Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kimberly A. Barker Date

CYP2C19 Polymorphisms and Breast Cancer Recurrence in a Heterogeneous Population of Tamoxifen-Treated Women

By

Kimberly A. Barker

Master of Public Health

Epidemiology

_____

Timothy L. Lash

Committee Chair

CYP2C19 Polymorphisms and Breast Cancer Recurrence in a Heterogeneous Population of Tamoxifen-Treated Women

By

Kimberly A. Barker

B.S.

Massachusetts Institute of Technology

2013

Thesis Committee Chair: Timothy L. Lash, Dsc, MPH

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in epidemiology

2015

Abstract

CYP2C19 Polymorphisms and Breast Cancer Recurrence in a Heterogeneous Population of Tamoxifen-Treated Women

By Kimberly A. Barker

Though tamoxifen is a standard treatment to prevent recurrences in women suffering from breast cancer, the drug is still not effective for 20 to 30% of patients receiving it. The important role of cytochrome P450 (CYP) proteins in the metabolism of tamoxifen has led to speculation that polymorphisms in CYP2C19, among other P450s, may contribute to the observed variations in treatment efficacy. This study utilized the publicly available data of the International Tamoxifen Pharmacogenomics Consortium to assess whether the CYP2C19*2 loss-of-function and CYP2C19*17 gain-of-function variants are associated with disease-free survival time in 2,102 female breast cancer patients prescribed 2 or 5 years of adjuvant tamoxifen therapy. Kaplan-Meier curves and Cox proportional hazards analyses showed that observed associations between CYP2C19*2 or CYP2C19*17 genotypes and disease-free survival were not substantial and were imprecise. Carriers of one or two CYP2C19*2 alleles had adjusted hazard ratios of 1.05 (95% CI: 0.78, 1.42) and 0.79 (95% CI: 0.32, 1.94) compared to non-carriers, respectively. Carriers of one or two CYP2C19*17 alleles had adjusted hazard ratios of 1.02 (95% CI: 0.71, 1.46) and 0.57 (95% CI: 0.26, 1.24) compared to non-carriers, respectively. The lack of a noteworthy association was robust to stratification by CYP2D6 genotype or by menopausal status, and the null results were further confirmed by imputation of missing data. This study contains one of the largest sample sizes of any examining the association between CYP2C19 genotype and tamoxifen efficacy, and addresses the conflicting prior findings regarding this topic. The results of this study indicate that CYP2C19 genotype is not a useful predictor of the risk of recurrence with tamoxifen therapy, and future research should focus on alternative methods to improve the outlook for breast cancer patients.

CYP2C19 Polymorphisms and Breast Cancer Recurrence in a Heterogeneous Population of Tamoxifen-Treated Women

By

Kimberly Barker

B.S.

Massachusetts Institute of Technology

2013

Thesis Committee Chair: Timothy L. Lash, Dsc, MPH

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in epidemiology

2015

Introduction

Breast cancer is the most common cancer among women in the United States (excluding skin cancers), with approximately 232,000 cases projected to be diagnosed in 2015 (1). For patients with estrogen receptor positive breast cancer, five years of adjuvant treatment with the selective estrogen receptor modulator tamoxifen has long been shown to reduce recurrences by nearly 50% (2, 3). Currently, tamoxifen is the standard treatment for premenopausal women with estrogen receptor positive breast cancer and is used in postmenopausal women as well, often in addition to an aromatase inhibitor (4). Tamoxifen and its metabolites competitively inhibit the binding of estrogen to estrogen receptors, thereby inhibiting expression of genes regulated by estrogen, including growth and angiogenic factors, in tumor cells (5, 6). However, patient responses to tamoxifen vary, and around 20% to 30% of patients receiving 5 years of this therapy still experience a breast cancer recurrence (7).

The metabolites of tamoxifen most actively bind the estrogen receptor and suppress estrogen-dependent tumor cell proliferation are 4-hydroxytamoxifen (4-OH-TAM) and 4-hydroxy-N-desmethyl tamoxifen (endoxifen) (5, 8, 9). While tamoxifen and its major primary metabolite, N-desmethyl tamoxifen, both exhibit a weak affinity for the estrogen receptor (10), 4-OH-TAM and endoxifen have 100-fold higher estrogen receptor affinities and are 30 to 100-fold more effective in suppressing tumor cell proliferation (11). Additionally, because of the unique ability of endoxifen to decrease ERα protein levels (12), and the fact that its average plasma concentration is over 10-fold higher than that of 4-OH-TAM (11), endoxifen is likely the metabolite contributing the most to successful treatment with tamoxifen. Levels of tamoxifen and its metabolites vary extensively between patients (9). Endoxifen and the other metabolites of tamoxifen are produced via the action of cytochrome P450 (CYP) enzymes. Therefore, intrinsic patient expression levels of certain CYP proteins, among other factors, may play a role in the observed variation in tamoxifen metabolite levels and treatment efficacy.

Due to the key role of CYP2D6 in tamoxifen metabolism, especially in the hydroxylation of N-desmethyl tamoxifen to endoxifen, inter-patient variation in functional levels of this enzyme has been the subject of extensive research. Non-functional alleles of CYP2D6 or the use of CYP2D6-inhibiting drugs can result in a poor-metabolizer or intermediate-metabolizer phenotype, corresponding to lower levels of plasma endoxifen compared to levels in extensive-metabolizers and ultrarapid metabolizers (9, 13-16). Additionally, CYP2D6 poor and intermediate metabolizer phenotypes have been associated with less favorable clinical outcomes in breast cancer patients treated with tamoxifen (17-22). However, null or near-null associations have been found in other studies (23-26), leading to ongoing controversy over the value of using CYP2D6 genotyping to guide the prescription of tamoxifen (27-29). Currently, the American Society of Clinical Oncology recommends against using CYP2D6 genotype to guide decisions regarding prescribed tamoxifen regimens (30).

The enzyme CYP2C19 functions in the metabolism of numerous drugs, including antiplatelets, antidepressants, and proton pump inhibitors (31). For many of these drugs, guidelines for prescription doses based on CYP2C19 genotype exist (32). CYP2C19 also plays a role in the metabolism of tamoxifen but is less well studied with regards to this pathway than CYP2D6 due to the more minor metabolic contributions of CYP2C19 (11). CYP2C19 contributes to the metabolism of tamoxifen to both primary metabolites (33-37) and in the conversion of 4-OH-TAM to endoxifen (Figure 1) (38). Specific inhibition of CYP2C19 in human liver microsomes decreases the rate of endoxifen formation, and the enzymatic activity of CYP2C19 positively correlates with the ratio of 4-OH-TAM to tamoxifen, indicating that CYP2C19 may contribute substantially in the metabolism of tamoxifen despite its reportedly minor roles (39, 40). However, other studies have indicated that CYP2C19 genotype variants are not associated with altered plasma levels of tamoxifen and its metabolites (41, 42).

The CYP2C19 gene is highly polymorphic. According to the star nomenclature for CYP single nucleotide polymorphisms (SNPs), CYP2C19*1 designates the wild-type allele (43), with individuals homozygous for the CYP2C19*1 allele phenotypically considered extensive metabolizers (44). Loss of enzyme activity is associated with CYP2C19*2 (defined by a 681G>A substitution, rs4244285) (43, 45) and CYP2C19*3 (defined by a 636G>A substitution, rs4986893) (43, 46). The *2 allele is found at a frequency of approximately 23-39% in Asians, 10-20% in Caucasians, and 15% in Africans (38, 47). The *3 allele, on the other hand, is only common among Asians, with an allele frequency of 5-10% in the Asian population (38). Carriers of one loss-of-function allele are designated intermediate metabolizers (48, 49), and those with two loss-of-function alleles are considered poor metabolizers (31, 44).

More recently, CYP2C19*17 (defined by two polymorphisms: -806C>T, rs12248560; and -3402C>T, rs11188072) has been characterized as leading to enhanced transcription of the gene (43, 50). The *17 allele is found at a frequency of approximately 4% in Asians (31, 38, 47) but is more common among Caucasians and Africans, occurring with a frequency of 18-24% in both groups (21, 31, 38, 47). Many studies consider both CYP2C19*17 heterozygotes and homozygotes to phenotypically be ultrarapid metabolizers (44, 51), though some studies claim that assignment of CYP2C19*17 carriers to an extensive metabolizer phenotype is sufficient (31, 52). The metabolizer phenotype of individuals possessing one loss-of-function allele and one *17 allele is also unclear (31, 53). While the phenotypic assignments described above for CYP2C19*2, *3, and *17 are used in studies of tamoxifen efficacy based on CYP2C19 genotype (49, 54), the correspondence of these phenotypes to the metabolism of tamoxifen, specifically, has not been decisively shown in the literature.

Though fewer studies have examined the association of CYP2C19 genotype with differential outcomes in tamoxifen-treated patients than have done so for CYP2D6, similarly conflicting results have arisen in the literature, as reviewed by Binkhorst et al. (55) . In some

studies, the presence of the CYP2C19*2 allele has been associated with longer relapse-free time or better survival in tamoxifen treated women (49, 56, 57), while other research has yielded a null result for this association (21, 54, 58-61). Interestingly, in breast cancer patients not treated with tamoxifen, the CYP2C19*2 allele has been found to correlate with a worse prognosis (49, 56), though a null association in these patients has also been reported (57).

