by about 20% (100 times FRL) both for $rr_1=0.5$ and $rr_2=1.6$ and for $rr_1=0.4$ and $rr_2=1.6$ (Figure 3), compared to the better of the two strategies that don't rely on risk models, namely giving tamoxifen to all women or giving tamoxifen to no women. If $rr_1=0.5$ and $rr_2=1.28$, the risk reduction is about 9%. With a poorly discriminating model for stroke ($AUC2 = 0.587$), the fractional reduction in loss is less than 5% for all three sets of relative risks (Figure 4) for $AUC1 < 0.75$. Only for unrealistically high values of $AUC1$ can FRL approach 0.2. The relative reduction in loss, RRL , compared to a perfect model for selecting women destined to develop breast cancer, is plotted in web appendix Figures A.1 and A.2. If $AUC2=0.973$, using both risk models achieves values RRL above 0.7 for all three intervention scenarios and across a wide range of values of $AUC1$ (Figure A. 1). If $AUC2 = 0.587$, values of RRL above 0.6 are only attained for $AUC1$ above 0.9, which is unlikely to be achieved by foreseeable risk models for breast cancer.

If the model for stroke has intermediate discriminatory accuracy ($AUC=0.745$), the improvements from adding the model for stroke to the model for breast cancer are somewhat less than the improvements from increasing the efficacy or reducing the toxicity from tamoxifen (web appendix Figure A.3).

Figure 5 is a heat-map with color codes for 100,000 times the reduction in expected loss compared to $\min(L_{norx}, L_{allrx})$ for various choices of $AUC1$ and $AUC2$ at baseline values of the intervention parameters, $rr_1=0.5$ and $rr_2=1.6$. This graph indicated that maximal benefit was achieved if either the risk model for breast cancer or the model for stroke had very high discriminatory accuracy (e.g. AUC above 0.95). For $AUC1$ in the range 0.58-0.65, which is realistic for current breast cancer models, joint modeling with a stroke model with $AUC2 = 0.75$, which is plausible currently, reduced expected losses from 141 to 175 events. For $AUC2 = 0.86$, joint modeling reduced expected losses by 280 to 300 events with $AUC1$ in the range 0.58-0.65.

3.3 Correlated risks, R_1 and R_2 , with rescaled marginal beta distributions

Risks of breast cancer and stroke may be correlated because some risk factors, such as age, may influence both risks, and some risk factors for breast cancer may be correlated with other risk factors for stroke. The web appendix Figures A.4 and A.5 are analogous to Figures 1 and 2 for expected loss times 100,000, but R_1 and R_2 have mean correlation 0.152 with values ranging from 0.086 to 0.247, depending on the marginal scaled beta distributions used in Figures A.4 and A.5. These correlations are induced by a correlation 0.25 in the underlying normal variates described in Section 3.1. Comparing Figure A.4 with Figure 1, one finds that the positive correlation had little effect on expected losses when both risk models are used and when the risk model for stroke is highly discriminating ($AUC2=0.973$). However, if only the breast cancer model was used (dashed lines in these

figures), the expected loss with AUC1 near 0.75 was greater in the presence of correlated risks (Figure A.4) than for independent risks (Figure 1) for $rr_1=0.5$ and $rr_2=1.6$ and for $rr_1=0.4$ and $rr_2=1.6$. In contrast, when the stroke model had poor discrimination (AUC2 = 0.587), the expected losses in the presence of positive correlation (Figure A.4) were only very slightly higher than those for independent risks (Figure 2). Moreover, the qualitative conclusions from the analysis of independent risks in Section 3.1 continued to hold in the presence of these moderate positive correlations (Figures A.4 and A.5).

Unreported analyses indicate that if the correlation was 0.5 in the underlying normal model, results were similar to those in Figure A.4 for AUC2 = 0.973, but expected losses increased to about 2300 expected events if both models were used and $rr_1=0.4$ and $rr_2=1.6$. With correlation 0.5, results very similar to Figure A.5 for AUC2 = 0.587 were obtained. The correlations between absolute risks with normal correlation 0.5 averaged 0.323 and ranged from 0.223 to 0.496 over the risk models in Figures A.4 and A.5.

Although it is less plausible that risks would be negatively correlated, correlations -0.25 in the underlying normal model led to an average correlation of risks of -0.080 and range of values from -0.046 to -0.116 across the parameter values used in creating web appendix Figure A.6. For a highly discriminating model for stroke risk with AUC2=0.973, expected losses decreased faster with increasing AUC1 (Figure A.6) than in Figure 1 when only the breast cancer risk model was used (dashed lines). When the stroke model had poor discriminatory accuracy AUC2=0.587, expected losses were hardly changed by negative correlations (compare Figure A.7 with Figure 2).

