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:

_____ _____

Sadaf A. Bhai Date

The Effect of CYP2C19 Polymorphisms on Breast Cancer Recurrence among a Danish Cohort of Premenopausal Tamoxifen Treated Women

By

Sadaf A. Bhai
Master of Public Health

Department of Epidemiology

Timothy L. Lash, DSc, MPH
Committee Chair

The Effect of CYP2C19 Polymorphisms on Breast Cancer Recurrence among a Danish Cohort of Premenopausal Tamoxifen Treated Women

By

Sadaf A. Bhai

B.S.
University of Illinois at Urbana- Champaign
2015

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
2020

Abstract

The Effect of CYP2C19 Polymorphisms on Breast Cancer Recurrence among a Danish Cohort of Premenopausal Tamoxifen Treated Women

By Sadaf A. Bhai

Tamoxifen is the guideline therapy for premenopausal women who are diagnosed with estrogen receptor positive breast cancer. Cytochrome P450 enzymes are responsible for breaking tamoxifen down to its active metabolites. The CYP2C19 enzyme is on the path for producing 4-OH-TAM and its conversion to endoxifen, which are two key metabolites that compete with estrogen to bind to the estrogen receptor. However, polymorphisms in the *CYP2C19* gene can affect the amount of metabolite produced. This study used 5,959 Danish premenopausal women enrolled in the Predictors of Breast Cancer Recurrence cohort to determine if there was an association between *CYP2C19* genotype and breast cancer recurrence after tamoxifen treatment. Additionally, this study aimed to find if CYP2C19 inhibiting drug therapy was an effect modifier for this association. The genotypes examined were *CYP2C19*2*, which causes loss of enzyme function, and *CYP2C19*17*, which allows for gain of enzyme function. Cox proportional hazard ratios were calculated along with the corresponding 95% confidence intervals for univariate and adjusted analysis in both ER+/ TAM+ and ER-/ TAM- cohorts. The interaction assessment found that hazard ratios between women who took CYP2C19 inhibiting drugs and those who did not were not meaningfully different between genotypes in both cohorts. In the ER+/ TAM+ cohort, adjusted analysis gave hazard ratios of 1.17 (0.93, 1.49) and 1.01 (0.55, 1.86) for hetero- and homozygotes with *CYP2C19*2* genotypes and hazard ratios of 1.09 (0.86, 1.39) and 0.77 (0.32, 1.86) for hetero- and homozygotes with *CYP2C19*17* genotypes. In the ER-/ TAM- cohort, adjusted analysis gave hazard ratios of 1.38 (0.91, 2.11) and 1.29 (0.59, 2.24) for hetero- and homozygotes with *CYP2C19*2* genotypes and hazard ratios of 1.00 (0.60, 1.66) and 0.84 (0.20, 3.50) for hetero- and homozygotes with *CYP2C19*17* genotypes. The results of this analysis indicate that there is no association between *CYP2C19*2* and *CYP2C19*17* genotypes and breast cancer recurrence, and that CYP2C19 inhibiting drugs do not modify this association. This study utilized a large cohort of premenopausal women, which most previous studies did not do. Future studies should aim to focus on the effect different dosages of tamoxifen might have on recurrence.

The Effect of CYP2C19 Polymorphisms on Breast Cancer Recurrence among a Danish Cohort of Premenopausal Tamoxifen Treated Women

By

Sadaf A. Bhai

B.S.
University of Illinois at Urbana- Champaign
2015

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
2020

Acknowledgements

I would like to thank Dr. Timothy Lash, the faculty thesis advisor for this work, for his guidance throughout this study. I would also like to thank Lindsay Collin for her support in completing this thesis. Finally, I would like to thank my family and friends who supported me throughout this process.

Table of Contents

Introduction:

Breast cancer is the most common cancer in women outside of certain skin cancers (1). It is estimated that there were 268,600 new cases and 41,760 deaths associated with female breast cancer in the United States in 2019 (2). Hormone receptor positive (HR+) breast cancer is the most common worldwide, with approximately 60 to 75% of women having estrogen receptor positive (ER+) cancers (3).

Tamoxifen, a selective estrogen receptor modulator (SERM), is the standard drug therapy given to premenopausal women with ER+ breast cancer (4). This SERM acts as an anti-estrogen among breast cells, while continuing to act as an agonist in other body tissues such as the uterus (5). Tamoxifen and its metabolites prevent estrogen binding to the estrogen receptor (ER) through competitive inhibition, suppressing the growth of the tumor (6). A five-year course of treatment with tamoxifen reduces breast cancer recurrence by nearly 50%, and more recent findings show treatment up to ten years might have additional benefit against recurrence and mortality (4,7,8).