Although the CYP2C19*17 allele results in a metabolizer phenotype opposite that of CYP2C19*2, the *17 allele has also been associated with more favorable outcomes in breast cancer patients treated with tamoxifen (21), though null results have also been found (54, 56, 60, 62). Contradictory results were obtained in the context of monotherapeutic tamoxifen use for advanced breast cancer, where an association between the *17 allele and shorter times to treatment failure was reported (49, 57). Adding to the complex results, a protective effect of the *17 allele against breast cancer risk has been observed in individuals not using tamoxifen (63), and an association between carrying the allele and longer disease free survival has also been reported in ER+ breast cancer patients not treated with tamoxifen (57). Other studies have found no effect of the *17 allele on clinical outcomes in untreated patients (21, 56).

In an attempt to clarify the contradictory results already published regarding this topic, the current study used a large publicly available dataset to investigate the association of CYP2C19*2 and CYP2C19*17 variants with breast cancer recurrence in both pre and postmenopausal women treated with adjuvant tamoxifen therapy for ER+ breast cancer.

Methods

Data source and study population

Data were obtained from the International Tamoxifen Pharmacogenomics Consortium (ITPC) (64). This group is comprised of research projects from 12 sites representing 9 countries, all designed to prospectively assess potential associations between genes involved in tamoxifen metabolism or transport and breast cancer recurrence risk. Of the 4,969 tamoxifen-treated female breast cancer patients included in the data, 2,102 met the criteria for study inclusion (Figure 2). These criteria required that patients were prescribed an intended dose of 20 mg/day tamoxifen for an intended duration of either 2 or 5 years, were not known to previously have received systemic therapy (e.g. tamoxifen or raloxifene) for breast cancer prevention, had no known history of invasive or in situ breast cancer occurrence, were not known to have used a first adjuvant therapy other than tamoxifen, and had a length of time between breast cancer surgery and initial tamoxifen use that was not known to be longer than 182 days (approximately six months). Additionally, only those patients with non-metastatic estrogen receptor positive tumors (93.2% of tumors in the ITPC data were estrogen receptor positive) for whom the data contained some information on at least one of the genotypes of interest, whether a recurrence occurred, and follow-up time were included. After restriction, studies from 7 sites included in the ITPC contained eligible data. These data came from: Belgium (site 2); Indiana, USA (site 4); Manchester, UK (site 6); Tuebingen, Germany (site 8); San Francisco, USA (site 9); Dundee, UK (site 11); and Korea (site 12).

Analytic variables

Disease-free survival time (DFS), the outcome of interest, was measured as the number of months from diagnosis until a breast cancer recurrence, defined as an ipsolateral local or regional recurrence (invasive or non-invasive), a distant recurrence, or a contralateral breast cancer (invasive or non-invasive). Patients who did not experience a recurrence were censored on

the date of death from another cause or on the day of last disease-free evaluation. One individual who did not have a recurrence and was missing data on the last disease-free evaluation was censored on the date she was last known to be alive.

The genotypes of interest, representing exposure variables, were CYP2C19*2 (SNP 681G>A) and CYP2C19*17 (SNP -806C>T). Various methods of genotyping were used in the seven studies comprising the data, with the majority of genotypes (60.5%) ascertained by the AmpliChip test platform (Roche Molecular Diagnostics, California, USA). In three instances where the blood genotype data variable did not match the CYP2C19*2 genotype variable, preference was given to the AmpliChip blood genotype data due to the very high sensitivity and specificity of this test (65). The CYP2C19*3 allele was not assessed in this study because no variants were detected in the included data.

Covariables

Potential covariates of interest were: age, race, menopausal status, tumor grade and stage, progesterone receptor status, use of other adjuvant therapies (radiation, chemotherapy, and aromatase inhibitors), and CYP2D6 metabolizer phenotype. One moderate CYP2C19 inhibitor, Fluoxetine, was assessed in the ITPC data, however it was not considered in this study because data on its use were missing in nearly 70% of included patients. Age (single-year), menopausal status (pre, post, or peri), progesterone receptor status, use of other adjuvant therapies, and Nottingham tumor grades were recorded directly in the ITPC dataset. Perimenopausal women (n = 57) were combined with post-menopausal women for all analyses. Racial categories as defined by the Office of Management and Budget (66) were used to divide patients into three racial groups: white, asian or pacific islander, and any other race.

Tumor stage was derived from information on both tumor diameter and the number of positive-testing lymph nodes. A lack of information on in situ tumors and distant metastases

prohibited use of the TNM staging system, however the primary tumor and pathologic guidelines of the TNM system (67) were used to classify tumors into five stages. Stage 1 was comprised of tumors less than or equal to 20 mm in diameter with no observed metastases to lymph nodes. Stage 2A tumors were either less than or equal to 20 mm with 1 to 3 lymph node metastases or were greater than 20 mm but less than or equal to 50 mm in diameter with no lymph node metastases. Stage 2B was comprised of tumors either greater than 20 mm but less than or equal to 50 mm in size with 1 to 3 lymph node metastases or tumors larger than 50 mm with no lymph node metastases. Tumors included in stage 3A included: those less than or equal to 20 mm in size with 4 to 9 lymph node metastases, those greater than 20 mm but less than or equal to 50 mm in diameter with 4 to 9 lymph node metastases, those larger than 50 mm with 1 to 3 lymph node metastases, and those larger than 50 mm with 4 to 9 lymph node metastases. Finally, stage 3C consisted of all tumors with 10 or more lymph node metastases, regardless of tumor size. As no women with distant metastases were included in this analysis, no stage 4 classification was used.

A variable encoding individuals' CYP2D6 genotype in terms of metabolizer status (ultra-UM, extensive- EM, intermediate- IM, or poor- PM) was availabe in the ITPC data, and accounted for both genetic factors and the use of CYP2D6 inhibiting drugs. This information was used to generate a variable describing metabolizer level as a function of both CYP2D6 phenotypes and CYP2C19 genotypes as done by Schroth et al. (Table 1) (21).

Statistical analyses

Descriptive analyses including all covariates of interest were conducted for all included women, the subset who experienced a breast cancer recurrence, and the subset who did not experience a recurrence. Survival curves for disease-free survival stratified by different combinations of CYP2C19 genotypes and by metabolic level as a function of both CYP2C19 genotypes and CYP2D6 phenotypes were generated using the Kaplan-Meier method, and a log-rank test was used to estimate the extent of differences between genotype or phenotype groups.

A directed acyclic graph was drawn to assess the potential for confounding of the association between CYP2C19 genotypes and breast cancer recurrence rate (Figure 3). This graph indicated that all potential confounding of the association could be eliminated by controlling for race, assuming the directed acyclic graph is correct. However, because several of the hypothesized associations in the graph are not well characterized, a model-based approach to confounding assessment was also performed. The adjuvant aromatase inhibitor therapy and tumor grade variables were found to violate the proportional hazards assumption when assessed using log-log survival curves, and were therefore excluded from all models. The initial cox proportional hazards model for each CYP2C19 genotype of interest contained indicator variables for genotype as well as terms for age, menopausal status, tumor stage, chemotherapy use, radiation use, and progesterone receptor status. The intial model for CYP2C19*2 also included indicator variables for white and Asian race, which were excluded from the CYP2C19*17 model because all individuals with information on this variant were white. An all-possible-subsets analysis was done, with confounding indicated in models where the removal of that combination of variables led to a hazard ratio changed by greater than 10% of the hazard ratio for the initial model. This analysis indicated that race and stage were important variables to control. Combining the information obtained from the directed acyclic graph, the all-possible-subsets analysis, and including age as a traditionally modeled covariate led to final models for Cox proportional hazards analyses that included age at diagnosis of primary breast cancer, tumor stage, and race (white or Asian, for CYP2C19*2 only) as covariates. These models were applied to compute hazard ratios for recurrence and corresponding 95% confidence intervals. Analyses stratified by CYP2D6 phenotype and menopausal status were also performed.

For the main multivariate analyses, individuals with missing values for an included variable were excluded. To assess the potential for bias due to the use of complete case analyses, a supplemental analysis invoking the missing data imputation function of SAS was used to

impute missing values of CYP2C19*2 genotype, CYP2C19*17 genotype, race, age at breast cancer diagnosis, and tumor stage. Data were assumed to be missing at random in an arbitrary pattern. The imputation model contained the imputed variables listed above as well as the indicator of breast cancer recurrence and the Nelson-Aalen estimator of disease hazard (H (T)), as suggested by White and Royston (68). Imputed values were drawn using the multivariate normal approach via the Markov Chain Monte Carlo (MCMC) algorithm. Five imputations were run after 200 burn-in iterations to establish convergence to the stationary distribution. Parameters of successive iterations were visually inspected to confirm the model had reached a stationary distribution. Autocorrelation between imputed datasets was assessed via visual inspection of autocorrelation plots. The same Cox proportional hazards models described above were applied to each of the five imputed datasets to compute hazard ratios for recurrence and corresponding 95% confidence intervals.

All analyses were carried out in SAS version 9.4 (Cary, NC).

Results

Study population

The seven sites containing eligible patients provided 2,102 women for analysis. Of these women, a total of 296 experienced a breast cancer recurrence. Patient characteristics by site are given in Table 2 for numbers of recurrences, median survival time, average age at diagnosis, menopausal status, genotype distributions, and source of genetic material. Characteristics of the pooled data are presented in Table 3 for all women and separately for the recurrence and no recurrence groups, describing the age at diagnosis, race, menopausal status, intended tamoxifen duration, tumor grade, tumor stage, progesterone receptor status, additional therapy use, CYP2D6 phenotypes, and CYP2C19 genotypes for each group. The median disease-free survival time was 61 months for all women, 45 months for women experiencing a recurrence, and 63 months for women with no recurrence.

