DOI: 10.1093/jnci/djr172

JNCI Journal of the National Cancer Institute Advance Access published June 24, 2011

© The Author 2011. Published by Oxford University Press. All rights reserved.
For Permissions, please e-mail: journals.permissions@oup.com.

ARTICLE

Risk Factor Modification and Projections of Absolute Breast Cancer Risk

Elisabetta Petracci, Adriano Decarli, Catherine Schairer, Ruth M. Pfeiffer, David Pee, Giovanna Masala, Domenico Palli, Mitchell H. Gail

Manuscript received July 22, 2010; revised March 8, 2011; accepted March 9, 2011.

Correspondence to: Elisabetta Petracci, PhD, National Cancer Institute, 6120 Executive Plaza South, EPS 8049, Bethesda, MD 20892-7244 (e-mail: elisabetta.petracci@gmail.com) or Mitchell H. Gail, MD, PhD, National Cancer Institute, 6120 Executive Plaza South, EPS 8032, Bethesda, MD 20892-7244 (e-mail: gailm@mail.nih.gov).

Background Although modifiable risk factors have been included in previous models that estimate or project breast cancer risk, there remains a need to estimate the effects of changes in modifiable risk factors on the absolute risk of breast cancer.

Methods Using data from a case–control study of women in Italy (2569 case patients and 2588 control subjects studied from June 1, 1991, to April 1, 1994) and incidence and mortality data from the Florence Registries, we developed a model to predict the absolute risk of breast cancer that included five non-modifiable risk factors (reproductive characteristics, education, occupational activity, family history, and biopsy history) and three modifiable risk factors (alcohol consumption, leisure physical activity, and body mass index). The model was validated using independent data, and the percent risk reduction was calculated in high-risk subgroups identified by use of the Lorenz curve.

Results The model was reasonably well calibrated (ratio of expected to observed cancers = 1.10, 95% confidence interval [CI] = 0.96 to 1.26), but the discriminatory accuracy was modest. The absolute risk reduction from exposure modifications was nearly proportional to the risk before modifying the risk factors and increased with age and risk projection time span. Mean 20-year reductions in absolute risk among women aged 65 years were 1.6% (95% CI = 0.9% to 2.3%) in the entire population, 3.2% (95% CI = 1.8% to 4.8%) among women with a positive family history of breast cancer, and 4.1% (95% CI = 2.5% to 6.8%) among women who accounted for the highest 10% of the total population risk, as determined from the Lorenz curve.

Conclusions These data give perspective on the potential reductions in absolute breast cancer risk from preventative strategies based on lifestyle changes. Our methods are also useful for calculating sample sizes required for trials to test lifestyle interventions.

J Natl Cancer Inst 2011;103:1–12

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

To assess the potential impact of lowering exposure to risk factors, risk models should include modifiable risk factors. Many breast cancer risk projection models include non-modifiable risk factors such as family history (1–3), mammographic density (4,5), and reproductive and medical factors such as age at menarche and number of breast biopsies (6). Some models also include potentially modifiable risk factors, such as body mass index (BMI) (7) and alcohol consumption (8,9). Boyle et al. (10) developed breast cancer risk models that included dietary information, BMI, alcohol consumption, physical activity, and hormone replacement therapy, in addition to non-modifiable factors. To the best our knowledge, despite these efforts, no systematic estimates of the reduction in absolute breast cancer risk that might result from reducing exposure to modifiable factors have been reported.

Epidemiologists often use attributable risk to estimate the proportion of cancers in the population that can be attributed to an exposure or a set of exposures (11,12). However, the reduction in absolute risk from a prevention strategy can be more important than the relative reduction in risk for counseling individual women before diagnosis and for assessing the potential public health impact of the strategy. Therefore, to determine the potential benefit from modifying behavior and lifestyle factors, we used data from a multicenter case–control study of Italian women with invasive breast cancer to develop a model of absolute breast cancer risk that includes standard non-modifiable as well as modifiable risk factors including BMI, alcohol consumption, and leisure-time physical activity. The model takes competing risks into account to estimate the probability that a woman of a given age and with specific risk factors will develop breast cancer in a defined period. We assessed the calibration and discriminatory accuracy of our model in independent data from the Florence-European Prospective Investigation into Cancer and Nutrition (EPIC)

CONTEXT AND CAVEATS

Prior knowledge

Breast cancer risk models to predict the impact of non-modifiable risk factors and potentially modifiable risk factors (body mass index, physical activity, and alcohol use) on cancer risk have been previously described. However, there are no previous reports quantifying the potential impact of changing modifiable risk factors on absolute breast cancer risk.

Study design

A model of absolute breast cancer risk was developed. The model included standard non-modifiable risk factors as well as modifiable risk factors including body mass index, alcohol consumption, and physical activity. Using independent data from an Italian cohort, the model was used to determine the potential impact of reducing modifiable exposures on absolute breast cancer risk for individuals, the whole population, and population subgroups.

Contribution

The model was well calibrated overall but overestimated the absolute risk of breast cancer in some subgroups, and the discriminatory accuracy was similar to that of other absolute risk models reported in the literature. The absolute risk reduction including exposure modifications was nearly proportional to the risk without including these factors and increased with age and risk projection time.

Implications

The absolute risk model developed in this study could help clinicians make decisions about implementing interventions to reduce a patient's exposure to modifiable risk factors, thereby reducing their absolute risk of breast cancer. Also, the methods used in this study could be used to determine sample sizes needed for clinical trials investigating modifiable risk factor intervention strategies.

Limitations

Data analysis indicated that the model overestimates absolute risk in the highest quintile and that the estimated absolute risk reductions are sensitive to the estimated odds ratios for modifiable risk factors. Also, when applied to individuals, the model's ability to estimate absolute risk reductions was less accurate. Finally, the details of the specific interventions were not given for the participants, so a causal relationship between an intervention and absolute breast cancer risk was undetermined.

From the Editors

cohort (13,14). Using criteria to evaluate reductions in absolute risk that apply to the individual, the entire population, and subgroups of the population, we used the absolute risk model to assess the impact of reducing modifiable exposures on absolute risk.

Methods

We developed a relative risk model and estimates of attributable risk from case–control data. We combined this information with breast cancer incidence rates and mortality rates from Florence Registries to produce a model of absolute breast cancer risk. The model was validated using independent data from the EPIC cohort. We applied the model to the risk factor distributions in EPIC to estimate the effect of reductions in modifiable risk factors on absolute risk in the population. The study populations, mathematical models, and statistical methods are described below.

Study Populations

We estimated absolute breast cancer risks from data from a multicenter case–control study of invasive breast cancer conducted from June 1, 1991, to April 1, 1994, in six Italian regions: Milan, Genoa, the provinces of Pordenone and Gorizia in northern Italy, the provinces of Forlì and Latina in central Italy, and Naples in southern Italy (15). The case–control study included 2569 female case patients with breast cancer, aged 23–74 years (median age: 55 years) who were admitted to the major hospitals in the study areas with histologically confirmed breast cancer that was diagnosed in the year before the interview, and with no history of breast cancer. Women aged 20–74 years (median age: 56 years) without breast cancer and admitted for acute conditions to hospitals in the same catchment areas as the case patients were used as control subjects (n = 2588). Women admitted for gynecologic, hormonal, or neoplastic diseases, or for diseases related to known risk factors of breast cancer were not eligible as control subjects. Control subjects were admitted for trauma (mostly fractures and sprains, 569 subjects, 22%), nontraumatic orthopedic diseases (854 subjects, 33%), surgical conditions (388 subjects, 15%), eye diseases (466 subjects, 18%), or other conditions such as ear, nose, throat, skin, or dental conditions (311 subjects, 12%). The distributions of age and area of residence were similar among case patients and control subjects, although case patients and control subjects were not individually matched. All the 5157 study subjects signed a consent form at enrollment, and 80 (1.6%) case patients and 98 (1.9%) control subjects refused to participate in the study. The interviewers were trained centrally, and the same structured questionnaire and coding manual were used at all study centers. The questionnaire included information on sociodemographic characteristics such as education, occupation, and socioeconomic indicators; lifelong smoking habits; physical activity at work and in leisure-time at selected ages; anthropometric measurements before diagnosis (study subjects were asked to report their weight and height before cancer or interview) and weight at various ages; alcohol and coffee consumption; dietary habits; personal medical history and selected questions regarding family history of cancer; gynecological and reproductive history; and history of use of oral contraceptives, hormone replacement therapy, and female hormone preparations for other indications.