Cytochrome P450 enzymes, especially from the *CYP2* and *CYP3* gene families, are responsible for metabolizing most prescribed drugs and endogenous steroids in the body (9). CYP2C19 plays a role in the formation of tamoxifen's primary metabolites; *trans*-4-hydroxy-tamoxifen (4-OH-TAM), which has 30 to 100 times more affinity to the ER than tamoxifen, and N-desmethyl-tamoxifen (10,11). CYP2C19 also is responsible for converting 4-OH-TAM into 4-hydroxy-N-desmethyl-tamoxifen (endoxifen), a metabolite that binds to the ER with a similar affinity as 4-OH-TAM (11). The concentration of endoxifen in human plasma samples of breast cancer patients on tamoxifen therapy was found to be six times higher than the concentration of 4-OH-TAM, implying that endoxifen is a key metabolite in preventing recurrence (12).

Polymorphisms in the *CYP2C19* gene can cause differences in amount of metabolite produced, which may have therapeutic implications (13). This metabolic inhibition could also partially explain why almost 30% of patients receiving tamoxifen therapy according to guidelines still experience recurrence (14,15).

The most common variants that cause loss of enzyme function are *CYP2C19*2* and *CYP2C19*3*. The *CYP2C19*2* variant involves a guanine (G) to adenine (A) SNP at nucleotide 681 (681 G > A, rs4244285) in exon 5, which results in turning a proline into a cryptic aberrant splice site. The *CYP2C19*3* variant involves a G to A SNP at nucleotide 636 (636 G > A, rs4986893) in exon 3, which results in a premature TGA stop codon. In both variants, a truncated and catalytically inactive enzyme is produced. *CYP2C19*17*, which involves a cystine (C) to thymine (T) SNP in the promoter region of exon 5 (-806 C > T, rs12248560 or -3402 C > T, rs11188072), has been associated with higher transcription rates, possibly due to changing the interaction between transcription factors (16). The *CYP2C19*2* variant is seen at a frequency of 16% in African, 13.3% in Caucasian, and 28.4% in Asian populations, while the *CYP2C19*17* variant is seen at frequencies of 26.3%, 19.1%, and 1.5% respectively (17). The CYP2C19 enzyme is one of many that are responsible for metabolizing tamoxifen, however, producing an active or inactive variant of this enzyme has potential to affect the amount of metabolite produced.

While tamoxifen is the guideline treatment among premenopausal women with ER+ breast cancer, there have been few studies done on recurrence with this group (18). This is a concern that needs to be addressed because premenopausal women have higher levels of estrogens, specifically estradiol (E2), that compete for the ER against tamoxifen metabolites, which could change the effectiveness of tamoxifen (15). However, the

results of previous studies vary across this topic from polymorphisms showing impact on tamoxifen metabolism to having a null effect (19-21). The current analysis is aiming to view if the *CYP2C19*2* or *CYP2C19*17* variants are associated with breast cancer recurrence among a cohort of Danish premenopausal women treated with adjuvant tamoxifen therapy, and if CYP2C19 inhibiting drugs modify this association.

Methods:

Data Source and Population:

The study population includes women in The Predictors of Breast Cancer Recurrence (ProBe CaRe) cohort (18). This population-based cohort was established using patients from the Danish Breast Cancer Group (DBCG) registry, which was established in 1976 by the Danish Surgical Society and began to register patients nationally in Denmark in 1977. Over 90% of new breast cancer patients in Denmark are registered with the DBCG. New patients are registered to the group using standardized forms. The registry also has an established standardized method to collect patient information, tumor data, and treatment choice. All data for a patient is linked to their assigned Civil Personal Register (CPR) number, and this number is used in all national registries. The ProBe CaRe cohort is comprised of premenopausal women diagnosed with stage I-III primary breast cancer between 2002 and 2010 that were reported to the DBCG.

There were 8,047 premenopausal women who received a breast cancer diagnosis during the time period of interest. Of these, 5,959 patients were eligible to be included in the study cohort. Eligibility criteria for the ProBe CaRe cohort included having a primary stage I-III breast cancer diagnosis and being untreated with neoadjuvant therapy. These patients were then divided into two sub cohorts: patients who were ER+ and received tamoxifen (n=4,600) and those who were ER- and did not receive tamoxifen (n=1,359). Patients who were missing ER expression status, tamoxifen treatment status, or did not meet the sub cohort criteria were excluded (n=2,088) (18). Women missing *CYP2C19* genotype information were dropped from the survival analysis.

Analytic Variables:

The event of interest in this study is breast cancer recurrence. Among the ProBe CaRe cohort, recurrence was defined as any type of breast cancer diagnosed subsequent to the initial course of therapy, which is consistent with the DBCG definition. Follow up time is defined as days until the event of interest.

