Title: Risk Factors and Outcomes in a Multiple Tumor Marker Setting: The Issues of Correlated and Missing Markers

Kathryn C Fitzgerald,[1] Robert J Glynn,[2] Rulla M Tamimi,[1,3] Wendy Y Chen,[3] Graham C Colditz,[1,3] Susan E Hankinson,[1,3] and Bernard Rosner[2,3]

[1]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, Massachusetts
[2]Departmend of Biostatistics, Harvard TH Chan School of Public Health, Boston, Massachusetts
[3]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School,

INTRODUCTION

Risk profiles for breast cancer generally differ according to various characteristics of the tumor. For example, pregnancy is generally inversely associated with estrogen receptor positive (ER+) breast cancer and equivocal (or possibly deleterious) for ER- breast cancer[2-4]. The most common approach to address tumor heterogeneity uses methodology for competing risks; markers are cross-classified and each combination of markers is considered individually[5]. However, as an increasing number of tumor markers are identified (e.g. tumor size, morphology or multiplicity) and as these characteristics are often correlated, it becomes difficult to assess the marker-specific effect of a particular risk factor without considering other tumor characteristics. To address these issues, Rosner et. al introduced the concept of an *adjusted* hazard ratio. Here, hazard ratios from sub-type specific risk factor models are computed and interpreted as the marker-specific effect of a particular risk factor while "controlling" for the other tumor characteristics [1].

Another added layer of complexity introduced by considering many tumor characteristics simultaneously is the inevitable increase in missing marker information - that is – a case will have information on only a subset of the markers being considered. In this context, often the analysis is restricted to disease cases with complete data on each of the markers in question, though other traditional approaches for missing data (i.e. including a missing indicator, using inverse probability weighting or implementing a multiple imputation algorithm) could also be considered. It's unclear as to which missing data approach is most appropriate, especially in the context of correlated markers.

As such, a combination approach that both properly addresses missing markers as well as integrates correlation among them would provide the most accurate estimate of a risk factor's effect on marker-specific disease.

In what follows, we apply Rosner et. al's methodology to calculate *adjusted* hazard ratios accounting for missing markers using each of the following approaches: the complete case, missing indicator, inverse probability weighting and multiple imputation. Each method is applied in an analysis conducted in the Nurses' Health Study (NHS) of associations of traditional breast cancer risk factors with subtype-specific disease considering 5 different markers. We also compare and assess performance of each approach in extensive simulation studies with respect to bias, standard error and coverage probabilities.

METHODS
Study population and risk factor assessment

The Nurses' Health Study began in 1976 and includes 121,701 female nurses who were aged 30 to 54 responded to a mail questionnaire inquiring about lifestyle and disease characteristics[6]. Follow-up for the NHS continues through biennial questionnaires where cohort participants continue to report lifestyle characteristics such as information on risk factors for breast cancer including reproductive factors (menopausal status, age at menarche, parity), hormone use (including type and duration of use), anthropometric variables (body weight and height), benign breast disease (BBD) and family history of breast cancer. Alcohol intake at age 18 was ascertained in 1988. Current alcohol intake was assessed in 1980 and assessed every 4 years thereafter.

Ascertainment of Breast Cancer
On each biennial questionnaire, women were asked to report whether breast cancer had occurred and the date of occurrence. For those women reporting breast cancer occurrence, we requested permission (including next of kin for decedents) to contact the treating physician and for release of relevant medical records. To ascertain estrogen receptor (ER), progesterone receptor (PR) and HER2 status, we reviewed pathology reports. Overall, 4380 cases of invasive breast cancer cases occurring from 1980 to 2010 with pathology reports were included; 1395 breast cancer cases with missing ER and/or PR status were censored at the time of diagnosis. We also excluded women with type of menopause other than natural or bilateral oophorectomy, prevalent cancer (other than non-melanoma skin cancer) in 1980, or missing data for weight at age 18, age at first birth, parity, age at menarche, age at menopause or hormone use leaving 77,232 women who were eligible for the analysis. For tumors diagnosed prior to 2002, HER2 status was determined by immunohistochemical staining performed on paraffin sections of tumor tissue microarray (TMA) according to a standard protocol as it was not routine clinical practice to assess HER2 status during these years. A more detailed description of TMA construction and ER, PR and HER2 immunohistochemical staining have been previously reported [4]. Following 2000, HER2 status was obtained from pathology and medical reports [as primarily determined by IHC or a subgroup also tested using fluorescent in situ hybridization (FISH)]. We censored women with ER-/PR+ breast cancer as a subset of 71 women with ER/PR also determined by TMA, only 4 (6%) were confirmed as ER-/PR+. Pathology reports were also used to define tumor grade (as well or moderately differentiated vs. poorly differentiated) and size (defined as large ≥2cm vs. small <2cm).