To compute absolute risks, we used 5-year age-specific incidence rates in the age range of 0–84 years for invasive breast cancer from the Florence Cancer Registry collected from January 1, 1989, to December 12, 1993. Rates were based on an estimated population of 1 190 516 residents in the provinces of Florence and Prato in 2006. Estimated age-specific hazard rates from competing mortality from causes other than breast cancer were also obtained from the Florence Cancer Registry. These rates are given in Appendix Table 2.

Independent data from the Florence-EPIC cohort study from 1998 to 2004 were used to assess the validity of our breast cancer absolute risk model. The Florence-EPIC cohort included 10 083 women aged 35–64 years who resided in the Italian provinces of Florence and Prato, which are covered by the Florence Cancer

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

Registry. These women were recruited to the Florence portion of the EPIC-Italy prospective study on diet and cancer (13,14). All participants gave written informed consent at enrollment. The Florence-EPIC project was approved by the local ethics committee, Comitato Etico della Azienda Sanitaria Fiorentina. No overlap occurred between the women enrolled in the case–control study and those recruited for the Florence-EPIC cohort study.

Detailed information on sociodemographic factors, dietary and lifestyle habits, reproductive history, and family history of breast cancer was obtained from a standardized questionnaire, and body measurements were assessed during a physical examination at the study entry (16). Data on current occupational physical activity included employment status and the level of physical activity at work (nonworker, sedentary, standing, manual, heavy manual, and unknown). Information on the frequency and duration of non-occupational physical activity during the past year included housework, home repair, gardening, stair climbing, recreational activities, and vigorous physical activity. Walking, cycling, and sports activities were combined to derive overall recreational activity.

We coded the three modifiable risk factors, alcohol consumption, BMI, and leisure-time exercise, in the Florence-EPIC cohort to obtain categories as consistent as possible with those used to estimate the relative risks from the case–control data. For physical activity at work, we combined standing occupation with nonworker to represent the intermediate exercise category in the case–control study because separate analyses yielded similar odds ratios for standing occupation and nonworker. We also combined manual work and heavy manual work into a high occupational physical activity category.

We excluded 30 women who had prevalent breast cancer at the time of recruitment and 12 women who were diagnosed with incident breast cancer within 6 months after recruitment. We also excluded 1605 (16%) of the 10 031 women who did not have complete covariate data needed for the final model. Follow-up started 6 months after recruitment and continued through December 31, 2004, when the cohort follow-up was last updated. Nineteen women were lost to follow-up.

Statistical Methods

Relative Risk Model. Data from the case–control study were used to select the non-modifiable and modifiable risk factors and to estimate the relative risks and corresponding 95% confidence intervals (CIs) from unconditional multiple logistic regression models (17). We used data from 2523 (98%) of the 2569 case patients and 2504 (97%) of the 2588 control subjects with complete covariate information. The relationship between continuous predictors and the logit of risk was investigated by fitting restricted cubic splines and using a Wald test for linearity (18). Likelihood ratio χ^2 tests for linear trend and for interactions were calculated by comparing models with and without the corresponding parameters and setting *df* equal to the difference in the numbers of parameters. The final model included the following variables with coding for logistic regression given in Table 1: age at menarche (AgeMen), number of previous breast biopsies (NBiops), number of first-degree female relatives with breast cancer (NumRel), age at first live birth (Age1st), body mass index for women aged 50 years and older (Bmi), body mass index for women younger than age 50 years (InvBmi), alcohol consumption in three categories (never, current, and former for women who stopped drinking at least 1 year before the interview), occupational physical activity at ages 30–39 years (OccAct), leisure-time physical activity at ages 30–39 years (LeiAct), educational level (Educat), and age at interview (age) and age^2. Because BMI was inversely associated with breast cancer risk in women aged less than 50 years and positively associated in older women, Bmi and InvBmi were not included as main effects but only through the products InvBmi × AgeLT50 and Bmi × AgeGE50, in which AgeLT50 is an indicator that takes the value 1 if age is less than 50 years and 0 otherwise, and AgeGE50 is an indicator that takes the value 1 if age is 50 years or older. We coded all covariates to yield positive relative risk estimates compared with the reference level to facilitate the calculations of attributable risk and the baseline hazard. To select variables for inclusion in our relative risk model, the association between the risk of breast cancer and other risk factors such as oral contraceptives, hormone replacement therapy, age at menopause, marital status, and parity was also assessed. These factors were not included in the final model because the associations were weak or non-statistically significant (data not shown).

Absolute Risk Model. The absolute risk, $r(a, \tau, x)$, for a woman of age a and with risk factors x to a subsequent age τ was obtained from the formula for a piecewise constant hazard model on each year [see Equation 6 in Gail et al. (19)]:

$$r(a,\tau,x) = \sum_{j=a}^{a+\tau-1} \{h_{1j} rr_j/(h_{1j} rr_j + h_{2j})\} \times \{1 - \exp[-(h_{1j} rr_j + h_{2j})]\} \times \exp\left[-\sum_{l=a}^{j-1}(h_{1l} rr_l + h_{2l})\right] \qquad [1]$$

In this formula, h_{1j} is the baseline cause-specific hazard for breast cancer for a woman of age j (in years) with risk factors $x = 0$; h_{2j} is the hazard from non-breast cancer causes of mortality at age j and rr_j is the relative risk at age j for a woman with risk factors x compared with a woman with $x = 0$. By convention, the last factor in Equation 1 equals 1 if $j - 1 < a$. We estimated h_{1j} by multiplying the invasive breast cancer age-specific incidence rates from the Florence Cancer Registry, h^*_{1j}, by 1 − [the estimated age-specific attributable risks], as described in Gail et al. (19). The age-specific attributable risks were obtained from the distribution of risk factors in case patients and were separately obtained for women younger than 50 years of age and for women aged 50 years and older from the formula in Bruzzi et al. (20). The mortality rates h_{2j} were also obtained from the Florence Cancer Registry. Because rates were constant for 5-year intervals in the Florence Cancer Registry, we used this same value for each year j on the 5-year interval.

Confidence intervals for the projected probabilities were obtained from a nonparametric bootstrap (21) with 1000 bootstrap replications. We assumed that h^*_{1j} and h_{2j} were known without error, so that all variability in Equation 1 arose from uncertainty in relative and attributable risks. Each bootstrap sample was drawn with replacement from the case patients and separately from the control subjects in the case–control study, with the original number of case patients and control subjects in each replication. For each bootstrap replication, we applied the logistic regression

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

Table 1. Distribution of case patients (n = 2523) and control subjects (n = 2504) in the Florence-European Prospective Investigation into Cancer and Nutrition study, and odds ratios of breast cancer risk*