The genotypes that are the exposure of interest are *CYP2C19*2* (SNP 681 G > A, rs4244285) and *CYP2C19*17* (SNP -806 C > T, rs12248560). These are two of 32 variants genotyped in this cohort that are thought to potentially affect tamoxifen metabolism. Tissues samples were obtained by sending the CPR number and hospital of diagnosis of cohort members to a medical research technician at the Institute of Pathology. The technician was able to send out requests to the hospital to retrieve the formalin-fixed paraffin-embedded (FFPE) tissue samples of interest. The *CYP2C19*3* allele is not considered in this analysis due to information not being available for this variant in the cohort.

Covariables:

Potential covariates of interest include age at diagnosis, ER status, additional cancer therapies the patient received (mastectomy or breast conserving surgery, radiation therapy, or chemotherapy), statin therapy, cancer stage, cancer grade, and if the patient was taking a CYP2C19 inhibiting drug. CYP2C19 inhibiting drugs will be considered as a potential effect modifier in this analysis. Information about all covariates was obtained from the DBCG registry data. The cohort of interest are women who are ER+/TAM+, however, analysis will also be done on ER-/TAM- women to determine if the genotypes of interest are predictive or prognostic markers of recurrence. Data for CYP2C19

inhibiting therapy and statin therapy was obtained by linking the CPR number for the patients to the Danish National Prescription Registry.

Statistical Analysis:

Descriptive statistical analysis was conducted for the analytic variables and covariates of interest for the full ProBe CaRe cohort, along with for women who did and did not experience breast cancer recurrence. Descriptive analysis was also conducted for both the ER+/TAM+ and ER-/ TAM- cohorts. A directed acyclic graph (DAG) was used to identify potential confounding between the variables (Supplement 1). The DAG indicated there was no confounding because there are no ancestors to the *CYP2C19* genotype, however, CYP2C19 inhibiting therapy will be studied as a potential effect modifier.

Univariate and adjusted survival analysis was conducted to determine the hazard ratio between genotype expression and recurrence. The proportional hazard (PH) assumptions for each variable were tested using graphical methods. For purposes of testing the PH assumptions, age was dichotomized to women under 45, and women 45 and older. Univariate and adjusted Cox proportional hazard ratios were computed, along with their corresponding 95% confidence intervals for homozygous and heterozygous genotypes for each variant, along with for CYP2C19 inhibiting drug therapy. For the sake of completeness, the final Cox proportional hazard model included the variables age, additional cancer therapies received, statin therapy, cancer stage, and if CYP2C19 inhibiting drugs were used. Cancer grade and HER2 status were excluded from the final model due to a significant portion of the population having missing data. Age at diagnosis was kept as a continuous variable in the adjusted models. Due to limited

numbers of some variants, homozygous and heterozygous variants were collapsed into one category, and the hazard ratio and 95% confidence intervals were obtained. Cell counts of less than 5 on tables were suppressed in compliance with the Danish Data Protection Board rules. This study was approved by the Emory Institutional Review Board, the Danish Data Protection Board, and the ethical committee for Aarhus University. All analysis was performed using SAS version 9.4.

Results:

Study Population

Of the 5,959 patients eligible in this cohort for analysis, 612 experienced breast cancer recurrence. Of this, 396 were in the ER+/ TAM+ cohort, while 216 were in the ER-/ TAM- cohort. Table 1 shows characteristics of the full ProBe CaRe cohort, the ER+/ TAM+ cohort, and the ER-/ TAM- cohort. The characteristics explored include breast cancer recurrence, follow up time, age at diagnosis, cancer stage, cancer grade, HER2 status, statin therapy, *CYP2C19*2* or *CYP2C19*17* genotype, CYP2C19 inhibiting therapy, and other cancer therapies including mastectomy and breast conserving surgery, chemotherapy, and radiation therapy. The ER+/ TAM+ cohort had 806 and 67 women hetero- and homozygous for *CYP2C19*2*, and 979 and 115 women hetero- and homozygous for *CYP2C19*17*. The ER-/ TAM- cohort had 148 and 14 women hetero- and homozygous for *CYP2C19*2*, and 242 and 32 women hetero- and homozygous for *CYP2C19*17*. Additionally, there are 2,096 women on CYP2C19 inhibiting drug therapy, of which 1,707 are in the ER+/ TAM+ cohort and 389 are in the ER-/ TAM- cohort.