Statistical Analysis
(a) General Competing Risks Framework
We initially performed analyses considering only ER and PR status to demonstrate the use of the *adjusted* hazard ratio (as each of these markers is completely observed). We then expanded our analysis to consider 5 tumor characteristics (some with incompletely observed data): ER status, PR status, HER2 status, tumor grade (well or moderately differentiated vs. poorly differentiated) and size (large ≥2cm vs. small <2cm) and included the following breast cancer risk factors: reproductive factors (duration of premenopause, duration following natural menopause, duration following bilateral oophorectomy, gynecologic age at first birth = $age_{first\ birth}$-$age_{menarche}$, birth index = $\sum_{i=1}^{S_t}(t^* - t_i)b_{it}$, where $b_{it} = 1$ if parity $\geq i$ at age t, = 0 else, t_i =age at i^{th} birth, $t^* =$ min(age, age at menopause), anthropometric and lifestyle characteristics (both before and after menopause - body mass index, height, alcohol intake, and use/type of hormones), and disease history (family history of breast cancer, benign breast disease). For a nulliparous woman, her gynecologic age at first birth and birth index are set to 0. In the following equations, *k* denotes tumor marker, *l* denotes subtype (i.e. those defined by the *k* tumor markers), and *j* denotes risk factor. Using the approach outlined by Lunn and

McNeil, levels of each tumor marker are cross-classified and considered individually. The hazard for a woman with tumor subtype l relative to no breast cancer is given by the following Cox-proportional hazards model:

$$h_l(t|\boldsymbol{x}) = h_{0l}(t)\exp(\boldsymbol{\beta_l x}), l = 1, \dots, L \tag{1}$$

Hazard ratios can be estimated using data duplication approaches detailed by Lunn and McNeil by creating separate records for each tumor type. A test of whether the regression coefficient for the jth risk factor varies among subtypes is given by $H_0: \beta_{1j} = \cdots = \beta_{Lj}$ vs. H_1: at least some β_{ij} are different is performed using a likelihood ratio test. Using this approach, some risk factors can have the same regression coefficients for different tumor types, but others can have different regression coefficients. We can assess heterogeneity by testing using likelihood ratio tests or alternatively, we can test the effect of the jth risk factor on tumor subtype l.

(b) Description of Calculation of Adjusted Hazard Ratios
Alternatively, we can write the original competing risks model for the lth subtype as

$$h(t|\boldsymbol{x}, \boldsymbol{w_l}) = h_{0\boldsymbol{l}}(t)\exp\left(\sum_{j=1}^{J}\beta_j x_j + \sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{jk} w_k x_j\right) \tag{2}$$

where x_j is the jth risk factor and w_k=score of the kth tumor marker for the lth subtype and $\exp\left(\beta_j + \sum_{k=1}^{K}\gamma_{jk} w_k\right)$ is the hazard ratio for a 1 unit increase in the jth risk factor for tumor type=$\mathbf{w}_1$. Using γ_{jk} as a measure of heterogeneity of the effect of the jth risk factor by the kth tumor marker, we can calculate the *adjusted* hazard ratio as

$$HR_{adj}(x_j|w_k) = exp\left(\beta_j + \gamma_{jk}w_k + \sum_{\boldsymbol{W_{-k}}}\left[\sum_{\substack{k_1=1 \\ k_1 \neq k}}^{K}\gamma_{jk_1}w_{k_1}Pr(W_{k_1} = w_{k_1})\right]\right) \tag{3}$$

where the *adjusted* hazard ratios can be interpreted as the effect of the jth risk factor at level w_k of the the kth tumor marker while controlling for the other markers.

(c) Methods for Handling Missing Subtype Information
For our analysis, we considered only those subtypes with at least 5 cases. In what follows, we provide a brief description of each of the methods to address missing marker information considered in this paper.

1. Complete Case Method
For the complete case analysis, we considered only 1551 of the 4380 cases of breast cancer with information available on all 5 tumor markers. Women with breast cancer and missing subtype information are censored at diagnosis.

2. Missing Indicator Method
For the missing indicator method, missing is considered an additional level of each tumor marker with missing information. Then in the linear transformation used to calculate the *adjusted* hazard ratio, additional binary variables are added to indicate missingness for a specific marker. For example, a woman's HER-2 status can take one of the following values: positive, negative or missing, which would be coded using 2 indicator variables in the calculation of the *adjusted* hazard ratio. Under missing at random, for a specific marker, the *adjusted* hazard ratio with missing indicators is interpreted as a weighted average over the observed markers.

Inverse Probability Weighting (IPW)

The complete case dataset is reweighted to account for missing markers. To create the weights, we fit the following logistic model and obtain the predicted probabilities of a tumor being a complete case (C; i.e having all 5 tumor markers present)

$$logit[\Pr(C=1)] = \theta_0 + \theta_1(ER_+) + \theta_2(PR_+) + \theta_3(calendar\ year) + \theta_4(age_{diagnosis}) \qquad (4)$$

The complete case analysis is then rerun weighted by the weight calculated as $w = \frac{1}{p}$, where p is calculated as the probability $\Pr(C=1)$ from the above model.