Risk factor category†	Code	No. of case patients (%), n = 2523	No. of control subjects (%), n = 2504	OR (95% CI)
Age at menarche, y				
≥14	0	837 (33.1)	929 (37.1)	1.0 (referent)
12–13	1	1200 (47.6)	1105 (44.1)	1.04 (0.96 to 1.13)
7–11	2	486 (19.3)	470 (18.8)	1.09 (1.00 to 1.19)
Age at first live birth, y				
<20	0	118 (4.7)	209 (8.3)	1.0 (referent)
20–24	1	775 (30.7)	949 (37.9)	1.30 (1.21 to 1.40)
25–29	2	1221 (48.4)	1051 (42.0)	1.69 (1.46 to 1.96)
≥30	3	409 (16.2)	295 (11.8)	2.20 (1.76 to 2.75)
No. of affected first-degree relatives				
0	0	2268 (89.9)	2387 (95.3)	1.0 (referent)
≥1	1	255 (10.1)	117 (4.7)	2.35 (1.86 to 2.96)
No. of biopsies				
0	0	2480 (98.3)	2474 (98.8)	1.0 (referent)
≥1	1	43 (1.7)	30 (1.2)	1.32 (0.81 to 2.14)
Occupational physical activity level				
High	0	373 (14.8)	455 (18.2)	1.0 (referent)
Intermediate	1	1882 (74.6)	1881 (75.1)	1.10 (0.97 to 1.24)
Low	2	268 (10.6)	168 (6.7)	1.21 (0.95 to 1.54)
Education, y				
<7	0	1265 (50.1)	1580 (63.1)	1.0 (referent)
7–12	1	700 (27.7)	606 (24.2)	1.37 (1.26 to 1.49)
≥12	2	558 (22.1)	318 (12.7)	1.88 (1.59 to 2.23)
Alcohol drinking habits				
Never drinker		748 (29.6)	860 (34.3)	1.0 (referent)
Current drinker		1632 (64.7)	1494 (59.7)	1.27 (1.12 to 1.43)
Former drinker		143 (5.7)	150 (6.0)	1.23 (0.95 to 1.59)
BMI at age <50 y, kg/m^2				
≥30.0	0	63 (7.9)	93 (13.4)	1.0 (referent)
25.0–29.9	1	182 (22.7)	181 (26.1)	1.26 (1.08 to 1.48)
<25.0	2	557 (69.4)	419 (60.5)	1.60 (1.16 to 2.20)
BMI at age ≥50 y, kg/m^2				
<25.0	0	799 (46.4)	868 (47.9)	1.0 (referent)
25.0–29.9	1	639 (37.1)	652 (36.0)	1.13 (1.03 to 1.24)
≥30.0	2	283 (16.4)	291 (16.1)	1.28 (1.06 to 1.54)
Leisure-time physical activity, h/wk				
≥2	0	774 (30.7)	816 (32.6)	1.0 (referent)
<2	1	1749 (69.3)	1688 (67.4)	1.08 (0.96 to 1.22)

* BMI = body mass index, CI = confidence interval, OR = odds ratio.

† The variables used in Equation 3 have the following equivalences to the variables in this table: AgeMen = age at menarche, Age1st = age at first live birth, NumRel = number of affected first-degree relatives, NBiops = number of biopsies, OccAct = occupational physical activity, Educat = education, CurrnDrnk = current drinker, ExDrnk = former drinker, InvBmi = body mass index at age <50 years (reference category = "≥ 30.0"), Bmi = body mass index at age ≥50 years (reference category = "< 25.0"), LeiAct = leisure-time physical activity.

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

model to obtain new relative and attributable risk estimates. By saving 1000 such sets of these quantities, we could compute 1000 estimates of absolute risk and obtain 95% confidence intervals as the 2.5th and the 97.5th percentiles of the bootstrap distribution. Bootstrap confidence intervals on other quantities, such as absolute risk reductions, were likewise based on the stored sets of relative and attributable risks.

Validation. To test calibration, we computed the expected number of invasive breast cancers (*E*) and compared them with the corresponding observed number (*O*) in the Florence-EPIC cohort. For each woman, the projected probability of breast cancer was obtained from the age enrollment (initial age) to the final age. The final age was defined as the younger of either the age at lost to follow-up for women who left the Prato region or the age on December 31, 2004. Follow-up did not end on the date of breast cancer incidence or death, because these events are already accounted for in Equation 1. Calibration was evaluated overall and for subgroups of women, defined in terms of either risk factor levels or quintiles of the distribution of the expected absolute risks in the total population. The χ^2 test of goodness of fit, based on the squared Pearson residuals $(O–E)^2/E$, and the sum of this quantity over the risk factor categories or probability levels (with *df* equal to the number of mutually exclusive and exhaustive categories for each factor) were also calculated. We used the concordance statistic *c*, which is the area under the receiver operating characteristic curve, to measure the discriminatory accuracy of the model (22). We calculated *c* from the independent Florence-EPIC cohort data with the SAS macro at http://support.sas.com/dsearch?ct=&qt=macro+%25roc&col=suppprd&nh=10&qp=&qc=suppsas&ws=1&qm=1&st=1&lk=1&rf=0&oq=&rq=0. This macro provides an SE based on calculations for the Mann–Whitney statistic adapted for ties. All 1 *df* statistical tests were two-sided. *P* values less than .05 were considered statistically significant.

Criteria for Assessing Effects of Modifying Risk Factors

Definition of Risk Factor Modifications. We distinguished plausible lifestyle modifications for an individual woman from those of the general population. For an individual woman, we changed the exposure for current drinkers to that of former drinkers; no changes were made for former drinkers or for never drinkers. We set BMI to less than 25 kg/m^2 for women aged 50 years and older; for women aged 49 years and younger, for whom breast cancer risk varied inversely with BMI, we did not modify BMI. We did not modify levels of leisure activity for women older than age 39 years, because the leisure activity variable was defined for women in the age range 30–39 years. For women in their 30's, the variable was set to exercising at least 2 hours per week. Our methods allow the evaluation of intermediate changes, as for example reducing BMI from greater than 30 kg/m^2 to the overweight level (ie, 25.0 kg/m^2 ≤ BMI ≤ 29.9 kg/m^2), but we did not include such intermediate values in this study.

For the population, it is possible to imagine additional modifications. For example, one might hope to increase the proportion of the population that exercised at least 2 hours per week when aged 30–39 years. Likewise, one could hope to increase the proportion of "never drinkers" in a population. Thus, for lifestyle modifications at the population level, we performed calculations on the optimistic assumptions that all current or former drinkers became never drinkers, all women who exercised less than 2 hours per week began exercising at least 2 hours per week, and all women aged 50 years and older maintained a BMI less than 25 kg/m^2. These optimistic modifications give an upper boundary of risk reductions that could be achieved by lifestyle changes in the population, in our model.

Criteria for Individual Counselees and for the Population

We studied the reduction in the absolute risk projections from lowering or eliminating exposure to modifiable risk factors, whereas keeping the values of non-modifiable risk factors unchanged. The effects of modifying risk factors on absolute risk projections in individual women were evaluated by means of the absolute risk reduction, $d_{(X_1,X_2)} = \{r_{(X_1,X_2)} - r_{(X_1,X_{20})}\} \times 100$. The quantities $r_{(X_1,X_2)}$ and $r_{(X_1,X_{20})}$ represent the non-modified and modified absolute risks obtained by setting the values for the modifiable risk factors, X_2, to their modified levels, X_{20}. Vector X_1 denotes the non-modifiable factors in the model. We computed 95% bootstrap confidence intervals as previously described.