Genotypes

Data on CYP2C19*2 were available from all seven included sites for 2,055 women, and data on CYP2C19*17 were reported from sites 2, 4, and 8, for 1,253 women. A total of 80 CYP2C19*2 homozygotes, 643 CYP2C19*2 heterozygotes, 86 CYP2C19*17 homozygotes, and 469 CYP2C19*17 heterozygotes were included in this study. Overall minor allele frequencies of CYP2C19*2 and CYP2C19*17 were 17% and 24%, respectively. Hardy-Weinberg chi-squared statistics and corresponding p values within each site for the CYP2C19*2 and CYP2C19*17 alleles are provided in Table 2. Both variants were in Hardy-Weinberg equilibrium for each site at which they were measured, except for CYP2C19*2 at site 9 ($p = 0.02$), which was observed to have a minor allele frequency of 18.5%, a value within the range expected for the population of site 9 (77% white, 21% Asian, data not shown), based on previous reports (38, 47).

CYP2C19 genotypes and disease-free survival

All log-rank test statistics for Kaplan-Meier analyses of both the CYP2C19*2 and CYP2C19*17 alleles revealed minimal differences between the disease-free survival curves for different genotypes (Figures 4 and 5). This was true regardless of the comparison made, and held for variant homozygotes vs. variant heterozygotes vs. non-carriers (Figures 4a and 5a), for variant allele carriers vs. non-carriers (Figures 4b and 5b), and for variant heterozygotes and homozygotes combined vs. wild-type individuals (Figures 4c and 5c). No Kaplan-Meier curves were markedly different when analyses involving the CYP2C19*17 genotype were restricted to only those sites providing genotype information on CYP2C19*17 (data not shown).

For the CYP2C19*2 allele, the cox proportional hazards ratios for the association of variant heterozygotes and homozygotes with disease-free survival, adjusted for age at diagnosis, tumor stage, and race, were 1.05 (95% CI: 0.78, 1.42) and 0.79 (95% CI: 0.32, 1.94), respectively (Table 4). For the CYP2C19*17 allele, the cox proportional hazards ratios for the association of variant heterozygotes and homozygotes with disease-free survival, adjusted for age at diagnosis and tumor stage, were 1.02 (95% CI: 0.71, 1.46) and 0.57 (0.26, 1.24), respectively (Table 4). Stratification by menopausal status and CYP2D6 phenotype also did not yield any notable associations between CYP2C19 genotype and disease-free survival for either CYP2C19 allele (Table 5). Restriction of the cox proportional hazards analyses for CYP2C19*17 to sites providing genotype information on the *17 allele did not yield any substantially different results (data not shown). The cox proportional hazards ratios observed in the imputed data sets were not substantially changed from those found in the complete case analysis but in general tended to be more precise and closer to the null (Supplemental Tables S1 and S2). Autocorrelation between imputed datasets was observed for the CYP2C19*17 genotype variable.

CYP2D6 phenotype/CYP2C19 genotype combinations

When CYP2D6 phenotypes are considered in combination with CYP2C19 genotypes, Kaplan-Meier analyses show that the visual distinction between these survival curves for CYP2C19*17 is more pronounced than for the CYP2C19*17 allele alone, with the group designated as level 3 in Table 1 having lower disease-free survival probabilities than the other two groups (Figure 6a). The log-rank test statistic for the difference between the survival curves for the combination of CYP2D6 phenotypes and CYP2C19*2 genotypes yielded a p-value of 0.0036 (Figure 6b). The group combining CYP2D6 homozygous extensive metabolizers who possess a *2 allele with CYP2D6 heterozygous extensive metabolizers with no *2 allele (level 2) appears to have the lowest probability of disease-free survival at all times. When CYP2D6 phenotypic groups are compared independently of CYP2C19 genotypes, CYP2D6 heterozygous extensive metabolizers are observed to have the lowest probability of disease-free survival at all times, though the resulting Kaplan-Meier survival curves are not as appreciably different between the phenotype groups as in the analysis also considering CYP2C19 genotypes (data not shown).

For the metabolizer levels assigned to CYP2D6 phenotype/CYP2C19*2 genotype combinations, the cox proportional hazards ratios for the association of levels 1 and 2 with disease-free survival, adjusted for age at diagnosis, tumor stage, and race, were 0.78 (95% CI: 0.52, 1.15) and 1.19 (95% CI: 0.86, 1.66), respectively. For the CYP2D6 phenotype/CYP2C19*17 genotype combinations, the cox proportional hazards ratios for the association of levels 1 and 2 with disease-free survival, adjusted for age at diagnosis and tumor stage, were 0.70 (95% CI: 0.43, 1.14) and 0.96 (95% CI: 0.73, 1.25).

Discussion

Here it has been shown that any associations between CYP2C19*2 or CYP2C19*17 genotypes and disease-free survival in tamoxifen-treated breast cancer patients in the ITPC dataset are highly imprecise and likely a product of random error. Secondary Kaplan-Meier analyses indicated that CYP2C19*2 genotypes in the context of CYP2D6 metabolizer phenotypes may produce notably different survival curves, especially for the group combining CYP2D6 extensive metabolizer homozygotes carrying a *2 allele with CYP2D6 extensive metabolizer heterozygotes not carrying a *2 allele. However, in a cox proportional hazards analysis of the combined CYP2D6/CYP2C19*2 groups, the observed associations were marginal and again imprecise.

This study takes advantage of the large sample size offered by the ITPC data and includes a large number of CYP2C19 variants compared to the variant sample size of most prior studies on the association of interest here. This study also benefits from the inclusion of a substantial number of premenopausal patients, allowing for an analysis of the effect of CYP2C19 genotype variation in both pre- and postmenopausal women. To our knowledge, only two other studies have examined the association between CYP2C19 genotype and breast cancer recurrence stratified by menopausal status (58, 60), and those studies included a combined total of only 85 premenopausal patients. The inclusion of a substantial premenopausal cohort is especially relevant as tamoxifen is the primary adjuvant treatment used in this group.

At sites not testing for the CYP2C19*17 allele, misclassification could have occurred if carriers were mistakenly assigned a wild-type genotype. To address this possibility, sub-analyses within sites testing for CYP2C19*17 were performed, and did not provide substantially different results. Because the CYP2C19*3 allele was not investigated in this study, some misclassification could have also occurred if carriers of this allele were labeled as wild-type. However, this allele is only expected to be common at site 12, and therefore its absence in these analyses is not likely to

have influenced the overall results. The lack of ability to control for CYP2C19 inhibitor use could have contributed to some amount of bias in observed associations. An additional study limitation is that CYP2C19 genotyping using tumor-derived DNA, as done at three of the seven included sites, may not provide optimal genotyping results due to the potential for loss-of-heterozygosity in tumor cells (69, 70). However, results of testing for Hardy-Weinberg equilibrium indicate that any loss-of-heterozygosity had a minimal impact on observed CYP2C19 genotypes in this study.

While previous reports have found the presence of CYP2C19*2 to be associated with elevated efficacy of tamoxifen treatment, this study agrees with the findings of other research that has reported no such association. The hazard ratio obtained here for the association of CYP2C19*17 homozygotes with a more favorable disease-free survival (HR = 0.57, 95% CI: 0.26, 1.24) is similar to the ratio found in an influential study by Schroth et al. for the association of carrying CYP2C19*17 with relapse-free time (HR = 0.45, 95% CI: 0.21, 0.92) (21). Just over 40% of the patient population and the majority of *17 allele data in the ITPC dataset was contributed by Schroth et al., so similarities in results may be expected. However, the comparison made in our study that best corresponds to that made by Schroth, in which CYP2C19*17 homozygotes and heterozygotes were combined and compared to non-carriers, is 0.94 (95% CI = 0.67, 1.32, data not shown). The present study contains more *17 variants that that of Schroth, lending credence to other studies which have found a null association between the presence of *17 and breast cancer outcomes, and indicating that CYP2C19 genotypes may not be useful as determinants of which patients are most likely to benefit from tamoxifen treatment. Our findings for the *17 allele are consistent with those of a very similar, though smaller, study by Moyer et al., which reported a hazard ratio for the association a *17 allele with disease-free survival of 0.93 (95% confidence interval 0.64-1.37), and also found a near-null association among those with impaired CYP2D6 (62). Despite the biologic plausibility of CYP2C19 playing an important role

in patients with reduced CYP2D6 function, the stratified analyses performed here also presented no evidence that this is the case.

Because no patients who were not treated with tamoxifen are included in the ITPC data, we were unable to compare the effects of these variants in those treated with tamoxifen to the effects in those not treated. Several other studies have indicated that CYP2C19 variants are associated with differences in baseline breast cancer risk due to the inherent role of CYP2C19 in the metabolism of estrogen. However, this association has not been consistently observed in studies exploring it, so it is unclear whether this issue could have impacted our results. Additionally, we did not examine advanced breast cancer patients, so previously reported results of an association between CYP2C19*2 alleles and a longer time to treatment failure on tamoxifen among these patients could not be evaluated here.