Effect Modification from CYP2C19 Inhibiting Therapy

Tables 2a and 2b show the Cox proportional hazard ratios for collapsed genotype variants in both cohorts. The hazard ratios among women with *CYP2C19*2* genotypes in the ER+/ TAM+ who took CYP2C19 inhibiting drugs was 1.31 (95% CI: 0.89, 1.94) and the hazard ratio among those who did not was 1.07 (95% CI: 0.81, 1.43). For women with *CYP2C19*17* genotypes in the ER+/ TAM+ cohort, women who took CYP2C19 inhibiting drugs had a hazard ratio of 0.79 (95% CI: 0.51, 1.24), while those who did not

had a hazard ratio of 1.22 (95% CI: 0.92, 1.61). Among the ER-/ TAM- cohort, women with *CYP2C19*2* genotypes who were on CYP2C19 inhibiting therapy had a hazard ratio of 1.35 (95% CI: 0.55, 3.31), while those not on inhibiting therapy had a hazard ratio of 1.39 (95% CI: 0.89, 2.18). Women with *CYP2C19*17* genotypes who were on CYP2C19 inhibiting therapy had a hazard ratio of 0.20 (95% CI: 0.03, 1.46), while those not on inhibiting therapy had a hazard ratio of 1.22 (95% CI: 0.74, 2.02). There were not enough women to obtain hazard ratios for all genotypes tested, however genotype specific hazard ratios are in Supplement 2a and 2b.

Survival Analysis

All variables in the adjusted model met the PH assumption from the graphical test. Since the interaction assessment determined CYP2C19 inhibiting therapy is not an effect modifier, the adjusted analysis omitted CYP2C19 inhibiting therapy from this category. However, a univariate and adjusted analysis was also conducted on the association between CYP2C19 inhibiting drugs therapy and breast cancer recurrence.

Tables 3a and 3b show the univariate and adjusted analysis for women with *CYP2C19*2* and *CYP2C19*17* genotypes, and women who were on CYP2C19 inhibiting drug therapy in the ER+/ TAM+ cohort. The analysis was adjusted for age at diagnosis, cancer stage, statin therapy and other adjuvant cancer therapies. For *CYP2C19*2*, the adjusted hazard ratio for heterozygotes is 1.17 (95% CI: 0.93, 1.49) and for homozygotes is 1.01 (95% CI:0.55, 1.86). For *CYP2C19*17*, the adjusted hazard ratio for heterozygotes is 1.09 (95% CI: 0.86, 1.39) and for homozygotes is 0.77 (0.32, 1.86).

Tables 4a and 4b show the univariate and adjusted analysis for women with *CYP2C19*2* and *CYP2C19*17* genotypes, and women who were on CYP2C19 inhibiting

drug therapy in the ER-/ TAM- cohort. The analysis was adjusted for age at diagnosis, cancer stage, statin therapy and other adjuvant cancer therapies. For *CYP2C19*2*, the adjusted hazard ratio for heterozygotes is 1.38 (95% CI: 0.91, 2.11) and the adjusted hazard ratio for homozygotes is 1.29 (95% CI: 0.59, 2.24). For *CYP2C19*17*, the adjusted hazard ratio for heterozygotes is 1.00 (95% CI: 0.60, 1.66) and the adjusted hazard ratio for homozygotes is 0.84 (95% CI: 0.20, 3.50).

Discussion

The results from this analysis shows that there is likely no association between the *CYP2C19*2* and *CYP2C19*17* genotypes and breast cancer recurrence among a cohort of Danish women treated with adjuvant tamoxifen therapy. There is also likely no effect modification by CYP2C19 inhibiting drug therapy between a *CYP2C19* genotype and breast cancer recurrence. Therefore, *CYP2C19* genotype should not be considered when deciding on adjuvant tamoxifen therapy for a patient.

For both ER+/ TAM+ and ER-/ TAM- cohorts, the interaction analysis shows that the hazard ratios between women with *CYP2C19*2* genotypes who took CYP2C19 inhibiting drugs were not meaningfully different from women who were not taking the CYP2C19 inhibiting drugs, indicating that this therapy is likely not an effect modifier for these genotypes. For the *CYP2C19*17* genotypes, the hazard ratio among women who were taking CYP2C19 inhibiting drugs in both cohorts indicated that time to breast cancer recurrence was actually slower as compared to women who had wild type alleles. However, the confidence intervals were not precise due to the genotype being rare. Additionally, for the ER-/ TAM- cohort, the heterozygous *CYP2C19*17* population was censored, as shown in Supplement 2, due to a cell count of less than 5, which indicates that a small sample size potentially influenced these results. For the *CYP2C19*17* genotypes that were not on CYP2C19 inhibiting therapy, the result of the interaction assessment was similar to the *CYP2C19*2* genotypes, indicating no effect modification by this therapy.

In both cohorts, the hazard ratio for *CYP2C19*2* variants and the associated confidence intervals are very similar between the univariate and adjusted analysis. This

implies that the factors adjusted for (age at diagnosis, cancer stage, other adjuvant therapies, and statin therapy) do not affect the relationship between *CYP2C19* genotype and breast cancer recurrence. Between the ER+/ TAM+ and ER-/TAM- cohorts, the hazard ratio and corresponding confidence intervals are also very similar between each genotype. This indicates that *CYP2C19* genotype variant is not a predictive marker for decreased breast cancer recurrence with tamoxifen therapy.