To evaluate the effects of risk modification at the population level, we averaged the risk reduction and fractional risk reduction over the entire population of women or within high-risk subgroups. High-risk subgroups were defined by particular risk factors and by using the Lorenz curve (23) to identify risk factor combinations that conferred high risk and accounted for a given percentage of the total population risk. We calculated the mean risk reduction for a specific subset from the formula:

$$\overline{d}^{S}_{(X_1,X_2)} = E\{d_{(X_1,X_2)} | (X_1,X_2) \in S\} = \frac{\int\limits_{X_1,X_2} \{r_{(X_1,X_2)} - r_{(X_1,X_{20})}\} I_{\{(X_1,X_2)\in S\}} dF_{(X_1,X_2)}}{\int\limits_{X_1,X_2} I_{\{(X_1,X_2)\in S\}} dF_{(X_1,X_2)}} \qquad [2]$$

where $I_{\{(X_1,X_2)\in S\}}$ is an indicator function that takes the value 1 if $(X_1,X_2) \in S$ and 0 otherwise. $F_{(X_1,X_2)}$ denotes the joint distribution of the covariates X_1 and X_2 estimated from the Florence-EPIC cohort. In Equation 2, integration is over all the values of X_1 and X_2, but each X_2 value is changed to X_{20} in computing $r(X_1,X_{20})$. We estimated different joint distributions for women younger than age 50 years and women aged 50 years and older. The fractional risk reduction is obtained by dividing the quantity in Equation 2 by the mean non-modified absolute risk, $[\int\limits_{X_1,X_2} r_{(X_1,X_2)} I_{\{(X_1,X_2)\in S\}} dF_{(X_1,X_2)}][\int\limits_{X_1,X_2} I_{\{(X_1,X_2)\in S\}} dF_{X_1,X_2}]^{-1}$ and multiplying by 100. Equation 2 simplifies when S corresponds to the entire space of X values, namely the whole population. Then the denominator is equal to $\int\limits_{X_1,X_2} dF_{(X_1,X_2)} = 1$ and Equation 2 reduces to $\overline{d}_{(X_1,X_2)} = \int\limits_{(X_1,X_2)} \{r_{(X_1,X_2)} - r_{(X_1,X_{20})}\} dF_{(X_1,X_2)}$.

Definition of Subgroups in the Population. The easiest way to define subgroups is through combinations of non-modifiable risk factors, namely $S = \bigcup_{x_1} \{(X_1,X_2) : X_1 = x_1\}$, where x_1 ranges over a set of fixed values. Another way to define high-risk groups is by means of the Lorenz curve (23), which describes the proportion of ranked population risk that is possessed by the given proportion q of population with lowest risk. Assume there are K mutually exclusive and exhaustive combinations of levels of risk factors X_1 and X_2. Not all of these K patterns have different risks. Suppose there are only J unique absolute risks r_j with probabilities P_j (with $J \le K$) estimated from the Florence-EPIC cohort data. The probabilities P_j can be estimated by summing over the corresponding probabilities of the contributing risk factor combinations. We ordered the J absolute risks r_j from smallest to largest and reordered the P_j accordingly so that P_1 is the probability of the smallest absolute risk and P_J is the probability of the largest absolute risk. Denoting by $i(q)$ the smallest integer such that $\sum_{j=1}^{i(q)} P_j \ge q$, we computed the Lorenz curve by means of the formula: $L\{i(q)\} = \frac{1}{\mu}\sum_{j=1}^{i(q)} r_j P_j$, where $\mu = \sum_{j=1}^{J} r_j P_j$ is the mean absolute risk in the population. We determine the unique absolute risks r_j that account for the top 10% of the total population absolute risk by determining the largest j^* such that $\sum_{j=j^*}^{J} P_j \ge p_{10}$, where $p_{10} = 1 - L^{-1}(0.9)$. Because $L^{-1}(0.9)$, the inverse of the Lorenz curve, defines the proportion of the population at lowest absolute risk that contains 90% of the total population absolute risk, the quantity p_{10} is the proportion of the population at the highest absolute risk that contains 10% of the population absolute risk. For each $j = j^*, j^* + 1, \ldots, J$, we find all risk factors combinations $S_j = \{(X_1,X_2) : r(X_1,X_2) = r_j\}$. Then, the subset of women at the highest absolute risk who together account for 10% of total population absolute risk is given by $S_{10} = \bigcup_{j=j^*}^{J} S_j$. We similarly defined high–absolute

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

risk subgroups S_{20}, S_{40}, S_{60}, and S_{80} that accounted for 20%, 40%, 60%, and 80%, respectively, of the total population absolute risk.

Results

Relative Risk Model

Risk factor codes and distributions, and breast cancer relative risk estimates with 95% confidence intervals are given in Table 1. We used multivariable logistic regression that included all risk factors listed in Table 1 as well as age at interview and age^2. Unreported spline models for continuous modifiable exposures did not fit the data statistically significantly better than the simpler categorical coding in Table 1 (data not shown). We found that increased relative risk of breast cancer was associated with alcohol consumption compared with never drinkers, but no dose–response relationship was observed among current drinkers. Thus, we categorized alcohol consumption into never, current, and former drinkers. We found a statistically significant interaction ($P < .001$), which indicated that increased BMI was protective for women younger than age 50 years but increased breast cancer risk above that age. Thus, we included the terms InvBmi × AgeLT50 ($P = .004$) and BMI × AgeGE50 ($P = .009$) in the relative risk model in Equation 3. Our unreported data indicated no other important interactions between age and other risk factors or among the other risk factors. We included previous breast biopsies in the relative risk model despite the fact that its association with invasive breast cancer was non-statistically significant in our data, because it has been associated with breast cancer risk in many other studies (24). The final relative risk model is given by:

$$\begin{aligned}\log(\text{odds}) = {} & -5.3437 + 0.0411 \times \text{AgeMen} + 0.8531 \times \text{NumRel} \\ & + 0.2627 \times \text{Age1st} + 0.2759 \times \text{NBiops} \\ & + 0.2360 \times \text{CurrnDrnk} + 0.2041 \times \text{ExDrnk} \\ & + 0.3157 \times \text{Educat} + 0.0783 \times \text{LeiAct} \\ & + 0.0946 \times \text{OccAct} + 0.2350 \times \text{InvBmi} \times \text{AgeLT50} \\ & + 0.1247 \times \text{Bmi} \times \text{AgeGE50} + 0.1525 \times \text{age} - 0.0013 \times \text{age}^2\end{aligned} \qquad [3]$$

The variance–covariance matrix for the coefficients in Equation 3 is described in Appendix Table 1.

Absolute Risk

We used the rates in Appendix Table 2, relative risks from Equation 3, and estimates of attributable risk to estimate the absolute risk of breast cancer from Equation 1. The age-specific attributable breast cancer risk estimates were 0.77 (95% CI = 0.69 to 0.83), derived from the 802 case patients aged less than 50 years, and 0.66 (95% CI = 0.58 to 0.72), derived from the 1721 case patients aged 50 years or more.

We show absolute 10- and 20-year breast cancer absolute risks for four risk counselee profiles and for initial ages 45 and 65 years (Table 2). A high-risk counselee profile (designated as a) is distinguished from a low-risk counselee profile (designated as b). The latter corresponds to decreased alcohol consumption, increased leisure-time exercise, or reduced obesity in older women, with other factors unchanged. Counselee profile 3a corresponds to the highest absolute risk, because the age is 65 years and all non-modifiable risk factors are at their highest levels. Consider a highly educated 45-year-old woman (counselee profile 1a) with an

Table 2. Examples of 10- and 20-year non-modified and modified absolute risk estimates of breast cancer for women of different ages with different risk factor profiles*

	Counselee profile No.‡							
Risk factor category†	**1a**	**1b**	**2a**	**2b**	**3a**	**3b**	**4a**	**4b**
Age, y	45	45	45	45	65	65	65	65
Age at menarche, y	7–11	7–11	7–11	7–11	7–11	7–11	7–11	7–11
No. of affected first-degree relatives	≥1	≥1	≥1	≥1	≥1	≥1	≥1	≥1
No. of biopsies	≥1	≥1	≥1	≥1	≥1	≥1	≥1	≥1
Age at first live birth, y	≥30	≥30	≥30	≥30	≥30	≥30	≥30	≥30
Occupational physical activity level	Low	Low	Low	Low	Low	Low	Low	Low
Education, y	≥12	≥12	≥12	≥12	≥12	≥12	≥12	≥12
BMI, kg/m^2	<25	<25	25.0–29.9	25.0–29.9	≥30	<25	≥30	<25
Current drinker	Yes	No	Yes	No	Yes	No	No	No
Leisure activity, h/wk	<2	<2	≥2	≥2	<2	<2	≥2	≥2
10-y risk, % (95% CI)	15.0 (9.4 to 24.9)	14.6 (8.4 to 25.0)	13.5 (8.3 to 22.7)	13.2 (7.6 to 23.2)	22.9 (14.0 to 37.0)	17.8 (10.4 to 30.2)	17.3 (10.7 to 28.8)	13.8 (8.5 to 23.2)
20-y risk, % (95% CI)	28.4 (18.2 to 44.4)	27.7 (16.5 to 44.7)	27.7 (17.6 to 43.5)	27.0 (16.3 to 44.3)	37.5 (24.1 to 56.2)	30.0 (18.2 to 47.6)	29.3 (18.7 to 45.8)	23.7 (15.0 to 38.0)

* BMI = body mass index, CI = confidence interval.