A study conducted by Province et al. using the ITPC data to assess an association between CYP2D6 genotype and breast cancer recurrence reported poorer invasive disease-free survival among CYP2D6 poor metabolizers and a weak association between poor metabolizer status and breast cancer-free interval (71). However, these associations were not robust to variations in inclusion criteria, and the research has been critiqued for its reliance on statistical interpretations of ad hoc subset analyses (72). In light of these criticisms, the criteria for inclusion in this study were defined *a priori*. Province et al. also showed the wide variety of results obtained within each of the study sites, which is an additional challenge in interpreting the results of our study. However, the ITPC data seem to firmly indicate against the use of CYP2C19 genotyping alone in decision making for tamoxifen prescriptions. Beelen et al. detail alternative strategies to predict tamoxifen efficacy, and highlight the therapeutic drug monitoring approach, in which individuals' levels of endoxifen are directly monitored and used to guide tamoxifen dosing (56). Effective implementation of this approach would require standardization of endoxifen analytical assays, as well as an agreed upon therapeutic window for endoxifen in the

prevention of breast cancer recurrence. Given the contradictory results observed for CYP2D6 and CYP2C19 regarding tamoxifen metabolism, future work should consider focusing on alternative approaches such as therapeutic drug monitoring, rather than repeating previous studies in an attempt to find one genetic key to determining tamoxifen efficacy.

References

1. American Cancer Society. What are the key statistics about breast cancer? ; 2015. (http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics). (Accessed 1/5 2015).
2. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. . *Lancet* 1998;351(9114):1451-67.
3. Early Breast Cancer Trialists' Collaborative G. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;365(9472):1687-717.
4. Burstein HJ, Temin S, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: american society of clinical oncology clinical practice guideline focused update. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2014;32(21):2255-69.
5. Johnson MD, Zuo H, Lee KH, et al. Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast cancer research and treatment* 2004;85(2):151-9.
6. Osborne CK. Tamoxifen in the treatment of breast cancer. *The New England journal of medicine* 1998;339(22):1609-18.
7. Hackshaw A, Roughton M, Forsyth S, et al. Long-term benefits of 5 years of tamoxifen: 10-year follow-up of a large randomized trial in women at least 50 years of age with early breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;29(13):1657-63.
8. Furr BJ, Jordan VC. The pharmacology and clinical uses of tamoxifen. *Pharmacology & therapeutics* 1984;25(2):127-205.

9. Stearns V, Johnson MD, Rae JM, et al. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *Journal of the National Cancer Institute* 2003;95(23):1758-64.

10. Wakeling AE, Slater SR. Estrogen-receptor binding and biologic activity of tamoxifen and its metabolites. *Cancer treatment reports* 1980;64(6-7):741-4.

11. Klein DJ, Thorn CF, Desta Z, et al. PharmGKB summary: tamoxifen pathway, pharmacokinetics. *Pharmacogenetics and genomics* 2013;23(11):643-7.

12. Wu X, Hawse JR, Subramaniam M, et al. The tamoxifen metabolite, endoxifen, is a potent antiestrogen that targets estrogen receptor alpha for degradation in breast cancer cells. *Cancer research* 2009;69(5):1722-7.

13. Jin Y, Desta Z, Stearns V, et al. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *Journal of the National Cancer Institute* 2005;97(1):30-9.

14. Owen RP, Sangkuhl K, Klein TE, et al. Cytochrome P450 2D6. *Pharmacogenetics and genomics* 2009;19(7):559-62.

15. Saladores P, Murdter T, Eccles D, et al. Tamoxifen metabolism predicts drug concentrations and outcome in premenopausal patients with early breast cancer. *The pharmacogenomics journal* 2014.

16. Borges S, Desta Z, Li L, et al. Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. *Clinical pharmacology and therapeutics* 2006;80(1):61-74.

17. Goetz MP, Knox SK, Suman VJ, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast cancer research and treatment* 2007;101(1):113-21.

18. Goetz MP, Rae JM, Suman VJ, et al. Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *Journal of clinical*

oncology : official journal of the American Society of Clinical Oncology 2005;23(36):9312-8.

19. Kiyotani K, Mushiroda T, Imamura CK, et al. Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010;28(8):1287-93.
20. Lim HS, Ju Lee H, Seok Lee K, et al. Clinical implications of CYP2D6 genotypes predictive of tamoxifen pharmacokinetics in metastatic breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007;25(25):3837-45.
21. Schroth W, Antoniadou L, Fritz P, et al. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007;25(33):5187-93.
22. Schroth W, Goetz MP, Hamann U, et al. Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *Jama* 2009;302(13):1429-36.
23. Lash TL, Cronin-Fenton D, Ahern TP, et al. CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. *Journal of the National Cancer Institute* 2011;103(6):489-500.
24. Nowell SA, Ahn J, Rae JM, et al. Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients. *Breast cancer research and treatment* 2005;91(3):249-58.
25. Rae JM, Drury S, Hayes DF, et al. CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *Journal of the National Cancer Institute* 2012;104(6):452-60.

26. Regan MM, Leyland-Jones B, Bouzyk M, et al. CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the breast international group 1-98 trial. *Journal of the National Cancer Institute* 2012;104(6):441-51.

27. Brauch H, Schroth W, Goetz MP, et al. Tamoxifen use in postmenopausal breast cancer: CYP2D6 matters. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013;31(2):176-80.

28. Hertz DL, McLeod HL, Irvin WJ, Jr. Tamoxifen and CYP2D6: a contradiction of data. *The oncologist* 2012;17(5):620-30.

29. Lash TL, Lien EA, Sorensen HT, et al. Genotype-guided tamoxifen therapy: time to pause for reflection? *The Lancet Oncology* 2009;10(8):825-33.

30. Burstein HJ, Griggs JJ, Prestrud AA, et al. American society of clinical oncology clinical practice guideline update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *Journal of oncology practice / American Society of Clinical Oncology* 2010;6(5):243-6.

31. Li-Wan-Po A, Girard T, Farndon P, et al. Pharmacogenetics of CYP2C19: functional and clinical implications of a new variant CYP2C19*17. *British journal of clinical pharmacology* 2010;69(3):222-30.

32. PharmGKB. CYP2C19 Dosing Guidelines. (http://www.pharmgkb.org/gene/PA124). (Accessed January 7 2015).

33. Boocock DJ, Brown K, Gibbs AH, et al. Identification of human CYP forms involved in the activation of tamoxifen and irreversible binding to DNA. *Carcinogenesis* 2002;23(11):1897-901.

34. Coller JK, Krebsfaenger N, Klein K, et al. The influence of CYP2B6, CYP2C9 and CYP2D6 genotypes on the formation of the potent antioestrogen Z-4-hydroxy-tamoxifen in human liver. *British journal of clinical pharmacology* 2002;54(2):157-67.

35. Crewe HK, Ellis SW, Lennard MS, et al. Variable contribution of cytochromes P450 2D6, 2C9 and 3A4 to the 4-hydroxylation of tamoxifen by human liver microsomes. *Biochemical pharmacology* 1997;53(2):171-8.

36. Crewe HK, Notley LM, Wunsch RM, et al. Metabolism of tamoxifen by recombinant human cytochrome P450 enzymes: formation of the 4-hydroxy, 4'-hydroxy and N-desmethyl metabolites and isomerization of trans-4-hydroxytamoxifen. *Drug metabolism and disposition: the biological fate of chemicals* 2002;30(8):869-74.

37. Dehal SS, Kupfer D. CYP2D6 catalyzes tamoxifen 4-hydroxylation in human liver. *Cancer research* 1997;57(16):3402-6.

38. Kiyotani K, Mushiroda T, Nakamura Y, et al. Pharmacogenomics of tamoxifen: roles of drug metabolizing enzymes and transporters. *Drug metabolism and pharmacokinetics* 2012;27(1):122-31.

39. Desta Z, Ward BA, Soukhova NV, et al. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *The Journal of pharmacology and experimental therapeutics* 2004;310(3):1062-75.

40. Gjerde J, Geisler J, Lundgren S, et al. Associations between tamoxifen, estrogens, and FSH serum levels during steady state tamoxifen treatment of postmenopausal women with breast cancer. *BMC cancer* 2010;10:313.

41. Lim JS, Chen XA, Singh O, et al. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP2C19 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *British journal of clinical pharmacology* 2011;71(5):737-50.

42. Murdter TE, Schroth W, Bacchus-Gerybadze L, et al. Activity levels of tamoxifen metabolites at the estrogen receptor and the impact of genetic polymorphisms of phase I and II enzymes on their concentration levels in plasma. *Clinical pharmacology and therapeutics* 2011;89(5):708-17.

43. Sim SC. CYP2C19 allele nomenclature. (http://www.cypalleles.ki.se/cyp2c19.htm). (Accessed January 6 2015).

44. Scott SA, Sangkuhl K, Shuldiner AR, et al. PharmGKB summary: very important pharmacogene information for cytochrome P450, family 2, subfamily C, polypeptide 19. *Pharmacogenetics and genomics* 2012;22(2):159-65.

45. de Morais SM, Wilkinson GR, Blaisdell J, et al. The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. *The Journal of biological chemistry* 1994;269(22):15419-22.

46. De Morais SM, Wilkinson GR, Blaisdell J, et al. Identification of a new genetic defect responsible for the polymorphism of (S)-mephenytoin metabolism in Japanese. *Molecular pharmacology* 1994;46(4):594-8.

47. Scott SA, Sangkuhl K, Gardner EE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. *Clinical pharmacology and therapeutics* 2011;90(2):328-32.

48. Desta Z, Zhao X, Shin JG, et al. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. *Clinical pharmacokinetics* 2002;41(12):913-58.

49. Ruiter R, Bijl MJ, van Schaik RH, et al. CYP2C19*2 polymorphism is associated with increased survival in breast cancer patients using tamoxifen. *Pharmacogenomics* 2010;11(10):1367-75.