Multiple Imputation

We implemented a chained imputation approach where we first imputed missing HER-2 status, then imputed missing tumor grade and finally imputed missing tumor size. Briefly, to impute HER-2 status, we calculated the predicted probabilities of HER-2$_{positive}$ for women with missing HER-2 status from the results of the following logistic regression model.

$$logit[\Pr(HER2_+=1)] = \tau_0 + \tau_1(ER_+) + \tau_2(PR_+) + \tau_3(calendar\ year) + \tau_4(age_{diagnosis}) \quad (5)$$

Using these probabilities, we generated a binary HER-2 status using a Bernoulli trial. We then performed a similar procedure to impute missing tumor grade using the following logistic model and predicted probabilities for those with missing tumor grade

$$logit[\Pr(Grade_{Poor}=1)] = \gamma_0 + \gamma_1(ER_+) + \gamma_2(PR_+) + \gamma_3(HER2_+) + \gamma_4(calendar\ year)$$
$$+\gamma_5(age_{diagnosis}) \qquad (6)$$

Then, to estimate missing tumor size, we fit a final similar model to impute missing tumor size. We used the predicted probabilities for those with missing tumor size using estimates from the following logistic regression model:

$$logit[\Pr(Size_{large}=1)] =$$
$$\delta_0 + \delta_1(ER_+) + \delta_2(PR_+) + \delta_3(HER2_+) + \delta_3(Grade_{well_+}) + \delta_4(calendar\ year) + \delta_5(age_{diagnosis})$$
$$(7)$$

We then used the complete dataset generated from the above imputation procedure to calculate the *adjusted* hazard ratios. We performed the imputation and analysis for 10 different datasets and combined results using Rubin's rules where the reported *log*(HR) is the average of the *log*(HR) calculated from the 10 imputed datasets. The variance of the *log*(HR) is calculated as the sum of the within and between imputation variance.

Simulation Study

We attempted to simulate conditions similar to the actual conditions observed in the NHS where parameters for each distribution are estimated using data from the actual study. A full list of the specified parameters used to generate the simulated data is provided in an Appendix available upon request from authors. To start, we generated a cohort of 10,000 hypothetical postmenopausal women. For each woman, we generated her age using a normal distribution and randomly generated menopausal type (x_p) as bilateral oophorectomy or natural menopause) using a Bernoulli trial. Conditional on menopausal type, we generated parous status (Z) and if parous (Z=1), we generated her number of births using a Poisson distribution. For each menopausal type (and parous status), we

generated her age at menarche (x_{m1}), age at menopause (x_{m2}) and age at first birth (for parous women) using a multivariate normal distribution. These were used to calculate her gynecologic age at first birth (x_G) and birth index (x_{BI}). For x_{BI}, a woman's age at each birth was generated by drawing a sample equal to her number of children from a random uniform distribution on her interval [age$_{menarche}$, age$_{menopause}$]. For nulliparous women (e.g. Z=0), both x_{BI} and x_G are set to 0. Lastly, we generated a woman's BMI (x_{BMI}) conditional on her menopausal type using a normal distribution. We denote the set of these 6 covariates ($x_{m1}, x_{m2}, x_p, x_G, x_{BI}, x_{BMI}$) as the vector $\boldsymbol{x}$. This simulation procedure was performed to generate 10,000 simulated datasets.

Similar to the actual data analysis presented, we simulated 5 binary tumor characteristics and generated a woman's survival time from a proportional hazards model using the method described in Bender et al,[7]

$$h(t|\boldsymbol{x}, w_1, w_2, w_3, w_4, w_5) = h_{0w_1,w_2,w_3,w_4,w_5}(t)\exp\left(\sum_{j=1}^{6} \gamma_{0j} x_j + \sum_{j=1}^{6}\sum_{k=1}^{5} \gamma_{jk} w_k x_j\right) \qquad (8)$$

We specified parameters for the "common parameter" vector $\boldsymbol{\gamma_0}$; these parameters reflect the effect of a risk factor on the specific subtype when each marker is set to its referent value (i.e. a tumor of the subtype: ER negative, PR negative, HER2 negative, small and well-differentiated). We also set values for each γ_{jk}, and the full matrix of parameters is provided in Appendix Table 1. We then generated missing disease characteristics for 3 markers conditional on 2 fully observed markers (akin to having restricting to women with available ER and PR information but potentially missing HER2, size or grade information) such that 20% (or 50%) of breast cancer cases were missing information on at least 1 marker. We generated which combination of the 3 markers were missing using a multinomial distribution using patterns of missing (e.g. missing patterns we observed for HER2, grade and size) observed in the NHS (the full specification of the multinomial distribution used is provided in the in Appendix Table 2). We then applied each of the 4 methods to address missing data: the complete case, missing indicator, IPW and multiple imputation and calculated *adjusted* hazard ratios and their corresponding standard errors. This procedure was repeated for 10,000 simulated datasets for each level of missingness (20% and 50%). We calculated the average bias, the %change in standard error and the average coverage probability for each of the missing data methods relative to the estimate obtained using the completely observed data for a given simulation.