† The variables used in Equation 3 have the following equivalences to the variables in this table: AgeMen = age at menarche, Age1st = age at first live birth, NumRel = number of affected first-degree relatives, NBiops = number of biopsies, OccAct = occupational physical activity, Educat = education, CurrnDrnk = current drinker, LeiAct = leisure-time physical activity.

‡ Each counselee profile number is identified as a high-risk profile (a) and a low-risk profile (b) obtained by modifying the exposure to alcohol consumption and the BMI.

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

unfavorable profile of regarding both non-modifiable and modifiable risk factors. This woman started menstruating before age 11, had a mother with breast cancer, had a previous breast biopsy with benign histology, and had a first live birth at age 32 years. Her job was sedentary, her leisure activity lasted less than 2 hours each week, she was a current user of alcohol, and her BMI was less than 25 kg/m^2. On the basis of these data, her 10- and 20-year projected absolute risks of breast cancer are 15.0% (95% CI = 9.35% to 24.9%) and 28.4% (95% CI = 18.2% to 44.4%), respectively.

Validation of the Model in Independent Data From the Florence-EPIC Study

Data on the distribution of risk factors, women-years of follow-up, and the number of incident breast cancers in the Florence-EPIC study cohort are in the Supplementary Data (available online). A total of 252 incident breast cancers were diagnosed among 10 031 women included in the analyses. However, the validation study was based on the 8426 women, including 206 with incident invasive breast cancer, who had complete data for all risk factors in the model.

Overall, the model predicted 225.7 invasive breast cancers, whereas 206 invasive breast cancers were observed, resulting a ratio of expected to observed (*E/O*) breast cancers of 1.10 (95% CI = 0.96 to 1.26, $P = .190$) (Table 3). Numbers of breast cancers were overestimated by the model for women aged 60 years or more, women aged 30 or more at first live birth, and women with 12 years or more of education. For the other variables, the *E/O* ratios were non-statistically significantly different from unity. Estimated absolute risks of breast cancer were divided into quintiles, and the sum of the absolute risks in each quintile (*E*) was compared with the observed invasive breast cancers (*O*) (Table 3). Except for the highest quintile, the differences between expected and observed counts were non-statistically significant. The concordance statistic (discriminatory power) was 0.62 (95% CI = 0.555 to 0.689) for women younger than age 50 years, and 0.57 (95% CI = 0.519 to 0.614) for women aged 50 years and older.

Absolute Risk Reduction for Individual Counselees

The 20-year estimated absolute risk for counselee profile 1a (Table 2) decreased 0.7% when current and former drinkers were compared (current vs former drinkers, absolute risk = 28.4% vs 27.7%, difference = 0.7%, 95% CI = −5.86% to 7.17%). The 20-year absolute risk for a 65-year-old woman with a BMI greater than 30 kg/m^2 (counselee profile 4a) was 29.3%, compared with 23.7% for an otherwise identical counselee with a BMI less than 25 kg/m^2. The estimated absolute risk reduction was 5.6% (95% CI = −18.5% to 22.5%). The 95% confidence intervals of projected absolute risk reductions for individual counselees are much wider than for populations.

Risk Reductions in the Entire EPIC Population and in High-Risk Subgroups

We present population-averaged estimates of 10- and 20-year mean absolute risks of breast cancer before modifying any risk factors, together with the average absolute risk reduction and the fractional reduction in average absolute risks from reducing modifiable risk factors to their lowest levels (Table 4). The 10- and 20-year non-modified mean absolute risks for 45-year-old women in the entire population were 3.1% (95% CI = 2.8% to 3.6%) and 6.5% (95% CI = 5.8% to 7.4%), respectively; for 65-year-old women, the 10-year non-modified mean absolute risk was 3.6% (95% CI = 3.3% to 4.1%) and the 20-year non-modified mean absolute risk was 6.5% (95% CI = 6.0% to 7.4%). Despite the fact that 65-year-old women had higher breast cancer incidence rates than 45-year-old women (Appendix Table 2), the average absolute risks were only slightly higher in the 65-year-old women, because the older women had a more favorable distribution of risk factors. The corresponding absolute mean risk reductions were 0.6% (95% CI = 0.3% to 1.0%) and 1.4% (95% CI = 0.7% to 2.0%) for 45-year-old women and 0.9% (95% CI = 0.5% to 1.3%) and 1.6% (95% CI = 0.9% to 2.3%) for 65-year-old women.

The estimated mean absolute risks and mean absolute risk reductions were larger in women with a family history of breast cancer (Table 4). They were also larger in women at the highest absolute risk who account for 10% of the total population absolute risk (Table 4). Among these high-risk women aged 65 years, the 10- and 20-year non-modified mean absolute risks were almost three times higher than in the general population (10-year non-modified mean absolute risk = 10.7%, 95% CI = 8.4% to 14.0% and 20-year non-modified mean absolute risk = 18.6%, 95% CI = 14.9% to 24.0%). The mean 10- and 20-year absolute risk reductions for these women were 2.5% (95% CI = 1.5% to 4.1%), and 4.1% (95% CI = 2.5% to 6.8%), respectively. For 65-year-old women with a family history of breast cancer, the 10- and 20-year mean absolute risk reductions were 1.9% (95% CI = 1.1% to 2.9%) and 3.2% (95% CI = 1.8% to 4.8%), respectively.

In contrast to the absolute mean reductions, the fractional risk reductions do not depend strongly on the number of years of follow-up, age, or presence of strong risk factors (Table 4). These fractional risk reductions are about 20% for women younger than 50 years and approximately 24% for women aged 50 years and older (Table 4). Fractional risk reductions are larger in older women because weight reduction is not a recommended intervention for younger women, for whom increased BMI is associated with lower risk.

Mean absolute risk reduction (ordinate) is directly proportional to mean unmodified absolute risk (abscissa) for women aged 45 years, both in the general population and in high-risk subgroups, over 5-, 10-, 20-, and 30-year projection intervals (Figure 1). The unmodified absolute risk and the absolute risk reductions tend to be greater in women with a positive family history of breast cancer and in the highest absolute risk group that accounts for 10% of population absolute risk compared with the general population. However, there is overlap if the projection intervals differ. For example, the 10- and 20-year mean reductions in the whole population are nearly the same as the 5- and 10-year reductions for women with positive family history of breast cancer.