50. Sim SC, Risinger C, Dahl ML, et al. A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clinical pharmacology and therapeutics* 2006;79(1):103-13.

51. Wang G, Lei HP, Li Z, et al. The CYP2C19 ultra-rapid metabolizer genotype influences the pharmacokinetics of voriconazole in healthy male volunteers. *European journal of clinical pharmacology* 2009;65(3):281-5.

52. Sugimoto K, Uno T, Yamazaki H, et al. Limited frequency of the CYP2C19*17 allele and its minor role in a Japanese population. *British journal of clinical pharmacology* 2008;65(3):437-9.

53. Sibbing D, Gebhard D, Koch W, et al. Isolated and interactive impact of common CYP2C19 genetic variants on the antiplatelet effect of chronic clopidogrel therapy. *Journal of thrombosis and haemostasis : JTH* 2010;8(8):1685-93.

54. Serrano D, Lazzeroni M, Zambon CF, et al. Efficacy of tamoxifen based on cytochrome P450 2D6, CYP2C19 and SULT1A1 genotype in the Italian Tamoxifen Prevention Trial. *The pharmacogenomics journal* 2011;11(2):100-7.

55. Binkhorst L, Mathijssen RH, Jager A, et al. Individualization of tamoxifen therapy: Much more than just CYP2D6 genotyping. *Cancer treatment reviews* 2015.

56. Beelen K, Opdam M, Severson TM, et al. CYP2C19 2 predicts substantial tamoxifen benefit in postmenopausal breast cancer patients randomized between adjuvant tamoxifen and no systemic treatment. *Breast cancer research and treatment* 2013;139(3):649-55.

57. van Schaik RH, Kok M, Sweep FC, et al. The CYP2C19*2 genotype predicts tamoxifen treatment outcome in advanced breast cancer patients. *Pharmacogenomics* 2011;12(8):1137-46.

58. Chamnanphon M, Pechatanan K, Sirachainan E, et al. Association of CYP2D6 and CYP2C19 polymorphisms and disease-free survival of Thai post-menopausal breast cancer patients who received adjuvant tamoxifen. *Pharmacogenomics and personalized medicine* 2013;6:37-48.

59. Dezentje VO, van Schaik RH, Vletter-Bogaartz JM, et al. CYP2D6 genotype in relation to tamoxifen efficacy in a Dutch cohort of the tamoxifen exemestane adjuvant multinational (TEAM) trial. *Breast cancer research and treatment* 2013;140(2):363-73.

60. Mwinyi J, Vokinger K, Jetter A, et al. Impact of variable CYP genotypes on breast cancer relapse in patients undergoing adjuvant tamoxifen therapy. *Cancer chemotherapy and pharmacology* 2014;73(6):1181-8.

61. Okishiro M, Taguchi T, Jin Kim S, et al. Genetic polymorphisms of CYP2D6 10 and CYP2C19 2, 3 are not associated with prognosis, endometrial thickness, or bone mineral density in Japanese breast cancer patients treated with adjuvant tamoxifen. *Cancer* 2009;115(5):952-61.

62. Moyer AM, Suman VJ, Weinshilboum RM, et al. SULT1A1, CYP2C19 and disease-free survival in early breast cancer patients receiving tamoxifen. *Pharmacogenomics* 2011;12(11):1535-43.

63. Justenhoven C, Hamann U, Pierl CB, et al. CYP2C19*17 is associated with decreased breast cancer risk. *Breast cancer research and treatment* 2009;115(2):391-6.

64. Whirl-Carrillo M, McDonagh EM, Hebert JM, et al. Pharmacogenomics knowledge for personalized medicine. *Clinical pharmacology and therapeutics* 2012;92(4):414-7.

65. de Leon J. AmpliChip CYP450 test: personalized medicine has arrived in psychiatry. *Expert review of molecular diagnostics* 2006;6(3):277-86.

66. Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. 1997. (http://www.whitehouse.gov/omb/fedreg_1997standards/). (Accessed January 8 2015).

67. American Cancer Society. How is breast cancer staged? (http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-staging). (Accessed January 8 2015).

68. White IR, Royston P. Imputing missing covariate values for the Cox model. *Statistics in medicine* 2009;28(15):1982-98.

69. Johnson JA, Hamadeh IS, Langaee TY. Loss of heterozygosity at the CYP2D6 locus in breast cancer: implications for tamoxifen pharmacogenetic studies. *Journal of the National Cancer Institute* 2015;107(2).

70. Ratain MJ, Nakamura Y, Cox NJ. CYP2D6 genotype and tamoxifen activity: understanding interstudy variability in methodological quality. *Clinical pharmacology and therapeutics* 2013;94(2):185-7.

71. Province MA, Goetz MP, Brauch H, et al. CYP2D6 genotype and adjuvant tamoxifen: meta-analysis of heterogeneous study populations. *Clinical pharmacology and therapeutics* 2014;95(2):216-27.

72. Berry DA. Response. *Journal of the National Cancer Institute* 2014;106(2):djt380.

Tables

Table 1a. Inferred phenotype levels of combined CYP2D6 phenotypes[1] and CYP2C19*17 genotypes

Level	CYP2D6 phenotype	CYP2C19*17 allele presence
1	EM/EM	Yes
2	EM/EM	No
	EM/IM	Yes
	EM/PM	Yes
3	EM/IM	No
	EM/PM	No
	IM/PM	Yes or No
	IM/IM	Yes or No
	PM/PM	Yes or No

[1]CYP2D6 UM considered as EM for creation of these levels

Table 1b. Inferred phenotype levels of combined CYP2D6 phenotypes[1] and CYP2C19*2 genotypes

Level	CYP2D6 phenotype	CYP2C19*2 allele presence
1	EM/EM	No
2	EM/EM	Yes
	EM/IM	No
	EM/PM	No
3	EM/IM	Yes
	EM/PM	Yes
	IM/PM	Yes or No
	IM/IM	Yes or No
	PM/PM	Yes or No

[1]CYP2D6 UM considered as EM for creation of these levels

Table 1. Creation of a CYP2D6 phenotype/CYP2C19 genotype variable. Inferred metabolizer levels (1 = most extensive metabolizer, 3 = least extensive metabolizer) were assigned using a combination of CYP2D6 phenotype and CYP2C19*17 (1a) or CYP2C19*2 (1b) genotypes. UM: ultrarapid metabolizer, EM: extensive metabolizer, IM: intermediate metabolizer, PM: poor metabolizer.

Table 2. Selected ITPC patient characteristics by study site[1]

	Site 2	Site 4	Site 6	Site 8	Site 9	Site 11	Site12
Total women	186 (8.9)	217 (10)	191 (9.1)	875 (42)	73 (3.5)	305 (15)	255 (12)
Recurrences	45 (15)	3 (1.0)	68 (23)	119 (40)	2 (0.68)	43 (15)	16 (5.4)
Median DFS time[2] [range]	45 [7.2-173]	52 [1.2-81]	124 [3.1-207]	64 [2.1-244]	28 [7.1-132]	65 [0.33-140]	68 [4.2-121]
Age at diagnosis	54 (11)	51 (10)	61 (10)	65 (9.8)	48 (9.9)	62 (14)	45 (8.0)
Premenopausal	78 (30)	64 (25)	10 (3.8)	43 (17)	0 (0.0)	66 (25)	0 (0.0)
Postmenopausal	96 (6.4)	121 (8.0)	136 (9.0)	825 (55)	73 (4.8)	210 (14)	47 (3.1)
Variant Genotypes							
*1/*17	73 (16)	78 (18)	0 (0.0)	294 (66)	0 (0.0)	0 (0.0)	0 (0.0)
*17/*17	6 (7.6)	6 (7.6)	0 (0.0)	67 (85)	0 (0.0)	0 (0.0)	0 (0.0)
*1/*2	57 (10)	61 (11)	48 (8.5)	223 (39)	17 (3.0)	75 (13)	86 (15)
*2/*2	2 (2.8)	4 (5.6)	3 (4.2)	23 (32)	5 (7.0)	4 (5.6)	30 (42)
Missing *2 data	3 (6.4)	1 (2.1)	0 (0.0)	43 (92)	0 (0.0)	0 (0.0)	0 (0.0)
Missing *17 data	0 (0.0)	0 (0.0)	191 (23)	25 (2.9)	73 (8.6)	305 (36)	255 (30)
Source of *17 tumor DNA							
Blood	186 (17)	208 (19)	0 (0.0)	470 (42)	0 (0.0)	0 (0.0)	255 (23)
Fresh frozen plasma[3]	0 (0.0)	0 (0.0)	0 (0.0)	381 (100)	0 (0.0)	0 (0.0)	0 (0.0)
Fresh frozen tumor tissue	0 (0.0)	0 (0.0)	0 (0.0)	24 (100)	0 (0.0)	0 (0.0)	0 (0.0)
Source of *2 tumor DNA							
Blood	186 (17)	217 (19)	0 (0.0)	470 (42)	0 (0.0)	0 (0.0)	255 (23)
Fresh frozen plasma[3]	0 (0.0)	0 (0.0)	0 (0.0)	381 (100)	0 (0.0)	0 (0.0)	0 (0.0)
Fresh frozen tumor tissue	0 (0.0)	0 (0.0)	191 (37)	24 (4.6)	0 (0.0)	305 (59)	0 (0.0)
HWE[4] χ^2 (p-value)							
CYP2C19*17	1.3 (0.26)	0.86 (0.35)	N/A	0.67 (0.41)	N/A	N/A	N/A
CYP2C19*2	1.5 (0.23)	0.27 (0.60)	0.12 (0.73)	0.01 (0.91)	5.2 (0.02)	0.21 (0.65)	3.1 (0.08)