RESULTS
Real Data from the NHS
We fit separate Cox proportional hazards models considering initially two markers defining 3 subtypes (with complete case information) among 4380 breast cancer cases resulting in the following cross-classified subtypes: ER+/PR-, ER+/PR- and ER-/PR- (data available upon request from the authors). Notably, premenopausal BMI (BMI at age 50) was inversely associated with each of the three subtypes whereas postmenopausal BMI (BMI at age 70) was positively associated with ER+/PR+ cancer but potentially inversely associated with both ER+/PR- and ER-/PR- cancers cancer (Premenopausal BMI: P_{het}=0.08; Postmenopausal BMI: P_{het}<0.0001). In addition, a woman's birth index appeared to be inversely associated with both ER+/PR+ and ER+/PR- cancers but not ER-/PR- cancer. We then considered the adjusted HR model introduced by Rosner et. al [1] for these 2 tumor markers. In comparison to the unadjusted HR model, postmenopausal BMI was strongly positively associated with PR+ cancer but not PR- cancer after controlling for a woman's ER status (P_{het}<0.001). We did not observe any heterogeneity

for postmenopausal BMI and risk of ER+ vs. ER- cancer (P_{het}=0.60; data available upon request from the authors). Additionally, a woman's birth index appeared differentially associated with ER+ vs. ER- cancer suggesting pregnancy to only be inversely associated with ER+ cancer (P_{het}=0.03) but uniformly inversely associated with PR+ and PR- cancers (P_{het}=0.74).

We then expanded our analysis to consider HER-2 status, tumor size and tumor grade in addition to ER and PR status. Table 1 (A) depicts the frequency distribution for each of the 5 cross-classified tumor markers considering only cases with complete information on all possible subtypes (1551 total cases). An ER+/PR+/HER2-/Well/Small tumor was the most common subtype with 663 of the 1551 cases. The next most common subtype was ER+/PR+/HER2-/Well/Large tumors with 231 cases. However, if we consider missing tumor characteristics, 4380 cases of breast cancer are considered. Table 1 (B) depicts cross-classified tumors considering missing characteristics. Notably, while ER+/PR+/HER2-/Well/Small were still the most common subtype, a nearly equally common subtype was ER+/PR+/HER2-missing/Well/Small tumors with 624 cases. The fully cross-classified distribution considering missing case type information and marginal distribution of each marker are available upon request from the authors.

We then fit Cox-proportional hazards models and calculated multivariable-adjusted HR for each of the 4 methods to address missing tumor information. Results for birth index are depicted in Table 2. We observed generally similar results for each of the 4 methods where we observe heterogeneous effects of a woman's birth index for ER+ vs. ER- cancer adjusted for PR, HER-2, size and grade but no heterogeneity for any of the other adjusted tumor characteristics. For birth index, the standard error for each subtype-specific HR appeared largest for the complete case model and was generally smallest for the missing indicator and multiple imputation methods.

When evaluating the associations between postmenopausal BMI and the subtype-specific HR and corresponding test for heterogeneity for each of the missing data methods, the HR were generally similar; however, we did note a few differences with respect to the tests for heterogeneity (Table 3). For example, higher postmenopausal BMI appeared significantly heterogeneously associated with poorly differentiated tumors if missing-marker information is addressed using the missing indicator or multiple imputation methods. Also noteworthy was the HR associated with postmenopausal BMI for a large tumor was much smaller using multiple imputation when compared with the other 3 methods.

A comparison for each of the four methods for all risk factors considered is available upon request from the authors.

Simulation study
Table 4 depicts the results of the simulation study for a covariate used to simulate the effect of birth index. For birth index, we assume no effect (γ_0 =0) of birth index on the specific subtype when each marker is set to its referent value. We also assume a protective effect only for ER positive subtypes (γ_{ER}=-0.25) and well-differentiated (γ_{Grade}=-0.12) tumors and no effect for PR (γ_{PR}=0), HER2 (γ_{HER2}=0.0) and tumor size (γ_{Size}=0). For each of the methods for addressing missing data we observed negligible biases for the complete case, IPW and missing indicator methods. The multiple imputation analysis produced slightly more bias for both 20% and 50% missing tumor marker information; however, the extent of the increase was not substantial. In terms of

efficiency, the complete case and IPW had the largest increases in standard errors with a uniform average increase across tumor markers of 12% and 9%, respectively, under 20% missing and 42% and 33%, respectively, under 50% missing. However, for the multiple imputation and missing indicator methods, we only observed an increase in standard error roughly proportional to the degree of missing. Thus, we did not observe substantial increases in the standard errors for markers without any missing information (e.g. ER, PR) but did observe changes in standard error for HER2, grade and size where increases in standard errors were larger for markers with the most missing (HER2) and smaller for markers with a lesser degree of missing information (size). The coverage probabilities for the associated 95% confidence intervals were also close to the nominal level for the complete case and missing indicator methods. Coverage probabilities for analyses adjusting for missing using IPW were generally lower; for analyses adjusting for missing using multiple imputation, we only observed a decrease in coverage probability for the birth index parameter for tumor grade, which, notably, is the only non-zero coefficient with missing information.