Univariate and adjusted analysis in both cohorts of the effect of CYP2C19 inhibiting therapy shows that there is a protective association between women who are on the therapy versus those that are not. However, this study did not look into factors considered before a patient is prescribed one of these drugs. These results, which were not main hypothesized effects, were also likely affected by immortal person-time bias. Women prescribed the medicines had to survive free of recurrence until the prescription was filled, whereas there was no such pressure on women who received no prescription. This likely accounts for all or part of the reported protective associations.

The findings of this study are consistent with other studies that indicate there is no meaningful association between *CYP2C19* genotype and breast cancer recurrence when a patient is treated with tamoxifen. For instance, in a similar study, Damkier et. al. found hazard ratios of 1.05 (95% CI: 0.78, 1.42) and 0.79 (95% CI: 0.32, 1.94) for hetero- and homozygote individuals with *CYP2C19*2* genotypes, and hazard ratios of 1.02 (95% CI: 0.71, 1.46) and 0.57 (95% CI: 0.26, 1.24) for the respective *CYP2C19*17* genotypes (20). This study was done using the International Tamoxifen Pharmacogenomics Consortium dataset. However, Damkier et. al. was limited to only women who had been treated with tamoxifen in their analysis, while the current study was able to compare results to a sub-

cohort of women who did not take tamoxifen. The earlier study also included primarily postmenopausal women, whereas the present study was restricted to premenopausal women, for whom tamoxifen remains guideline endocrine therapy.

A strength of this study was using the ER-/ TAM- cohort to examine if *CYP2C19* genotype was a predictive or prognostic marker of recurrence, as previous studies did not have a comparison group within the cohort not treated with tamoxifen. This study also utilized a large cohort of only premenopausal women, which was not done in most previous studies. Additionally, since the data used in analysis was registry based, most patients had complete and accurate information on record that was collected in a standardized manner.

One limitation of this analysis is that the study cohort is a single race, which might cause concerns when applying these findings broadly. Another limitation is that free estrogen levels and tamoxifen metabolite levels in blood serum were not measured in the cohort. Depending on the variant, women with different genotypes could have different amounts of free estrogen and tamoxifen metabolites, which could impact if she experiences a recurrence of breast cancer. Future studies may want to consider the effect that tamoxifen dosages have on breast cancer recurrence by measuring free estrogen and tamoxifen metabolites in a multi-racial population. Using a quantitative measure in populations where *CYP2C19* polymorphisms are common could shed light on if certain variants would require dosages of tamoxifen that are higher than the current guideline doses.

References

1. Breast Cancer Statistics | CDC. 2019;(https://www.cdc.gov/cancer/breast/statistics/index.htm). (Accessed January 15, 2020)

2. Cancer of the Breast (Female) - Cancer Stat Facts. *SEER*. (https://seer.cancer.gov/statfacts/html/breast.html). (Accessed January 15, 2020)

3. 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. *J Clin Oncol*. 2014;32(21):2255–2269.

4. Dean L. Tamoxifen Therapy and CYP2D6 Genotype. In: Pratt V, McLeod H, Rubinstein W, et al., eds. *Medical Genetics Summaries*. Bethesda (MD): National Center for Biotechnology Information (US); 2012 (Accessed January 20, 2020)(http://www.ncbi.nlm.nih.gov/books/NBK247013/). (Accessed January 20, 2020)

5. Martinkovich S, Shah D, Planey SL, et al. Selective estrogen receptor modulators: tissue specificity and clinical utility. *Clin Interv Aging*. 2014;9:1437–1452.

6. Tamoxifen in the Treatment of Breast Cancer | NEJM. (https://www-nejm-org.proxy.library.emory.edu/doi/full/10.1056/NEJM199811263392207). (Accessed January 20, 2020)

7. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet*. 2011;378(9793):771–784.

8. Davies C, Pan H, Godwin J, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*. 2013;381(9869):805–816.

9. Nebert DW, Russell DW. Clinical importance of the cytochromes P450. *The Lancet*. 2002;360(9340):1155–1162.

10. 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):122–131.

11. Lim YC, Desta Z, Flockhart DA, et al. Endoxifen (4-hydroxy-N-desmethyl-tamoxifen) has anti-estrogenic effects in breast cancer cells with potency similar to 4-hydroxy-tamoxifen. *Cancer Chemother Pharmacol*. 2005;55(5):471–478.