[1]Numbers given as n (% total study pop with corresponding characteristic) or mean (sd), unless otherwise noted

[2]In months

[3]Of tumor tissue

[4]Hardy-Weinberg Equilibrium

Table 3. Demographic and disease characteristics of included women from the ITPC dataset (n (%) unless stated otherwise)

	All women (n = 2102)	No recurrence (n= 1806)	Recurrence (n = 296)
Mean age at diagnosis in years (sd)	59 (13)	59 (13)	61 (12)
Race			
White	1506 (72)	1271 (70)	235 (79)
Asian or Pacific Islander	276 (13)	258 (14)	18 (6.1)
Other	12 (0.57)	12 (0.67)	0 (0.0)
Missing or unknown	308 (15)	265 (15)	43 (15)
Menopause status at diagnosis			
Postmenopausal	1565 (75)	1333 (74)	232 (78)
Premenopausal	261 (12)	227 (13)	34 (12)
Missing or unknown	276 (13)	246 (14)	30 (10)
Intended tamoxifen duration			
2 years	33 (1.6)	30 (1.7)	3 (1.0)
5 years	2069 (98)	1776 (98)	293 (99)
Tumor Nottingham grade			
1	263 (13)	248 (14)	15 (5.1)
2	1078 (52)	936 (52)	142 (48)
3	354 (17)	256 (14)	98 (33)
Unknown	407 (19)	366 (20)	41 (14)
Tumor stage			
1	810 (39)	753 (42)	57 (19)
2a	605 (29)	523 (29)	82 (28)
2b	224 (11)	177 (9.8)	47 (16)
3a	146 (7.0)	106 (5.9)	40 (14)
3c	55 (2.6)	34 (1.9)	21 (7.1)
Unknown	262 (13)	213 (12)	49 (17)
Progesterone receptor status			
Positive	1754 (83)	1514 (84)	240 (81)
Negative	287 (14)	233 (13)	54 (18)
Missing or unknown	61 (2.9)	59 (3.3)	2 (0.68)
Additional cancer therapy use			
Radiation	1274 (61)	1085 (60)	189 (64)
Chemotherapy	429 (20)	369 (20)	60 (20)
Aromatase Inhibitor	119 (5.7)	112 (6.2)	7 (2.4)
CYP2D6 phenotype			
Extensive metabolizer	814 (39)	708 (39)	106 (36)
Intermediate metabolizer	1139 (54)	968 (54)	171 (58)
Poor metabolizer	116 (5.5)	100 (5.5)	16 (5.4)
Missing or unknown	33 (1.6)	30 (1.7)	3 (1.0)
CYP2C19*2 genotype[1]			
Two null function alleles	71 (3.4)	62 (3.4)	9 (3.0)
One null function allele	567 (27)	486 (27)	81 (27)
No null function allele	1417 (67)	1215 (67)	202 (68)
Missing or unknown	47 (2.2)	43 (2.4)	4 (1.4)
CYP2C19*17 genotype[1]			
Two gain of function alleles	79 (3.8)	70 (3.9)	9 (3.0)
One gain of function allele	445 (21)	387 (21)	58 (20)
No gain of function allele	729 (35)	632 (35)	97 (33)
Missing or unknown	849 (40)	717 (40)	132 (45)

[1] A *1 allele is defined as a functional allele.

Table 4. Cox proportional hazard ratios for CYP2C19*2 and CYP2C19*17 genotypes

Comparison	Hazard Ratio	95% Confidence Interval
CYP2C19*2		
Genotypes		
No *2 allele	1.0 (Reference)	---
*2/*1	1.05	0.78, 1.42
*2/*2	0.79	0.32, 1.94
CYP2C19*17		
Genotypes		
No *17 allele	1.0 (Reference)	---
*17/*1	1.02	0.71, 1.46
*17/*17	0.57	0.26, 1.24

Table 5. Cox proportional hazard ratios for CYP2C19*2 and CYP2C19*17 genotypes, stratified by menopausal status and CYP2D6 phenotype

Stratum and Comparison	Hazard Ratio	95% Confidence Interval	Stratum and Comparison	Hazard Ratio	95% Confidence Interval
Premenopausal Women			**Premenopausal Women**		
No *2 allele	1.0 (Reference)	---	No *17 allele	1.0 (Reference)	---
*2/*1	2.34	0.97, 5.64	*17/*1	1.32	0.52, 3.32
*2/*2[1]	---	---	*17*17[1]	---	---
Postmenopausal Women			**Postmenopausal Women**		
No *2 allele	1.0 (Reference)	---	No *17 allele	1.0 (Reference)	---
*2/*1	0.98	0.69, 1.39	*17/*1	0.98	0.66, 1.45
*2/*2	0.83	0.31, 2.26	*17/*17	0.57	0.25, 1.31
CYP2D6 EMs			**CYP2D6 EMs**		
No *2 allele	1.0 (Reference)	---	No *17 allele	1.0 (Reference)	---
*2/*1	1.58	0.97, 2.56	*17/*1	1.01	0.56, 1.82
*2/*2	0.47	0.065, 3.46	*17/*17	0.16	0.022, 1.16
CYP2D6 IMs			**CYP2D6 IMs**		
No *2 allele	1.0 (Reference)	---	No *17 allele	1.0 (Reference)	---
*2/*1	0.75	0.48, 1.17	*17/*1	1.15	0.70, 1.91
*2/*2	0.52	0.13, 2.16	*17/*17	0.71	0.17, 2.98
CYP2D6 PMs			**CYP2D6 PMs**		
No *2 allele	1.0 (Reference)	---	No *17 allele	1.0 (Reference)	---
*2/*1	0.92	0.41, 2.06	*17/*1	0.83	0.29, 2.43
*2/*2	1.65	0.21, 13	*17/*17	1.75	0.47, 6.55

[1]Not enough homozygotes were present in the data to provide an estimate. EMs: extensive metabolizers; IMs: intermediate metabolizers; PMs: poor metabolizers

Table 6. Cox proportional hazard ratios for combinations of CYP2D6 metabolizer phenotypes with CYP2C19*2 or CYP2C19*17 genotypes

Comparison	Hazard Ratio	95% Confidence Interval
CYP2C19*2		
Level 3[1]	1.0 (Reference)	---
Level 2	1.19	0.86, 1.66
Level 1	0.78	0.52, 1.15
CYP2C19*17		
Level 3	1.0 (Reference)	---
Level 2	0.96	0.73, 1.25
Level 1	0.70	0.43, 1.14

[1]Levels defined in Table 1.

Table S1. Cox proportional hazard ratios for CYP2C19*2 in imputed datasets

Genotypes	Hazard Ratio	95% Confidence Interval	Imputation
No *2 allele	1.0 (Reference)	---	**1**
*2/*1	1.03	0.80, 1.33	
*2/*2	0.99	0.51, 1.95	
No *2 allele	1.0 (Reference)	---	**2**
*2/*1	1.01	0.78, 1.31	
*2/*2	1.00	0.53, 1.89	
No *2 allele	1.0 (Reference)	---	**3**
*2/*1	1.02	0.79, 1.32	
*2/*2	1.05	0.57, 1.93	
No *2 allele	1.0 (Reference)	---	**4**
*2/*1	1.00	0.78, 1.30	
*2/*2	1.00	0.49, 2.05	
No *2 allele	1.0 (Reference)	---	**5**
*2/*1	1.00	0.77, 1.23	
*2/*2	0.99	0.51, 1.94	

Table S2. Cox proportional hazard ratios for CYP2C19*17 in imputed datasets

Genotypes	Hazard Ratio	95% Confidence Interval	Imputation
No *17 allele	1.0 (Reference)	---	**1**
*17/*1	1.17	0.92, 1.49	
*17/*17	1.24	0.78, 1.98	
No *17 allele	1.0 (Reference)	---	**2**
*17/*1	1.04	0.82, 1.33	
*17/*17	1.07	0.64, 1.79	
No *17 allele	1.0 (Reference)	---	**3**
*17/*1	1.00	0.78, 1.27	
*17/*17	0.79	0.46, 1.37	
No *17 allele	1.0 (Reference)	---	**4**
*17/*1	1.04	0.82, 1.32	
*17/*17	0.73	0.41, 1.31	
No *17 allele	1.0 (Reference)	---	**5**
*17/*1	1.05	0.83, 1.33	
*17/*17	0.59	0.33, 1.05	