Table 5 depicts the results of the simulation study for a covariate used to simulate the effect of postmenopausal BMI on subtype-specific breast cancer. We assume a common protective parameter across all subtypes as -0.14 (γ_0), a protective effect for ER positive (γ_{ER}=-0.20), a harmful effect for PR positive (γ_{PR}=0.65), HER2 positive (γ_{HER2}=0.14), and large (γ_{Size}=0.50) tumors. We assume no effect of postmenopausal BMI on tumor grade (γ_{Grade}=0). Similar to the coefficients for birth index, for each of the methods for addressing missing data we observed negligible biases for the complete case, IPW and missing indicator methods. The multiple imputation analysis produced more bias for both 20% and 50% missing tumor marker information with the extent of the bias increasing among markers with the most missing information. In terms of efficiency, the patterns of change in standard error were similar to patterns observed in Table 4 where the complete case and IPW had similarly large increases in standard errors uniformly across markers (regardless of the amount of missing tumor information for the specific marker) and where the missing indicator and multiple imputation methods produced increases in standard error specific to the degree of missing in the marker. Thus, we did not observe substantial increases in the standard errors for markers without any missing information (e.g. ER, PR) but did observe changes in standard error for HER2, grade and size. The coverage probabilities for the associated 95% confidence intervals were also close to the nominal level for the complete case and missing indicator methods. Coverage probabilities for analyses adjusting for missing using IPW were generally lower; for analyses adjusting for missing using multiple imputation, we observed a marked decrease in coverage probability for the BMI parameter for HER2 and size at 58% and 76%, respectively.

DISCUSSION

In this paper, we evaluated several approaches for addressing missing tumor marker data in the context of correlated markers to calculate subtype-specific "adjusted" hazard ratios. The missing data approaches we considered were the complete case, inverse probability weighting, missing indicator and multiple imputation methods. We applied each of the approaches to a study of breast cancer risk factors in the Nurses' Health Study considering 5 correlated tumor characteristics (ER, PR, HER2, grade and size). Overall, the results of these analyses suggest similar results for the complete case and inverse probability weighting approaches and similar results for the missing indicator and multiple imputation approaches. We also observed generally smaller standard errors for the missing indicator and multiple imputation approaches as well. Ultimately, our

subsequent extensive simulation studies suggested that the missing indicator method for addressing missing tumor marker may provide the least bias and smallest increase in variance while conserving the nominal 95% confidence limits.

Other investigators have proposed alternative methods to address missing information tumor marker data. One approach suggested by Chatterjee et al[8] implements an estimating equation approach that requires parametric specification of the interdependence of the effects of specific markers for the baseline hazard. With respect to our study and for breast cancer in general, this is challenging as with the exception of a few well-studied markers it's rarely known *apriori*. In our study specifically, we observed potential correlations among all 5 markers considered; however, we do not need to assume a specific functional relationship between them.

The analyses presented also represent an update of the initial paper introducing the concept of an *adjusted* hazard ratio but considering more than 1000 additional breast cancer cases and considering 2 additional tumor markers: size and grade. Our analysis confirms our initial findings for the heterogeneous effect of pregnancy for ER+ tumors after accounting for PR status where there is no such heterogeneity. Similarly, we also observed a significant adverse effect of postmenopausal BMI on PR+ tumors and a potentially adverse effect on larger tumors (although this could potentially reflect a detection bias, as smaller tumors may be more difficult to detect in a larger woman).

Our simulation studies suggest little bias for each of the methods used to address missing tumor characteristic information with the exception of the multiple imputation approach. Not surprisingly, in further sensitivity analyses, this bias was assuaged with the addition of improved predictors of characteristics with missing information (data not shown). Given these predictors are often unknown, the missing indicator approach may be the safest with respect to minimal bias, the smallest inflation in standard error and acceptable coverage probability.

In summary, though the analyses presented here relate specifically to breast cancer, general concepts of etiologic heterogeneity and correlated disease markers are documented in other cancers[9,10] and are not limited to neoplastic disease[11-14]. Therefore, with increasing interest in more individualized models of disease and the likely concomitant increase in missing information, the methods developed here may add important insights into more powerful ways to assess such etiologic heterogeneity.

Table 1. Frequency Distribution of ER, PR, HER2 status, Nurses' Health Study, 1980 – 2010

A			Complete Cases (1551 cases)		
ER	PR	HER2	Grade	Size	Count (%)
+	+	-	Well	small	663 (43)
+	+	-	Well	large	231 (15)
+	+	+	Well	small	80 (5)
+	-	-	Well	small	78 (5)
+	+	-	Poor	small	68 (4)
-	-	-	Poor	large	63 (4)
-	-	-	Poor	small	48 (3)
+	+	-	Poor	large	46 (3)
-	-	+	Poor	large	34 (2)
-	-	-	Well	small	33 (2)
+	-	-	Well	large	31 (2)
+	-	-	Poor	small	24 (2)
+	+	+	Well	large	23 (1)
+	-	-	Poor	large	17 (1)
+	+	+	Poor	large	16 (1)
-	-	+	Well	small	15 (1)
+	+	+	Poor	small	14 (1)
-	-	-	Well	large	13 (1)
+	-	+	Well	small	13 (1)
-	-	+	Well	large	12 (1)
-	-	+	Poor	small	8 (1)
+	-	+	Well	large	8 (1)
+	-	+	Poor	small	7 (<1)
+	-	+	Poor	large	6 (<1)
B			Considering Missing as a Category (top 24 classifications only; 4380 cases)		
ER	PR	HER2	Grade	Size	Count (%)
+	+	-	Well	small	663 (15)
+	+	Missing	Well	small	624 (14)
+	+	Missing	Missing	small	404 (9)
+	+	-	Well	large	231 (5)
+	+	Missing	Missing	Missing	166 (4)
+	+	Missing	Missing	large	146 (3)
+	+	Missing	Well	large	129 (3)
+	-	Missing	Well	small	115 (3)
+	-	Missing	Missing	small	102 (2)
+	+	Missing	Poor	small	101 (2)
-	-	Missing	Missing	small	93 (2)