12. Stearns V, Johnson MD, Rae JM, et al. Active Tamoxifen Metabolite Plasma Concentrations After Coadministration of Tamoxifen and the Selective Serotonin Reuptake Inhibitor Paroxetine. *J Natl Cancer Inst*. 2003;95(23):1758–1764.

13. Thota K, Prasad K, Basaveswara Rao MV. Detection of Cytochrome P450 Polymorphisms in Breast Cancer Patients May Impact on Tamoxifen Therapy. *Asian Pac. J. Cancer Prev.* 2018;19(2):343–350.

14. 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. *JCO*. 2011;29(13):1657–1663.

15. Johnson MD, Zuo H, Lee K-H, et al. Pharmacological Characterization of 4-hydroxy-N-desmethyl Tamoxifen, a Novel Active Metabolite of Tamoxifen. *Breast Cancer Res Treat*. 2004;85(2):151–159.

16. Brown S-A, Pereira N. Pharmacogenomic Impact of CYP2C19 Variation on Clopidogrel Therapy in Precision Cardiovascular Medicine. *Journal of Personalized Medicine*. 2018;8(1):8.

17. Gurusamy U, Shewade DG. Chapter 46 - Pharmacogenomics in India. In: Padmanabhan S, ed. Handbook of Pharmacogenomics and Stratified Medicine. San Diego: Academic Press; 2014 (Accessed April 17, 2020):1037–1059.(http://www.sciencedirect.com/science/article/pii/B978012386882400046). (Accessed April 17, 2020)

18. Collin LJ, Cronin-Fenton DP, Ahern TP, et al. Cohort Profile: the Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark. *BMJ Open* [electronic article]. 2018;8(7). (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6074634/). (Accessed February 3, 2020)

19. 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 Res Treat*. 2013;139(3):649–655.

20. Damkier P, Kjærsgaard A, Barker KA, et al. CYP2C19*2 and CYP2C19*17 variants and effect of tamoxifen on breast cancer recurrence: Analysis of the International Tamoxifen Pharmacogenomics Consortium dataset. *Sci Rep* [electronic article]. 2017;7.

(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5552748/). (Accessed January 22, 2020)

21. Schroth W, Antoniadou L, Fritz P, et al. Breast Cancer Treatment Outcome With Adjuvant Tamoxifen Relative to Patient CYP2D6 and CYP2C19 Genotypes. JCO. 2007;25(33):5187–5193.

Table 1: Population Characteristics of the ProBe CaRe cohort

Characteristic	Full Cohort (n=5,959)	ER+/ TAM+ (n= 4,600)	ER-/ TAM- (n=1, 359)
Breast Cancer Recurrence (%)			
Yes	612 (10)	396 (8.6)	216 (16)
No	5,347 (90)	4,204 (91)	1,143 (84)
Follow Up Time (Years)* (SD)	7.5 (2.6)	7.6 (2.5)	7.2 (2.8)
Age at diagnosis** (SD)	44.7 (6.0)	45.2 (5.6)	43.0 (6.8)
Cancer Grade			
I	976 (16)	955 (21)	21 (1.5)
II	2,607 (44)	2,391 (52)	216 (16)
III	1,801 (30)	950 (21)	884 (65)
Unsuitable or Unknown	542 (9.1)	304 (6.6)	238 (18)
Cancer Stage (%)			
I	1,586 (27)	1,184 (26)	402 (30)
II	3,178 (53)	2,476 (54)	702 (52)
III	1,163 (20)	917 (20)	246 (18)
Missing or Unknown	32 (0.5)	23 (0.5)	9 (0.7)
HER2 Status			
HER2+ (%)	973 (16)	619 (14)	354 (26)
HER2 – (%)	3,579 (60)	2,887 (63)	692 (51)
Missing or Unknown (%)	1,407 (24)	1,094 (24)	313 (23)
Surgery Type			
Mastectomy (%)	2,660 (45)	2,033 (44)	627 (46)
Breast Conserving (%)	3,299 (55)	2,567 (56)	732 (54)
Chemotherapy (%)			
Yes	5,413 (91)	4,163 (91)	1,250 (92)
No	546 (9.2)	437 (9.5)	109 (8.0)
Radiation Therapy (%)			
Yes	5,037 (85)	3,945 (86)	1,092 (80)
No	922 (15)	655 (14)	267 (20)
Statin Therapy (%)			
Yes	364 (6.1)	270 (5.9)	94 (6.9)
No	5,588 (94)	4,326 (94)	1,262 (93)
*CYP2C19*2* Genotype			
WT	3,487 (59)	2,728 (75)	759 (81)
1	954 (16)	806 (22)	148 (16)
2	81 (1.4)	67 (1.9)	14 (1.5)
Missing	43 (0.7)	29 (0.8)	14 (1.5)
*CYP2C19*17* Genotype			
WT	2,453 (41)	1,963 (54)	490 (52)
1	1,221 (20)	979 (27)	242 (26)
2	147 (2.5)	115 (3.2)	32 (3.4)
Missing	744 (12)	573 (16)	171 (18)
CYP2C19 Inhibiting Therapy (%)			
Yes	2,096 (35)	1,707 (37)	389 (29)
No	3,856 (65)	2,889 (63)	967 (71)