To see how the average absolute risk reduction varied as the high-risk subgroup based on the Lorenz curve was relaxed progressively to include women with lower risk, we considered subgroups accounting for 10%, 20%, 40%, 60%, 80%, and 100% of the total population absolute risk (Figure 2). Mean reductions in 10-year absolute risk are shown for women of 45 and 55 years, together with vertical lines to indicate 95% confidence intervals. As the high-risk groups were expanded to include progressively

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

Table 3. Expected and observed breast cancers in the Florence-European Prospective Investigation into Cancer and Nutrition cohort (N = 10 031) by risk factor categories and by quintiles of projected absolute risk*

Risk factor category†	*O*	*E*	*E/O* (95% CI)	Goodness of fit‡	*P*
Overall	206	225.7	1.10 (0.96 to 1.26)	$\chi^2_{(1)} = 1.72$	.19
Age at recruitment, y				$\chi^2_{(4)} = 9.77$	.04
30–39	6	10.97	1.82 (0.82 to 4.05)		
40–49	64	53.82	0.84 (0.66 to 1.07)		
50–59	105	115.30	1.10 (0.91 to 1.33)		
≥60	31	45.60	1.47 (1.03 to 2.09)		
Age at menarche, y				$\chi^2_{(3)} = 1.80$	.61
≥14	44	49.59	1.13 (0.84 to 1.52)		
12–13	109	117.08	1.07 (0.89 to 1.29)		
<12	53	59.03	1.11 (0.85 to 1.45)		
Age at first live birth, y				$\chi^2_{(4)} = 16.10$	.003
<20	1	3.04	3.04 (0.43 to 21.58)		
20–24	56	43.40	0.78 (0.60 to 1.01)		
25–29	113	117.24	1.04 (0.86 to 1.25)		
≥30	36	62.02	1.72 (1.24 to 2.38)		
No. of affected first-degree relatives				$\chi^2_{(2)} = 1.87$	.39
0	176	190.54	1.08 (0.93 to 1.25)		
≥1	30	35.17	1.17 (0.81 to 1.67)		
No. of biopsies				$\chi^2_{(2)} = 3.59$	.17
0	195	217.79	1.12 (0.97 to 1.29)		
≥1	11	7.91	0.71 (0.39 to 1.28)		
Occupational physical activity level				$\chi^2_{(3)} = 3.21$	.36
High	6	10.15	1.69 (0.76 to 3.76)		
Medium	119	130.59	1.10 (0.92 to 1.32)		
Low	81	84.97	1.05 (0.84 to 1.31)		
Education, y				$\chi^2_{(3)} = 9.83$	.02
<7	48	39.36	0.82 (0.62 to 1.08)		
7–12	67	64.44	0.96 (0.76 to 1.22)		
≥12	91	121.90	1.34 (1.09 to 1.65)		
Alcohol drinking habits				$\chi^2_{(3)} = 2.96$	.40
Never drinker	30	30.37	1.01 (0.70 to 1.44)		
Current drinker	171	186.51	1.09 (0.94 to 1.27)		
Ex-drinker	5	8.83	1.77 (0.74 to 4.25)		
Leisure-time physical activity, h/wk				$\chi^2_{(2)} = 3.66$	.16
≥2	165	189.23	1.15 (0.99 to 1.34)		
<2	41	36.48	0.89 (0.66 to 1.21)		
BMI at age <50 y, kg/m^2				$\chi^2_{(3)} = 1.14$	.77
<25.0	55	47.82	0.87 (0.67 to 1.13)		
25.0–29.9	13	13.62	1.05 (0.61 to 1.81)		
≥30	3	3.34	1.11 (0.36 to 3.44)		
BMI at age ≥50 y, kg/m^2				$\chi^2_{(3)} = 8.49$	.04
<25.0	67	73.37	1.10 (0.87 to 1.40)		
25.0–29.9	48	62.41	1.30 (0.98 to 1.73)		
≥30	21	25.13	1.20 (0.78 to 1.84)		
Quintile of risk (range), %§				$\chi^2_{(5)} = 15.13$	.03
1 (0–1.57)	26	20.64	0.79 (0.54 to 1.16)		
2 (1.57–2.10)	31	31.00	1.00 (0.70 to 1.42)		
3 (2.10–2.69)	43	40.03	0.93 (0.69 to 1.25)		
4 (2.69–3.53)	52	51.69	0.99 (0.75 to 1.30)		
5 (≥3.53)	54	82.34	1.52 (1.16 to 1.98)		

* BMI = body mass index, CI = confidence interval, E = expected, O = observed.

† The variables used in Equation 3 have the following equivalences to the variables in this table: AgeMen = age at menarche, Age1st = age at first live birth, NumRel = number of affected first-degree relatives, NBiops = number of biopsies, OccAct = occupational physical activity, Educat = education, CurrnDrnk = current drinker, LeiAct = leisure-time physical activity.

‡ The χ^2 test of goodness of fit with *df* as shown was used to compare the observed and expected number of cases.

§ Ranges show the cutoff values for risk in percent for quintiles of risk.

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

Table 4. Estimated 10- and 20-year non-modified mean absolute risks, mean absolute risk reductions and fractional reductions in mean risk for women of different initial ages in the whole population and in high-risk subgroups of the population*

	Non-modified mean absolute risk		Mean absolute risk reduction†		Fractional reduction in mean risk	
Population and population subsets investigated	**10 y, % (95% CI)**	**20 y, % (95% CI)**	**10 y, % (95% CI)**	**20 y, % (95% CI)**	**10 y, % (95% CI)**	**20 y, % (95% CI)**
Age of entire population, y						
45	3.1 (2.8 to 3.6)	6.5 (5.8 to 7.4)	0.6 (0.3 to 1.0)	1.4 (0.7 to 2.0)	20.5 (11.0 to 29.2)	20.9 (11.6 to 29.6)
55	3.1 (2.9 to 3.5)	6.5 (5.9 to 7.3)	0.8 (0.5 to 1.1)	1.6 (0.9 to 2.3)	24.5 (14.7 to 34.2)	24.0 (14.4 to 33.7)
65	3.6 (3.3 to 4.1)	6.5 (6.0 to 7.4)	0.9 (0.5 to 1.3)	1.6 (0.9 to 2.3)	24.4 (14.6 to 34.1)	24.0 (14.3 to 33.6)
Age of women with a positive family history, y						
45	6.6 (5.2 to 8.6)	13.3 (10.6 to 17.2)	1.3 (0.7 to 2.1)	2.7 (1.4 to 4.2)	20.3 (10.9 to 29.0)	20.3 (11.21 to 28.89)
55	6.8 (5.4 to 8.8)	13.7 (11.1 to 17.6)	1.6 (0.9 to 2.5)	3.2 (1.8 to 4.8)	24.1 (14.4 to 33.7)	23.2 (13.8 to 32.5)
65	7.8 (6.3 to 10.2)	13.8 (11.2 to 17.7)	1.9 (1.1 to 2.9)	3.2 (1.8 to 4.8)	24.0 (14.3 to 33.5)	23.1 (13.8 to 32.4)
Age of women in the top 10% population absolute risk, y‡						
45	8.1 (6.3 to 10.7)	16.1 (12.7 to 21.0)	1.6 (0.8 to 2.6)	3.2 (1.6 to 5.0)	20.1 (10.5 to 29.2)	19.7 (10.4 to 28.4)
55	9.3 (7.3 to 12.2)	18.5 (14.8 to 23.9)	2.4 (1.3 to 3.6)	4.4 (2.4 to 6.8)	25.4 (15.3 to 35.3)	24.3 (14.5 to 33.9)
65	10.7 (8.4 to 14.0)	18.6 (14.9 to 24.0)	2.5 (1.5 to 4.1)	4.1 (2.5 to 6.8)	23.3 (15.1 to 35.0)	22.2 (14.4 to 33.6)

* CI = confidence interval.

† Mean difference between non-modified absolute risk and absolute risk obtained by assuming that all current drinkers became never drinkers, all women who exercised less than 2 hours per week began exercising at least 2 hours per week, and all women aged 50 years and older maintained BMI less than 25 kg/m^2.

‡ Women who account for the highest 10% of absolute risk in the EPIC population.

more of the total population by lowering the absolute risk threshold for inclusion, the average absolute risk reduction decreased, finally reaching the average absolute risk reduction in the entire population (100%).

Discussion

We developed a new absolute risk prediction model for invasive breast cancer for Italian women. The model includes non-modifiable risk factors and three potentially modifiable factors, BMI, leisure-time physical activity, and alcohol consumption. The model was reasonably well calibrated overall in independent data from the Florence-EPIC cohort study, but overestimated absolute risk in some subgroups such as women aged 60 years or older and women whose first live birth occurred at age 30 years or later. The discriminatory accuracy (concordance) in the Florence-EPIC cohort data was 0.62 at age less than 50 years and 0.57 for older women, and is comparable with that of other absolute risk models for breast cancer (4–6,8,25–29).