-	-	Missing	Poor	small	85 (2)
-	-	Missing	Poor	large	83 (2)
+	+	+	Well	small	80 (2)
+	-	-	Well	small	78 (2)
+	+	-	Missing	small	75 (2)
+	+	Missing	Poor	large	69 (2)
+	+	-	Poor	small	68 (2)
-	-	-	Poor	Large	63 (1)
+	+	-	Missing	Missing	56 (1)
-	-	-	Poor	small	48 (1)
-	-	Missing	Missing	Missing	46 (1)
+	+	-	Poor	large	46 (1)
-	-	Missing	Well	small	45 (1)

Table 2: Multivariable-Adjusted Subtype Specific adjusted HR for Birth Index for each candidate method for adjusting for missing data
HR displayed denote a birth index of 102 (corresponding to 4 births at ages 20, 23, 26, 29 and age at menopause = 50) vs. nulliparous

Method of Adjustment	HR (95% CI)	HR (95% CI)	P_{het}
	ER Status		
	+	-	
Complete	0.70 (0.58, 0.85)	1.36 (0.77, 2.40)	0.05
Missing Indicator	0.70 (0.62, 0.78)	1.13 (0.80, 1.60)	0.01
IPW	0.69 (0.58, 0.84)	1.38 (0.81, 2.36)	0.03
Multiple Imputation	0.69 (0.62, 0.77)	1.10 (0.80, 1.50)	0.01
	PR Status		
	+	-	
Complete	0.81 (0.66, 1.00)	0.70 (0.47, 1.05)	0.58
Missing Indicator	0.81 (0.69, 0.94)	0.75 (0.60, 0.93)	0.96
IPW	0.81 (0.65, 1.00)	0.71 (0.49, 1.01)	0.57
Multiple Imputation	0.75 (0.66, 0.85)	0.74 (0.60, 0.92)	0.89
	HER2 Status		
	+	-	
Complete	0.91 (0.61, 1.35)	0.75 (0.63, 0.89)	0.39
Missing Indicator	0.78 (0.56, 1.09)	0.74 (0.62, 0.87)	0.76
IPW	0.89 (0.60, 1.32)	0.75 (0.63, 0.89)	0.43
Multiple Imputation	0.81 (0.66, 1.00)	0.73 (0.66, 0.81)	0.28
	Size		
	Small	Large	
Complete	0.82 (0.68, 1.00)	0.66 (0.45, 0.96)	0.32
Missing Indicator	0.79 (0.69, 0.91)	0.64 (0.50, 0.83)	0.17
IPW	0.83 (0.69, 1.00)	0.64 (0.44, 0.92)	0.24
Multiple Imputation	0.78 (0.70, 0.88)	0.65 (0.53, 0.79)	0.13
	Grade		
	Well	Poor	
Complete	0.83 (0.69, 1.01)	0.65 (0.48, 0.87)	0.16
Missing Indicator	0.75 (0.67, 0.83)	0.66 (0.54, 0.80)	0.10
IPW	0.83 (0.69, 1.01)	0.65 (0.48, 0.87)	0.17
Multiple Imputation	0.79 (0.70, 0.88)	0.65 (0.54, 0.78)	0.05

Additionally adjusted for all risk factors in Supplemental Table 1. - reproductive factors (duration of premenopause, duration following natural menopause, duration following bilateral oophorectomy, gynecologic age at first birth − age$_{first\ birth}$-age$_{menarche}$, birth index - $\sum_{i=1}^{s_t}(t^* - t_i)b_{it}$, where $b_{it} = 1$ if parity $\geq i$ at age t, = 0 else, t_i =age at ith birth), anthropometric and lifestyle characteristics (body mass index, height, alcohol intake, and use of hormones), and disease history (family history of breast cancer, benign breast disease).