*= continuous variable

Tables 2a and 2b: Effect of CYP2C19 inhibiting therapy on collapsed *CYP2C19* genotypes in ER+/ TAM+ and ER-/ TAM- cohorts

CYP2C19 Inhibiting Therapy in ER+/TAM+ Cohort	Hazard Ratio	95% Confidence Interval
***CYP2C19*2*- Inhibitor**		
No *2 allele	1.0 (reference)	-
Any *2 allele	1.31	0.89, 1.94
***CYP2C19*2*- No Inhibitor**		
No *2 allele	1.0 (reference)	-
Any *2 allele	1.07	0.81, 1.43
***CYP2C19*17*- Inhibitor**		
No *17 allele	1.0 (reference)	-
Any *17 allele	0.79	0.51, 1.24
***CYP2C19*17*- No Inhibitor**		
No *17 allele	1.0 (reference)	-
Any *17 allele	1.22	0.92, 1.61

CYP2C19 Inhibiting Therapy in ER-/ TAM- cohort	Hazard Ratio	95% Confidence Interval
***CYP2C19*2*- Inhibitor**		
No *2 allele	1.0 (reference)	-
Any *2 allele	1.35	0.55, 3.31
***CYP2C19*2*- No Inhibitor**		
No *2 allele	1.0 (reference)	-
Any *2 allele	1.39	0.89, 2.18
***CYP2C19*17*- Inhibitor**		
No *17 allele	1.0 (reference)	-
Any *17 allele	0.20	0.03, 1.46
***CYP2C19*17*- No Inhibitor**		
No *17 allele	1.0 (reference)	-
Any *17 allele	1.22	0.74, 2.02

Tables 3a and 3b: Univariate and Adjusted Analysis for ER+/ TAM+ cohort

Univariate Analysis	Hazard Ratio	95% Confidence Interval
CYP2C19*2		
No *2 allele	1.0 (reference)	-
*2/ WT	1.16	0.92, 1.47
*2/ *2	0.95	0.52, 1.74
Any *2 allele	1.14	0.90, 1.43
CYP2C19*17		
No *17 allele	1.0 (reference)	-
*17/ WT	1.20	0.86, 1.40
*17/ *17	0.74	0.30, 1.78
Any *17 allele	1.07	0.84, 1.35
CYP2C19 Drug Inhibiting Therapy		
No therapy	1.0 (reference)	-
Therapy	0.75	0.60, 0.93

Adjusted Analysis*	Hazard Ratio	95% Confidence Interval
CYP2C19*2		
No *2 allele	1.0 (reference)	-
*2/ WT	1.17	0.93, 1.49
*2/ *2	1.01	0.55, 1.86
Any *2 allele	1.16	0.92, 1.46
CYP2C19*17		
No *17 allele	1.0 (reference)	-
*17/ WT	1.09	0.86, 1.39
*17/ *17	0.77	0.32, 1.86
Any *17 allele	1.07	0.84, 1.35
CYP2C19 Drug Inhibiting Therapy		
No therapy	1.0 (reference)	-
Therapy	0.73	0.57, 0.92

***** Analysis adjusted for age at diagnosis, cancer stage, other adjuvant treatments, statin therapy, and CYP2C19 inhibiting drugs.

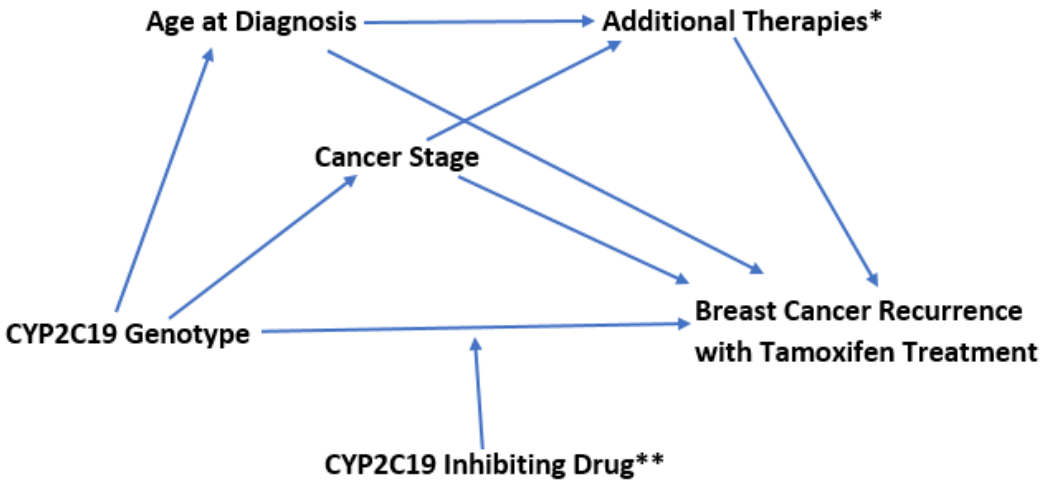
Tables 4a and 4b: Univariate and Adjusted Analysis for ER-/ TAM- cohort. Table 4b is adjusted for age at diagnosis, cancer stage, other adjuvant treatments, statin therapy, and CYP2C19 inhibiting drugs.