4. Discussion

Most interventions have beneficial and adverse effects. This fact motivates a “high risk” prevention strategy[1] in which the intervention is given only to those persons at high enough risk of the disease of interest such that the benefits of prevention outweigh the losses from adverse events. In this paper I show that modeling both the risk of the disease to be prevented (breast cancer) and the risk of an adverse health event associated with intervention (stroke) can substantially reduce expected losses, compared to modeling only the risk of the disease to be prevented. In favorable circumstances, modeling both risks has comparable or greater effect on expected losses than improving the intervention's preventive efficacy or adverse risk profile. How much benefit can be gained from joint modeling depends on the discriminatory accuracy of the various component models (Figure 5). Nonetheless, these findings indicate that more research effort is warranted to model the risks of the various health outcomes associated with a preventive intervention and to use these models together to target preventive interventions.

The numerical studies in this paper reflect realistic intervention effects arising in the use of tamoxifen to prevent breast cancer[2]. However, a more realistic analysis would include additional life-threatening health outcomes such as endometrial cancer, pulmonary embolus, and hip fracture[3], some of which are favorably and some unfavorably affected by tamoxifen. Mathematical extension of the methods in this paper to multiple outcomes is straightforward. However, obtaining good models for the several risks and estimating expected losses reliably remain a practical challenge. Ideally, data from randomized intervention trials would be available to estimate the preventive and adverse effects of the intervention, as for tamoxifen[2]. Further work is needed to develop methods of estimation and analysis from empirical studies of the joint distribution of risk factors for the several risk models. Such work would permit not only an estimate of expected loss, but construction of confidence intervals for expected loss. If the available data also included health outcomes, one could use it to test the calibration of the risk models. In any case, it would be important

to assess the robustness of risk-model-based intervention rules and estimated losses to plausible levels of miscalibration.

In the examples in this paper, we chose weights all equal to one. Thus, our expected loss was proportional to the expected number of adverse health events. The weights can also be chosen to represent the relative severity of the various health outcomes[3]. As long as the weights are positive, the intervention rule (3) will not be affected by multiplying or dividing all weights by any positive constant. If one divides by the weight corresponding to the outcome with smallest weight, the other weights become multiples of the loss for that outcome, and units of expected loss can be expressed in that unit. The losses in this paper are based on health outcomes, not costs of intervention. If there are cost constraints, one might modify the intervention criterion (3) to minimize expected health losses in the population subject to cost constraints, as for example, for a single risk model in [14].

It is a challenge to employ preventive interventions with a range of positive and adverse health outcomes. Yet the need to do so is evident from the global monitoring policy adopted for trials in the Women's Health Initiative, for example[15]. Obtaining consensus on how to weigh the various outcomes may be difficult, and obtaining needed data on intervention effects from randomized trials is costly and time-consuming. Moreover, depending on the discriminatory accuracy of the available risk models, the added value of including multiple risk assessments may not justify the added complexity and expense. Nonetheless, examples in this paper indicate that the use of multiple risk models can yield worthwhile reductions in expected loss and are worth exploring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health and Division of Cancer Epidemiology and Genetics of the National Cancer Institute. I thank Dr. Ruth M. Pfeiffer for detecting a previous error in the calculation of min loss.

References

1. Rose, GA. The Strategy of Preventive Medicine. Oxford University Press; Oxford: 1992.
2. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, Vogel V, Robidoux A, Dimitrov N, Atkins J, Daly M, Wieand S, Tan-Chiu E, Ford L, Wolmark N. Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *Journal of the National Cancer Institute*. 1998; 90:1371–1388. [PubMed: 9747868]
3. Gail MH, Costantino JP, Bryant J, Croyle R, Freedman L, Helzlsouer K, Vogel V. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *Journal of the National Cancer Institute*. 1999; 91:1829–1846. [PubMed: 10547390]
4. Gail MH. Personalized estimates of breast cancer risk in clinical practice and public health. *Statistics in Medicine*. 2011; 30:1090–1104. [PubMed: 21337591]
5. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. *Biostatistics*. 2005; 6:227–239. [PubMed: 15772102]
6. Pfeiffer RM, Gail MH. Two Criteria for Evaluating Risk Prediction Models. *Biometrics*. 67:1057–1065. [PubMed: 21155746]
7. Pharoah PDP, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BAJ. Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genetics*. 2002; 31:33–36. [PubMed: 11984562]