Table 3: Multivariable-Adjusted Subtype Specific adjusted HR for BMI at Age 70 for each candidate method for adjusting for missing data

HR is for difference in BMI from age 50 to 70 (Postmenopausal BMI) of 8 units (BMI 30 vs. BMI 22)

Method of Adjustment	HR (95% CI)	HR (95% CI)	P_{het}
	ER Status		
	+	-	
Complete	1.09 (0.89, 1.33)	1.11 (0.60, 2.07)	0.95
Missing Indicator	1.21 (1.09, 1.36)	0.85 (0.59, 1.24)	0.03
IPW	1.09 (0.90, 1.33)	1.13 (0.64, 2.01)	0.92
Multiple Imputation	1.20 (1.08, 1.34)	0.91 (0.64, 1.28)	0.15
	PR Status		
	+	-	
Complete	1.34 (1.07, 1.67)	0.68 (0.44, 1.05)	0.01
Missing Indicator	1.43 (1.16, 1.76)	0.79 (0.62, 0.99)	<0.0001
IPW	1.34 (1.07, 1.67)	0.70 (0.47, 1.03)	0.01
Multiple Imputation	1.37 (1.21, 1.55)	0.77 (0.61, 0.97)	<0.0001
	HER2 Status		
	+	-	
Complete	1.26 (0.84, 1.90)	1.06 (0.87, 1.27)	0.44
Missing Indicator	1.22 (0.85, 1.76)	0.95 (0.79, 1.14)	0.23
IPW	1.25 (0.83, 1.88)	1.07 (0.89, 1.28)	0.49
Multiple Imputation	1.27 (1.03, 1.57)	1.12 (1.01, 1.24)	0.21
	Size		
	Small	Large	
Complete	0.91 (0.73, 1.12)	1.83 (1.30, 2.58)	0.001
Missing Indicator	1.05 (0.92, 1.21)	1.87 (1.48, 2.37)	<0.0001
IPW	0.92 (0.75, 1.13)	1.81 (1.29, 2.55)	0.002
Multiple Imputation	1.04 (0.93, 1.17)	1.49 (1.23, 1.79)	0.002
	Grade		
	Well	Poor	
Complete	1.02 (0.83, 1.26)	1.29 (0.95, 1.75)	0.21
Missing Indicator	1.16 (1.05, 1.29)	1.37 (1.14, 1.65)	0.02
IPW	1.04 (0.85, 1.27)	1.27 (0.94, 1.73)	0.28
Multiple Imputation	1.05 (0.93, 1.17)	1.33 (1.11, 1.58)	0.02

Additionally adjusted for all risk factors in Supplemental Table 1. - reproductive factors (duration of premenopause, duration following natural menopause, duration following bilateral oophorectomy, gynecologic age at first birth $-$ age$_{first\ birth}$-age$_{menarche}$, birth index - $\sum_{i=1}^{s_t}(t^* - t_i)b_{it}$, where $b_{it} = 1$ if parity $\geq i$ at age t, $= 0$ else, $t_i =$age at ith birth), anthropometric and lifestyle characteristics (body mass index, height, alcohol intake, and use of hormones), and disease history (family history of breast cancer, benign breast disease).

Table 4: Results of Simulation Study for Birth Index

Method		No missing disease characteristics				
		ER	PR	HER-2	Grade	Size
γ_{true}^{*}		-0.25	0.00	0.00	-0.12	0.00
Full Cohort	Bias (x 10^2)	0.29	-0.09	0.04	-0.05	-0.09
	95% Coverage Probability	94.4	95.9	94.5	94.8	94.3
		With Missing Disease Characteristics				
		No Missing		20% Missing Information		
		ER	PR	HER-2	Grade	Size
Complete case	Bias (x 10^2)	0.18	0.39	0.10	-0.02	0.08
	Standard Error (% change)	12.5	12.6	11.7	11.7	11.6
	95% Coverage Probability	94.4	95.2	94.1	95.4	94.6
IPW	Bias (x 10^2)	0.27	-0.14	0.07	-0.09	-0.13
	Standard Error (% change)	8.2	9.0	8.1	8.4	9.2
	95% Coverage Probability	93.0	93.9	92.2	94.0	93.7
Missing Indicator	Bias (x 10^2)	0.30	-0.33	0.15	-0.03	-0.05
	Standard Error (% change)	-0.7	0.5	9.0	3.3	1.4
	95% Coverage Probability	94.5	95.9	95.2	95.1	94.1
Multiple Imputation	Bias (x 10^2)	0.47	-0.01	-0.01	-0.49	-0.09
	Standard Error (% change)	0.80	0.20	9.5	5.0	1.5
	95% Coverage Probability	94.3	95.9	97.8	94.3	94.9
		No Missing		50% Missing Information		
		ER	PR	HER-2	Grade	Size
Complete case	Bias (x 10^2)	0.75	0.96	0.32	-0.22	0.55
	Standard Error (% change)	43.9	44.3	41.3	41.5	41.0
	95% Coverage Probability	94.3	95.3	94.4	95.0	95.3
IPW	Bias (x 10^2)	1.15	-0.85	0.15	-0.45	-0.12
	Standard Error (% change)	33.4	35.3	32.3	33.4	34.7
	95% Coverage Probability	90.0	93.2	90.9	92.4	93.8
Missing Indicator	Bias (x 10^2)	0.77	-0.47	0.32	-0.49	0.08
	Standard Error (% change)	-1.3	0.3	31.2	10.3	3.8
	95% Coverage Probability	94.5	95.7	94.1	94.1	94.8
Multiple Imputation	Bias (x 10^2)	0.86	-0.02	-0.01	-1.46	0.03
	Standard Error (% change)	2.0	0.6	21.9	12.1	4.1
	95% Coverage Probability	94.8	95.9	99.8	86.4	95.9