Figures

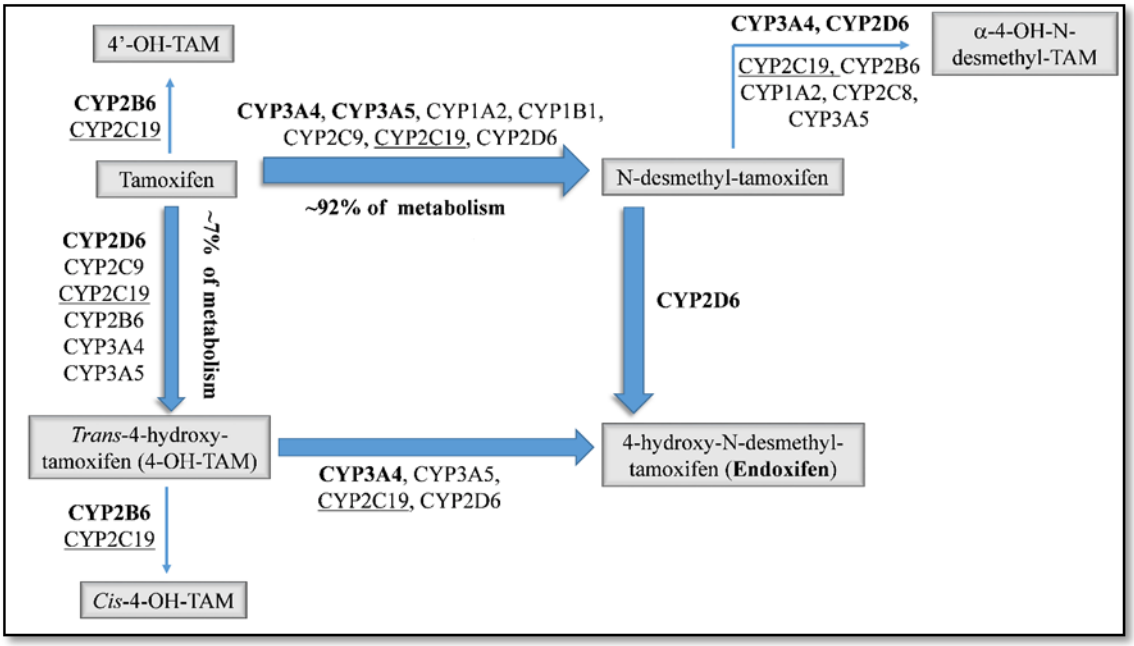

Figure 1. The role of CYP2C19 in the hepatic metabolism of tamoxifen. Enzymes that catalyze each reaction are listed beside the corresponding arrow, with those thought to be the primary catalysts of each pathway in bold. The CYP2C19 enzyme (underscored) is involved in the metabolism of tamoxifen to both of the primary metabolites N- desmethyl tamoxifen and 4-OH-TAM, as well as the demethylation of 4-OH-TAM to endoxifen (33-37). CYP2C19 is also involved in the metabolism of tamoxifen to other metabolites not considered as clinically relevant for breast cancer treatment (11). A more detailed pathway can be found in Klein et al. (2013) (11).

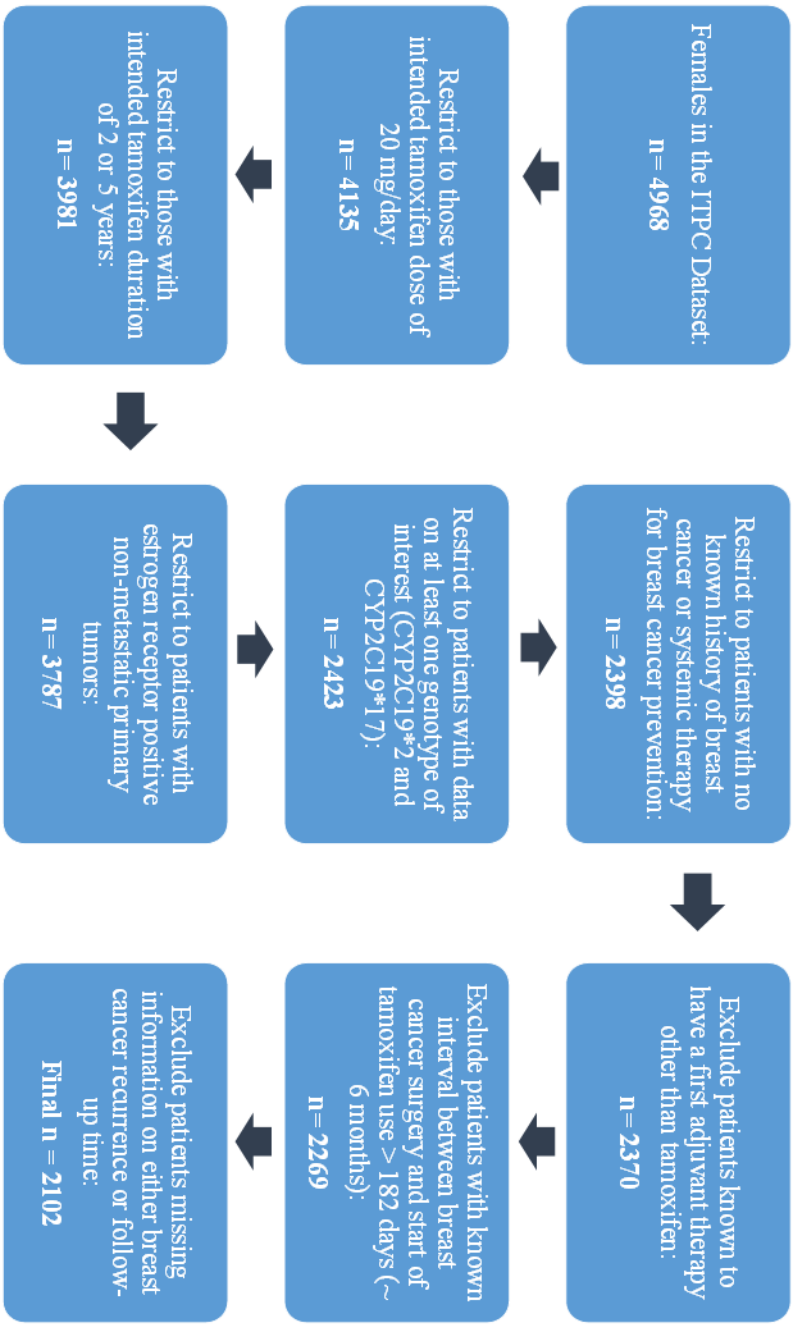

Figure 2. **Study inclusion flow diagram.** Of the 4,968 women included in the ITPC dataset, 2,102 met the inclusion criteria.

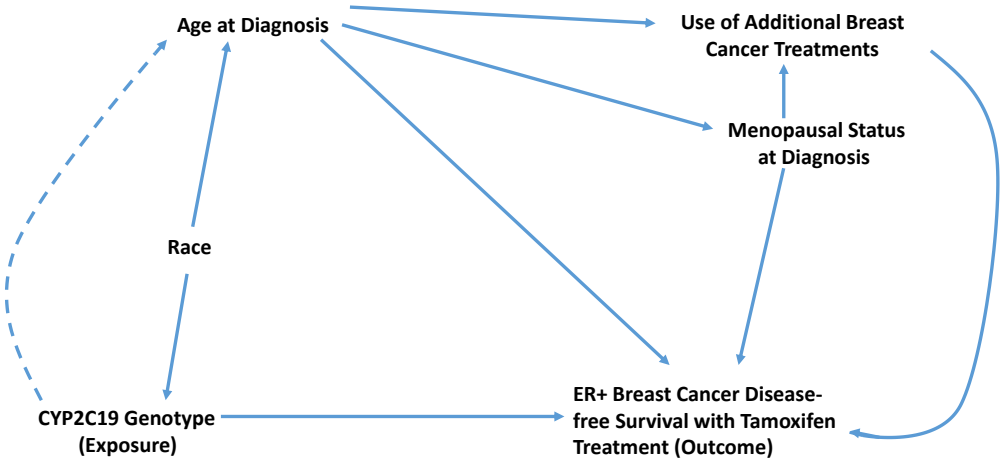

Figure 3. Directed acyclic graph. Hypothesized relationships among variables that may influence the observed association of interest between CYP2C19 genotype and disease-free survival among estrogen receptor positive breast cancer patients treated with tamoxifen. The arrow joining CYP2C19 genotype and age at diagnosis is dashed because it is inferred from evidence that CYP2C19 variants may experience altered baseline risks of breast cancer due to lifelong exposure to altered endogenous estrogen levels (63). ER: estrogen receptor.