A novel aspect of this work is the evaluation of the potential effects of reducing exposures from modifiable risk factors on absolute breast cancer risk, not only for the individual counselee but also for the entire population and high-risk subgroups. We developed methods based on the Lorenz curve of population absolute risk to identify high-risk subgroups. Assessment of the reduction in average absolute risk gave a different perspective than assessment of the fractional risk reduction, which is analogous to attributable risk. Indeed, in the entire population, 20-year fractional risk reductions are 20%−24%, whereas absolute risk reductions are 1.4%−1.6%. Fractional risk reductions are less useful for clinical and public health decisions than absolute risks and absolute risk reductions (30). Our methods are also useful for designing intervention trials, because the power of such trials depends on the average absolute risk with and without intervention (31,32). In our study, the absolute reduction is nearly proportional to absolute risk before modification of risk factors, reflecting proportional hazards assumptions. Thus the fractional risk reduction was nearly constant across categories of risk. However, the fractional risk reductions are greater in older women for whom lower BMI was associated with reduced risk.

Estimates of the potential effects of interventions on absolute risk can provide perspective on whether to pursue prevention research or implement interventions. For example, our estimates for risk factor modifications indicate an approximate 1.6% absolute risk reduction during 20 years in the general postmenopausal population, and an approximate 3.2% absolute risk reduction for women with a positive family history of breast cancer. For women at the highest absolute risk who account for 10% of total population absolute risk, the absolute risk reduction is approximately 4.4%. In a population of 1 million women, even a 1.6% absolute risk reduction amounts to 16 000 fewer cancers. Because programs to encourage less alcohol consumption, increase leisure activity, and encourage some weight control are likely to be safe, they can be widely administered. As emphasized by Rose (33,34), broadly applicable interventions can be more effective than interventions focused on high-risk subgroups. If these interventions were restricted to the 8% of postmenopausal women with a positive family

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

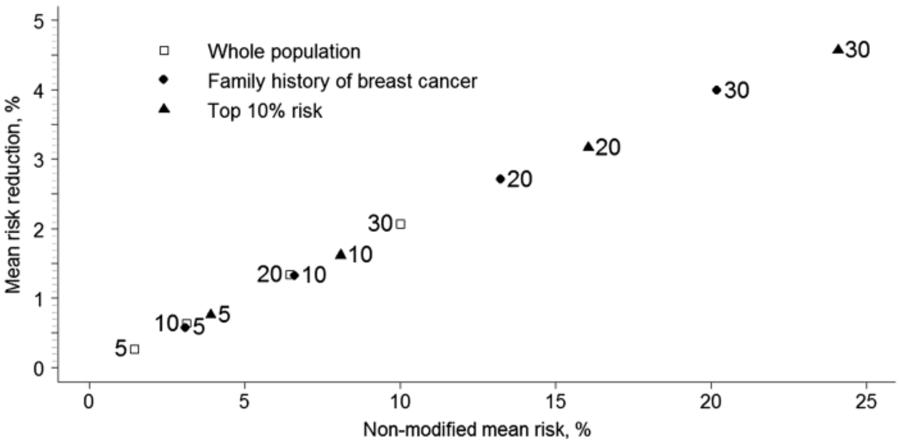

Figure 1. Estimated 5-, 10-, 20-, 30-year projections of the mean risk reduction vs the non-modified mean risk in the whole population of 45-year-old women and women in two high-risk subsets of the population. Women with a positive family history for breast cancer and women who account for the highest 10% of risk in the population make up the two high-risk subsets. The numbers next to the symbols denote the risk projection interval in years.

history of breast cancer, then in a population of 1 million women, only 2560 breast cancers would be prevented.

A strength of the study was the quality of the data used for developing the model and for validation. Selection bias was limited in the case–control data because the participation rate was high, and the catchment areas were comparable for case patients and control subjects. The comparability of recall between case patients and control subjects was improved by interviewing all study participants in a hospital setting (15).

A number of limitations need to be considered. Although our model was reasonably well calibrated, there was evidence of overestimation in the highest quintile of absolute risk. Recalibration (35) led to smaller odds ratios, improved fit to Florence-EPIC data in this quintile, and smaller estimates of the effects of modifying risk factors (unreported data). This analysis and other unreported numerical studies indicate that the estimated absolute risk reductions are sensitive to the estimated odds ratios for modifiable risk factors.

Another limitation was that the estimated absolute risk reductions are imprecise for the individual counselee. Population level estimates are more precise, but both individual- and population-level estimates are subject to systematic errors, which are not reflected in the confidence limits. An ideal study to estimate the effects of interventions would be a randomized intervention trial, such as the Woman's Health Initiative (36) or the Breast Cancer Prevention Trial (31). Such trials yield unbiased estimates of treatment effect and information on compliance. Although trial results may not generalize quantitatively to the general population, they provide a good guide to preventative strategy. Estimates from case–control data may yield associations that are confounded by other factors or biased by differential recall. The associations observed in such data may therefore not predict actual preventative effects. Without empiric data from intervention studies, we cannot test the assumptions underlying our model. A key assumption is the proportional hazards assumption, whereby a risk factor modification acts immediately and indefinitely to multiply the breast cancer hazard by a constant factor, unless the model includes an interaction with time. In fact, it is unknown how long it will take before an intervention affects breast cancer hazard rates or how long the effect will last. Under the proportional hazards assumption, our calculations give an idea of the largest reductions in absolute risk that might be achieved.

A further limitation is that the interventions are not specified. Knowing that a woman with elevated BMI is at increased risk does not define the intervention. Yet an estimation of the causal effect of the intervention on absolute risk is desired (37). We have made optimistic assumptions that interventions could reduce modifiable exposures to their lowest risk levels; thus our calculations would give an upper bound on the reductions in absolute risk.

With hospital-based controls, associations can be distorted by correlations between the risk factors and the control diseases; however, we chose control diseases to avoid this bias. The Florence-EPIC cohort was not a random sample of the population of Italian women,

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

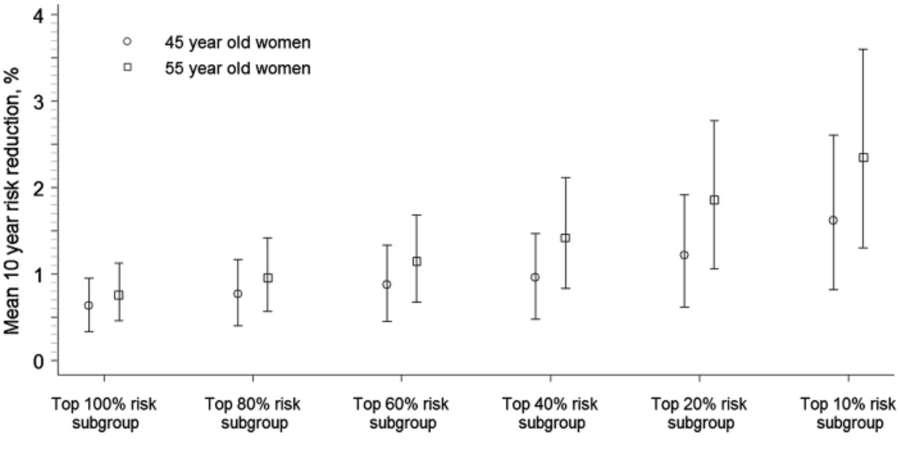

Figure 2. Estimated 10-year mean risk reductions in 45- and 55-year-old women in subsets that contain varying proportions of the total population risk. Top 10% risk subset = the subset of women that contains the highest 10% of population risk. Other subsets are defined similarly. **Vertical lines** represent 95% confidence intervals for the estimated risk reductions.

and the cohort may have had a more favorable distribution of lifestyle risk factors than the general population. Moreover, results may not generalize to other countries, where the prevalence of obesity may be larger or the frequency of mammographic screening greater.