*γ_{true} includes a "common" parameter among subtypes of $\gamma_0 = -0.14$ – that is the effect when all subtypes are set to their referent values (i.e. a tumor of the subtype: ER negative, PR negative, HER2 negative, small and well-differentiated)

Table 5: Results of Simulation Study for Postmenopausal BMI

Method		No missing disease characteristics				
		ER	PR	HER-2	Grade	Size
γ^*_{true}		-0.20	0.65	0.14	0.00	0.50
Full Cohort	Bias (x 10^2)	0.01	0.05	0.08	-0.02	0.02
	95% Coverage Probability	94.9	95.0	94.5	95.6	95.3
		With Missing Disease Characteristics				
		No Missing		20% Missing Information		
		ER	PR	HER-2	Grade	Size
Complete case	Bias (x 10^2)	-0.06	0.86	0.20	0.01	0.48
	Standard Error (% change)	12.2	12.3	11.3	11.4	11.3
	95% Coverage Probability	94.2	93.8	94.9	95.5	94.3
IPW	Bias (x 10^2)	0.05	0.05	0.09	-0.10	0.03
	Standard Error (% change)	8.4	9.4	8.1	8.6	9.3
	95% Coverage Probability	93.0	93.9	92.2	94.0	93.7
Missing Indicator	Bias (x 10^2)	0.00	0.11	0.05	-0.15	-0.08
	Standard Error (% change)	-0.9	0.5	8.9	3.3	1.4
	95% Coverage Probability	95.3	95.5	95.2	94.8	94.6
Multiple Imputation	Bias (x 10^2)	0.14	0.21	-2.38	-1.41	-1.10
	Standard Error (% change)	0.8	0.2	9.3	4.8	1.4
	95% Coverage Probability	95.2	95.0	91.6	94.7	93.1
		No Missing		50% Missing Information		
		ER	PR	HER-2	Grade	Size
Complete case	Bias (x 10^2)	0.11	2.42	0.42	0.13	1.4
	Standard Error (% change)	42.7	43.1	40.0	40.3	39.9
	95% Coverage Probability	94.4	92.2	93.4	95.1	91.7
IPW	Bias (x 10^2)	0.19	-0.07	0.04	-0.02	-0.05
	Standard Error (% change)	34.0	36.3	32.6	33.7	35.0
	95% Coverage Probability	91.0	91.6	89.9	92.7	92.6
Missing Indicator	Bias (x 10^2)	0.21	0.09	-0.19	-0.28	-0.43
	Standard Error (% change)	-1.6	0.2	31.2	10.4	3.8
	95% Coverage Probability	95.3	95.5	95.2	94.8	94.6
Multiple Imputation	Bias (x 10^2)	0.20	0.22	-6.12	-2.94	-2.95
	Standard Error (% change)	1.8	0.4	21.5	11.5	3.9
	95% Coverage Probability	95.6	95.5	57.8	88.8	76.4

*γ_{true} includes a "common" parameter among subtypes of $\gamma_0 = -0.14$ – that is the effect when all subtypes are set to their referent values (i.e. a tumor of the subtype: ER negative, PR negative, HER2 negative, small and well-differentiated)

1. Rosner B, Glynn RJ, Tamimi RM, et al. Breast cancer risk prediction with heterogeneous risk profiles according to breast cancer tumor markers. *Am J Epidemiol.* 2013;178(2):296-308.

2. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Res.* 2006;8(4).

3. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat.* 2008;109(1):123-139.

4. Tamimi RM, Colditz GA, Hazra A, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. *Breast Cancer Res Treat.* 2012;131(1):159-167.

5. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics.* 1995;51(2):524-532.

6. Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J Womens Health.* 1997;6(1):49-62.

7. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med.* Jun 15 2005;24(11):1713-1723.

8. Chatterjee N, Sinha S, Diver WR, Feigelson HS. Analysis of cohort studies with multivariate and partially observed disease classification data. *Biometrika.* Sep 2010;97(3):683-698.

9. Morton LM, Sampson JN, Cerhan JR, et al. Rationale and Design of the International Lymphoma Epidemiology Consortium (InterLymph) Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr.* 2014;48:1-14.

10. Liao X, Lochhead P, Nishihara R, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. *N Engl J Med.* 2012;367(17):1596-1606.

11. Glynn RJ, Rosner B, Christen WG. Evaluation of risk factors for cataract types in a competing risks framework. *Ophthalmic Epidemiol.* 2009;16(2):98-106.

12. Sandhu RK, Conen D, Tedrow UB, et al. Predisposing factors associated with development of persistent compared with paroxysmal atrial fibrillation. *J Am Heart Assoc.* 2014;3(3):000916.

13. Schiess MC, Suescun J. Clinical Determinants of Progression of Parkinson Disease: Predicting Prognosis by Subtype. *JAMA Neurol.* 2015;15(10).

14. Goutman SA, Feldman EL. Clinical Trials of Therapies for Amyotrophic Lateral Sclerosis: One Size Does Not Fit All. *JAMA Neurol.* 2015;11(10)