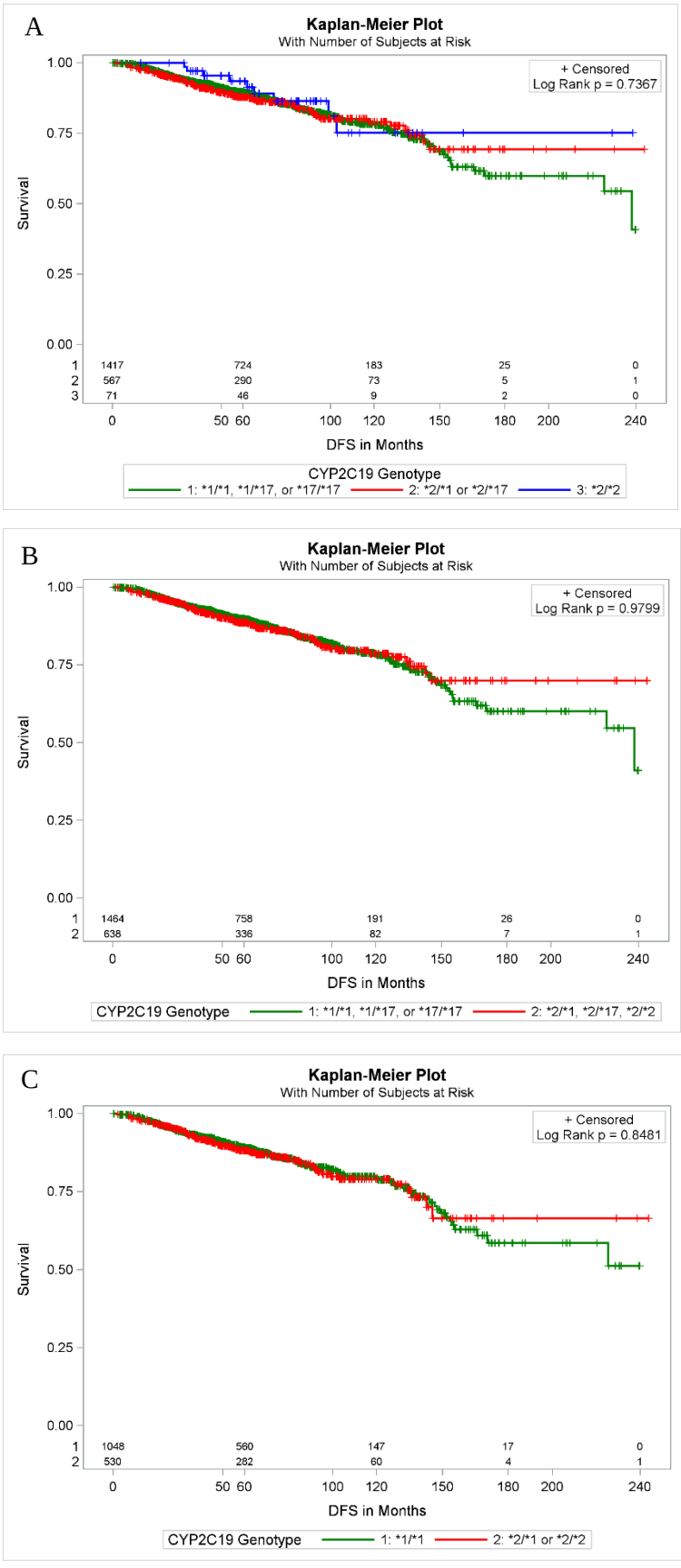

Fig 4. Kaplan-Meier analysis of disease-free survival (DFS) with tamoxifen use for CYP2C19*2 genotypes. The number of subjects at risk in each stratum at each interval of 60 months (5 years) is shown: (A) Carriers of no *2 allele vs. carriers of one *2 allele vs. carriers of two *2 alleles; (B) Carriers of no *2 allele vs. carriers of one or two *2 alleles; (C) Individuals with a CYP2C19 *1/*1 genotype vs. individuals with a *2/*1 or *2/*2 genotype

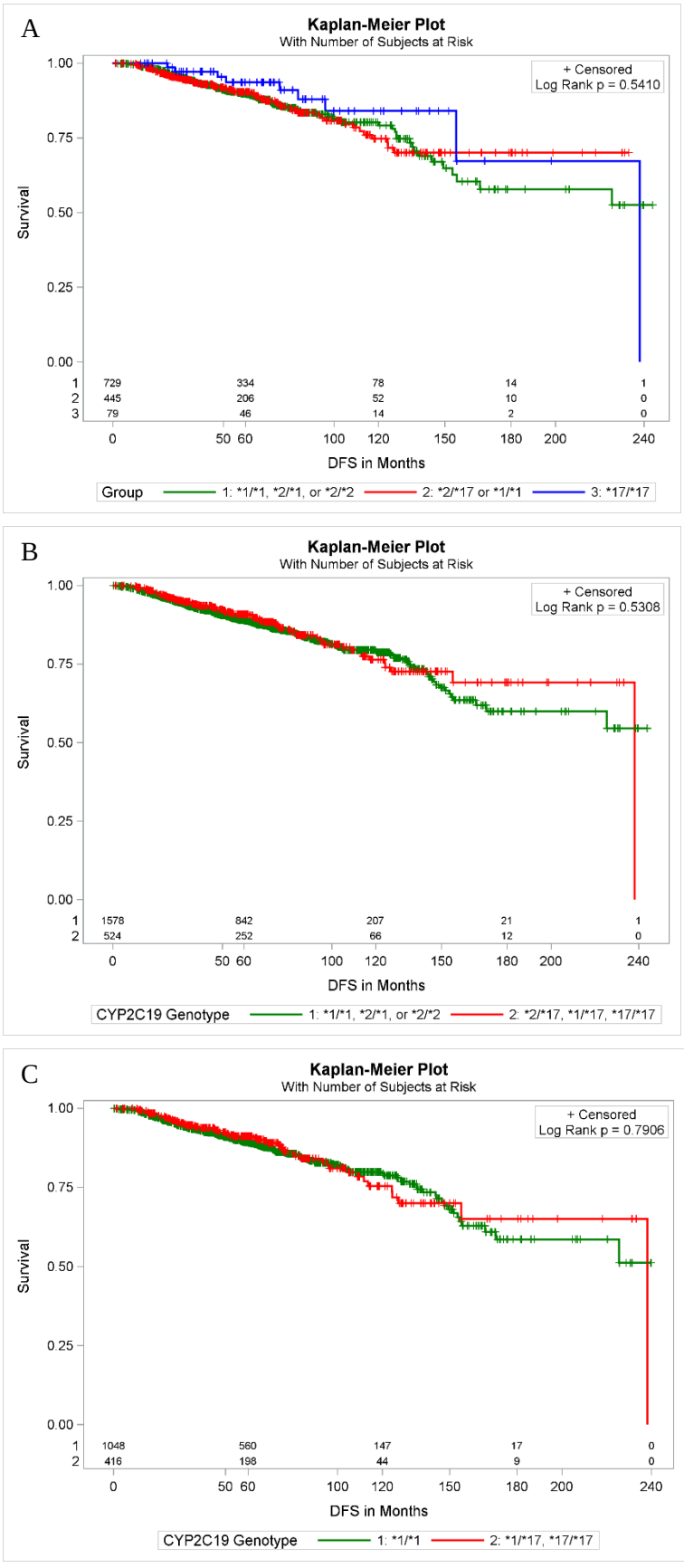

Fig 5. Kaplan-Meier analysis of disease-free survival (DFS) with tamoxifen use for CYP2C19*17 genotypes. The number of subjects at risk in each stratum at each interval of 60 months (5 years) is shown: (A) Carriers of no *17 allele vs. carriers of one *17 allele vs. carriers of two *17 alleles; (B) Carriers of no *17 allele vs. carriers of one or two *17 alleles; (C) Individuals with a CYP2C19 *1/*1 genotype vs. individuals with a *17/*1 or *17/*17 genotype

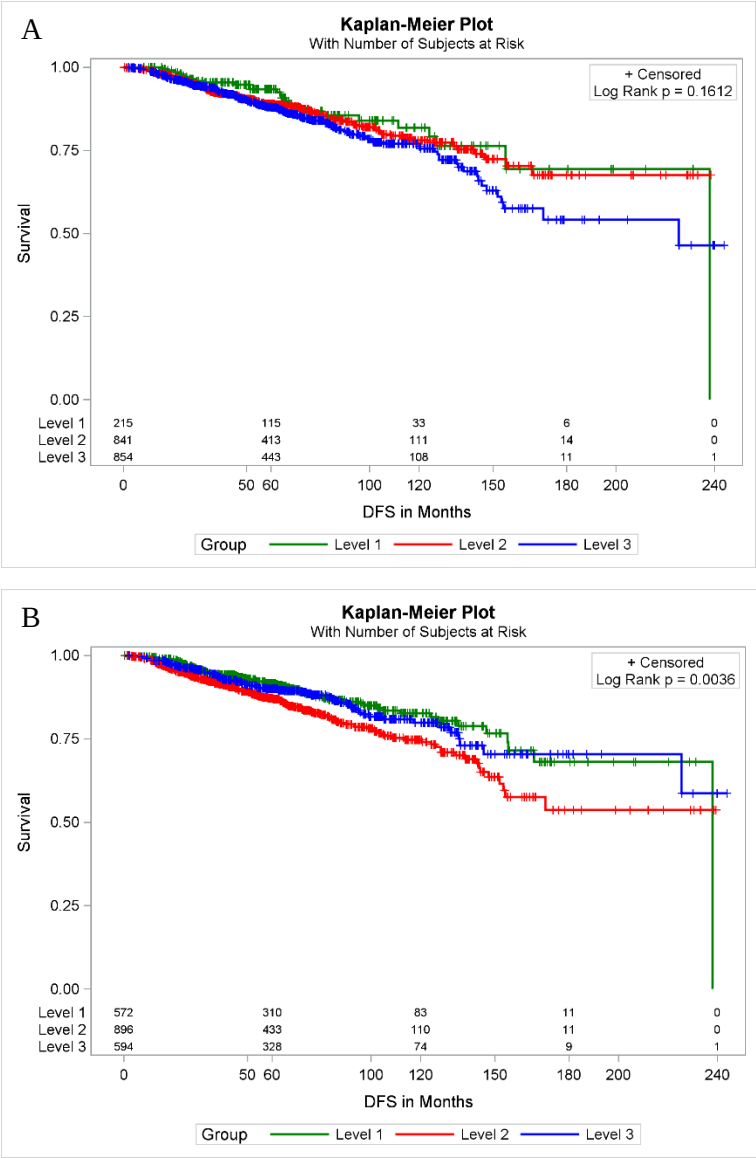

Fig 6. Kaplan-Meier analysis of disease-free survival (DFS) with tamoxifen use for CYP2D6 phenotypes in combination with CYP2C19 genotypes. Levels 1, 2, and 3 are defined in Table 1. The number of subjects at risk in each stratum at each interval of 60 months (5 years) is shown: (A) CYP2D6 phenotype and CYP2C19*17 genotype combinations; (B) CYP2D6 phenotype and CYP2C19*2 genotype combinations.