8. Baker SG, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. *Journal of the Royal Statistical Society Series A*. 2009; 172:729–748.
9. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *Journal of the National Cancer Institute*. 2009; 101:959–963. [PubMed: 19535781]
10. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*. 1989; 81:1879–1886. [PubMed: 2593165]
11. Wolf PA, Dagostino RB, Belanger AJ, Kannel WB. Probability of stroke - a risk profile from the Framingham Study. *Stroke*. 1991; 22:312–318. [PubMed: 2003301]
12. Chen JB, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute*. 2006; 98:1215–1226. [PubMed: 16954474]
13. Rosner BA, Colditz GA, Webb PM, Hankinson SE. Mathematical models of ovarian cancer incidence. *Epidemiology*. 2005; 16:508–515. [PubMed: 15951669]
14. Gail MH. Applying the Lorenz curve to disease risk to optimize health benefits under cost constraints. *Statistics and Its Interface*. 2009; 2:117–121. [PubMed: 19779595]
15. Freedman L, Anderson G, Kipnis V, Prentice R, Wang CY, Rossouw J, Wittes J, DeMets D. Approaches to monitoring the results of long-term disease prevention trials: Examples from the women's health initiative. *Controlled Clinical Trials*. 1996; 17:509–525. [PubMed: 8974210]

Figure 1.

Plot of 100,000 times expected ten-year loss against the AUC of the breast cancer (health outcome 1) risk model. Solid loci correspond to risk models for breast cancer and for stroke (health outcome 2). Dashed loci correspond to a risk model for breast cancer, but an assumption that all women have the same (average) stroke risk. When used, the stroke risk model has good discriminatory accuracy ($AUC2 = 0.973$). Open circles (o) represent the baseline relative risks from tamoxifen, $rr_1=0.5$ and $rr_2=1.6$. Solid circles (·) represent a 20% improvement in breast cancer prevention efficacy compared to tamoxifen, $rr_1=0.4$ and $rr_2=1.6$. Plus signs (+) represent a 20% reduction in stroke risk from compared to tamoxifen, $rr_1=0.5$ and $rr_2=1.28$. Risks of breast cancer and stroke are assumed to be independent.

Figure 2.

Plot of 100,000 times expected ten-year loss against the AUC of the breast cancer (health outcome 1) risk model. Solid loci correspond to risk models for breast cancer and for stroke (health outcome 2). Dashed loci correspond to a risk model for breast cancer, but an assumption that all women have the same (average) stroke risk. When used, the stroke risk model has poor discriminatory accuracy ($AUC2 = 0.587$). Open circles (o) represent the baseline relative risks from tamoxifen, $rr_1 = 0.5$ and $rr_2 = 1.6$. Solid circles (•) represent a 20% improvement in breast cancer prevention efficacy compared to tamoxifen, $rr_1 = 0.4$ and $rr_2 = 1.6$. Plus signs (+) represent a 20% reduction in stroke risk from compared to tamoxifen, $rr_1 = 0.5$ and $rr_2 = 1.28$. Risks of breast cancer and stroke are assumed to be independent.

Figure 3.

Plot of the fractional reduction in expected loss, FRL, (compared to giving tamoxifen to all women or to no women, whichever yields smaller expected loss) against the AUC of the breast cancer (health outcome 1) risk model. Solid loci correspond to risk models for breast cancer and for stroke (health outcome 2). Dashed loci correspond to a risk model for breast cancer, but an assumption that all women have the same (average) stroke risk. When used, the stroke risk model has good discriminatory accuracy ($AUC2 = 0.973$). Open circles (o) represent the baseline relative risks from tamoxifen, $rr_1=0.5$ and $rr_2=1.6$. Solid circles (•) represent a 20% improvement in breast cancer prevention efficacy compared to tamoxifen, $rr_1=0.4$ and $rr_2=1.6$. Plus signs (+) represent a 20% reduction in stroke risk from compared to tamoxifen, $rr_1=0.5$ and $rr_2=1.28$. Risks of breast cancer and stroke are assumed to be independent.

Figure 4.

Plot of the fractional reduction in expected loss, FRL, (compared to giving tamoxifen to all women or to no women, whichever yields smaller expected loss) against the AUC of the breast cancer (health outcome 1) risk model. Solid loci correspond to risk models for breast cancer and for stroke (health outcome 2). Dashed loci correspond to a risk model for breast cancer, but an assumption that all women have the same (average) stroke risk. When used, the stroke risk model has poor discriminatory accuracy ($AUC2 = 0.587$). Open circles (o) represent the baseline relative risks from tamoxifen, $rr_1 = 0.5$ and $rr_2 = 1.6$. Solid circles (•) represent a 20% improvement in breast cancer prevention efficacy compared to tamoxifen, $rr_1 = 0.4$ and $rr_2 = 1.6$. Plus signs (+) represent a 20% reduction in stroke risk from compared to tamoxifen, $rr_1 = 0.5$ and $rr_2 = 1.28$. Risks of breast cancer and stroke are assumed to be independent.

Figure 5.

Heat-map of 100,000 times the difference between the expected loss for joint risk models with discriminatory accuracies (AUC1, AUC2) and the expected loss with no risk modeling, namely the smaller of the expected losses from intervening on the entire population or not intervening at all. Baseline intervention effects $rr_1=0.5$ and $rr_2=1.6$ were used.