Univariate Analysis	Hazard Ratio	95% Confidence Interval
CYP2C19*2		
No *2 allele	1.0 (reference)	-
*2/ WT	1.34	0.88, 2.04
*2/ *2	1.56	0.67, 3.62
Any *2 allele	1.37	0.92, 2.04
CYP2C19*17		
No *17 allele	1.0 (reference)	-
*17/ WT	0.91	0.55, 1.51
*17/ *17	1.07	0.26, 4.35
Any *17 allele	0.93	0.57, 1.50
CYP2C19 Drug Inhibiting Therapy		
No therapy	1.0 (reference)	-
Therapy	0.47	0.30, 0.75

Adjusted Analysis*	Hazard Ratio	95% Confidence Interval
CYP2C19*2		
No *2 allele	1.0 (reference)	-
*2/ WT	1.38	0.91, 2.11
*2/ *2	1.29	0.59, 2.24
Any *2 allele	1.38	0.93, 2.07
CYP2C19*17		
No *17 allele	1.0 (reference)	-
*17/ WT	1.00	0.60, 1.66
*17/ *17	0.84	0.20, 3.50
Any *17 allele	0.98	0.61, 1.59
CYP2C19 Drug Inhibiting Therapy		
No therapy	1.0 (reference)	-
Therapy	0.55	0.34, 0.90

***** Analysis adjusted for age at diagnosis, cancer stage, other adjuvant treatments, statin therapy, and CYP2C19 inhibiting drugs.

Supplement 1: The directed acyclic graph (DAG) drawn to assess potential confounders in the relationship between CYP2C19 genotype and breast cancer recurrence.



* Additional therapies include surgery, chemotherapy, and radiation therapy as discussed in the study.
** CYP2C19 inhibiting drugs will be examined as an effect modifier between exposure and outcome.

Supplement 2a and 2b: Effect of CYP2C19 inhibiting therapy on specific *CYP2C19* genotypes in ER+/ TAM+ and ER-/ TAM- cohorts

CYP2C19 Inhibiting Therapy in ER+/TAM+ Cohort	Hazard Ratio	95% Confidence Interval
***CYP2C19*2*- Inhibitor**		
No *2 allele	1.0 (reference)	-
*2/ WT	9.38	0.08, 1,131.26
*2/ *2	1.26	0.55, 2.91
***CYP2C19*2*- No Inhibitor**		
No *2 allele	1.0 (reference)	-
*2/ WT	1.08	0.81, 1.46
*2/ *2	1.02	0.50, 2.08
***CYP2C19*17*- Inhibitor**		
No *17 allele	1.0 (reference)	-
*17/ WT	0.02	0.00, 3.15
*17/ *17	0.57	0.19, 1.73
***CYP2C19*17*- No Inhibitor**		
No *17 allele	1.0 (reference)	-
*17/ WT	1.24	0.93, 1.66
*17/ *17	0.88	0.33, 2.37

CYP2C19 Inhibiting Therapy in ER-/ TAM- cohort	Hazard Ratio	95% Confidence Interval
***CYP2C19*2*- Inhibitor**		
No *2 allele	1.0 (reference)	-
*2/ WT	0.52	0.00, 14,963.00
*2/ *2	1.17	0.31, 4.46
***CYP2C19*2*- No Inhibitor**		
No *2 allele	1.0 (reference)	-
*2/ WT	1.41	0.88, 2.26
*2/ *2	1.29	0.51, 3.28
***CYP2C19*17*- Inhibitor**		
No *17 allele	1.0 (reference)	-
17/ WT	-	-
*17/ *17	0.14	0.01, 1.76
***CYP2C19*17*- No Inhibitor**		
No *17 allele	1.0 (reference)	-
*17/ WT	1.27	0.75, 2.16
*17/ *17	0.88	0.21, 3.77

*: Not enough women to obtain a hazard ratio estimate in this genotype