Cummings et al. (24) reviewed the literature on possible interventions to reduce breast cancer incidence, and concluded that lifestyle interventions such as exercise, weight reduction, low-fat diet, and reduced alcohol intake should be included in programs to prevent breast cancer. Despite their limitations, calculations of reductions in absolute risk using our model potentially provide additional perspective on the possible benefits of such prevention strategies. If combined with operational definitions of interventions and their effect sizes, our methods can provide information needed to compute sample sizes for trials to evaluate such interventions.

APPENDIX

Appendix Table 1. Parameter estimates and covariances for Equation 3*

AgeMen	NumRel	Age1st	NBiops	Currndrnk	ExDrnk	Educat	LeiAct	OccAct	InvBmi	Bmi
0.0411	0.8531	0.2627	0.2759	0.2360	0.2041	0.3157	0.0783	0.0946	0.2350	0.1247
0.173	0.001	0.008	0.010	0.011	−0.003	−0.024	−0.008	−0.005	0.025	−0.010
	1.383	0.014	−0.028	0.008	0.001	0.005	−0.003	−0.008	0.009	0.0001
		0.145	0.023	−0.002	0.012	−0.028	0.004	−0.018	−0.006	0.007
			6.150	0.0001	0.033	−0.029	0.0455	0.027	−0.051	0.011
				0.407	0.272	0.003	0.0102	0.011	−0.041	0.028
					1.729	0.012	0.0583	0.035	−0.022	0.023
						0.189	0.0012	−0.065	−0.034	0.024
							0.3961	−0.040	0.007	−0.003
								0.388	−0.012	0.017
									0.661	0.928
										0.228

* All covariances are 10^{-2} times the numbers in the table. The covariances are in the triangular array beginning with the second row of numbers. Parameter estimates are in the first row. The variables used in Equation 3 and this table are: AgeMen = age at menarche, Age1st = age at first live birth, NumRel = number of affected first-degree relatives, NBiops = number of biopsies, OccAct = occupational physical activity, Educat= education, CurrnDrnk = current drinker, ExDrnk = former drinker, InvBmi = body mass index at age <50 years (reference category = "≥30.0"), Bmi = body mass index at age ≥50 years (reference category = "<25.0"), LeiAct = leisure-time physical activity.

Appendix Table 2. Age-specific composite breast cancer incidence rates per 100 000 women-years and age-specific mortality rates per 100 000 women-years. Data from the Florence Cancer Registry collected from January 1, 1989, to December 12, 1993 (N = 1 190 516)

Age, y	Incidence rates	Mortality rates
20–25	0	26.8
25–30	5.4	29.4
30–35	23.5	53.5
35–40	75.9	46.0
40–45	118.2	81.2
45–50	190.4	130.5
50–55	215.7	200.0
55–60	215.4	295.6
60–65	237.5	464.9
65–70	278.9	835.0
70–75	264.8	1514.5
75–80	304.8	2729.1
80–85	274.2	5984.3
85–90	275.9	14401.4

References

1. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet.* 1991;48(2):232–242.
2. Berry DA, Parmigiani G, Sanchez J, Schildkraut J, Winer E. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *J Natl Cancer Inst.* 1997;89(3):227–238.
3. Antoniou AC, Pharoah PP, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer.* 2004;91(8):1580–1590.
4. Chen J, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J Natl Cancer Inst.* 2006;98(17):1215–1226.
5. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst.* 2006;98(17):1204–1214.
6. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst.* 1999;91(18):1541–1548.
7. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med.* 2004;23(7):1111–1130.
8. Rockhill B, Byrne C, Rosner B, Louie MM, Colditz G. Breast cancer risk prediction with a log-incidence model: evaluation of accuracy. *J Clin Epidemiol.* 2003;56(9):856–861.

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011

9. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol.* 2000;152(10):950–964.
10. Boyle P, Mezzetti M, La Vecchia C, Franceschi S, Decarli A, Robertson C. Contribution of three components to individual cancer risk predicting breast cancer risk in Italy. *Eur J Cancer Prev.* 2004;13(3):183–191.
11. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol.* 1974;99(5):325–332.
12. Benichou J. A review of adjusted estimators of attributable risk. *Stat Methods Med Res.* 2001;10(3):195–216.
13. Calza S, Specchia C, Frasca G, et al. EPIC-Italy cohorts and multipurpose national surveys. A comparison of some socio-demographic and life-style characteristics. *Tumori.* 2003;89(6):615–623.
14. Masala G, Assedi M, Saieva C, et al. The Florence city sample: dietary and life-style habits of a representative sample of adult residents. A comparison with the EPIC-Florence volunteers. *Tumori.* 2003;89(6): 636–645.
15. Mezzetti M, La Vecchia C, Decarli A, Boyle P, Talamini R, Franceschi S. Population attributable risk for breast cancer: diet, nutrition, and physical exercise. *J Natl Cancer Inst.* 1998;90(5):389–394.
16. Palli D, Berrino F, Vineis P, et al.; EPIC-Italy. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of partecipants. *Tumori.* 2003;89(6): 586–593.
17. Breslow NE, Day NE. *Statistical Methods in Cancer Research. Volume I–The Analysis of Case-Control Studies.* Lyon, France: International Agency for Research on Cancer, Publication number 32; 1980.
18. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York, NY: Springer-Verlag; 2001.
19. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81(24):1879–1886.
20. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol.* 1985;122(5):904–914.
21. Efron B, Tibshirani R. *An Introduction to the Bootstrap.* New York, NY: Chapman & Hall; 1993.
22. Harrell FE Jr, Lee KL, Mark DB. Tutorial in biostatistics. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4): 361–387.
23. Dagum C. *Lorenz Curve.* In: Kotz S, Johnson NL, eds. *Encyclopedia of Statistical Sciences.* New York, NY: Wiley; 1985:156–161.
24. Cummings SR, Tice JA, Bauer S, et al. Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk. *J Natl Cancer Inst.* 2009;101(6):384–398.
25. Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst.* 2007;99(23):1782–1792.
26. Chlebowski RT, Anderson GL, Lane DS, et al. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J Natl Cancer Inst.* 2007;99(22):1695–1705.
27. Decarli A, Calza S, Masala G, Specchia C, Palli D, Gail MH. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. *J Natl Cancer Inst.* 2006;98(23):1686–1693.
28. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst.* 2001;93(5):358–366.
29. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med.* 2008;148(5):337–347.
30. Schwartz LM, Woloshin S, Dvorin EL, Welch HG. Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *BMJ.* 2006;333(7581):1248.
31. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst.* 1998;90(18):1371–1388.
32. Halperin M, Gordon T, Kjelsberg M, Neaton J, Sherwin R. Statistical design considerations in the NHLI multiple risk factor intervention trial (MRFIT). The Multiple Risk Factor Intervention Trial Group. *J Chronic Dis.* 1977;30(5):261–275.
33. Rose G. *The Strategy of Preventive Medicine.* New York, NY: Oxford University Press; 1992.
34. Rose G. Strategy of prevention: lessons from cardiovascular disease. *Br Med J (Clin Res Ed).* 1981;282(6279):1847–1851.
35. Cox DR. Two further applications of a model for binary regression. *Biometrika.* 1958;45(3–4):562–565.
36. Anderson G, Cummings S, Freedman LS, et al. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998;19(1):61–109.
37. Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes (Lond).* 2008;32(suppl 3):S8–S14.

Funding

This research was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health and the Associazione Italiana per la Ricerca sul Cancro (AIRC IG 10415 to A.D.).

Affiliations of authors: Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD (EP, CS, RMP, MHG); Occupational Health Department, Branch of Medical Statistics and Biometry, University of Milan and Fondazione IRCSS Istituto Nazionale Tumori, Milan, Italy (AD); Molecular and Nutritional Epidemiology Unit, ISPO Cancer Prevention and Research Institute, Florence, Italy (GM, DPa); Information Management Services, Rockville, MD (DPe).

Downloaded from jnci.oxfordjournals.org at University of Milan on October 4, 2011