BMJ Open

Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021805
Article Type:	Cohort profile
Date Submitted by the Author:	19-Jan-2018
Complete List of Authors:	Collin, Lindsay; Emory University School of Public Health, Epidemiology Cronin-Fenton, Deirdre; Aarhus Universitetshospital, Ahern, Thomas; The Robert Larner M.D. College of Medicine at The University of Vermont Christiansen, P; Aarhus Universitet Damkier, Per; University of Southern Denmark, Dept. Clinical Biochemistry and Pharmacology Ejlertsen, Bent; Copenhagen University Hospital, Oncology Hamilton Dutoit, Stephen; Aarhus Universitetshospital, Institute of Pathology Kjærsgaard, Anders; Aarhus Universitetshospital, Department of Clinical Epidemiology Silliman, Rebecca; Boston University School of Medicine; Aarhus Universitetshospital, Department of Clinical Epidemiology Sørensen, Henrik T.; Aarhus University Hospital, Department of Clinical Epidemiology Lash, Timothy; Emory University Rollins School of Public Health, Department of Epidemiology
Keywords:	Cohort Study, Breast tumours < ONCOLOGY, Pharmacogenetics, EPIDEMIOLOGY

Title: Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Author List: Lindsay J Collin[a], Deirdre P Cronin-Fenton[b], Thomas P Ahern[c], Peer M Christiansen[d,e,f], Per Damkier[g,h], Bent Ejlertsen[f,i], Stephen Hamilton-Dutoit[j], Anders Kjærsgaard[b], Rebecca A Silliman[b,k], Henrik Toft Sørensen[b,l], Timothy L Lash[a,b]

[a]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [b]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; [c]The Robert Larner, M.D. College of Medicine at The University of Vermont, Burlington, VT, USA; [d]Breast Unit, Surgical Department, Randers Regional Hospital, Randers, Denmark; [e]Department of Oncology, Odense University Hospital, Odense, Denmark; [f]Danish Breast Cancer Group, Copenhagen, Denmark; [g]Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark; [h]Department of Clinical Research, University of Southern Denmark, Denmark; [i]Rigshospitalet, Copenhagen, Denmark; [j]Institute of Pathology, Aarhus University Hospital, Aarhus, Denmark; [k]Boston University School of Medicine, Boston University, Boston, MA, USA; [l]Department of Health Research & Policy (Epidemiology), Stanford University, Stanford, CA, USA;

To whom correspondence should be addressed: Lindsay J Collin, Department of Epidemiology, 1518 Clifton Rd NE Atlanta GA 30322; Phone: 1-530-386-3341; Email: lindsay.jane.collin@emory.edu, ORCID: 0000-0003-0170-2592

Word Count: 3223

Keywords: Cohort study, breast cancer, pharmacogenetics, epidemiology

ABSTRACT

Purpose: The Predictors of Breast Cancer Recurrence (ProBe CaRe) study was established to evaluate modification of tamoxifen effectiveness in premenopausal women through reduced activity of tamoxifen-metabolizing enzymes. It comprehensively evaluates the effects of pharmacogenetic variants, use of concomitant medications, and biomarkers involved in oestrogen metabolism on breast cancer recurrence risk.

Participants: The ProBe CaRe study was established using resources from the Danish Breast Cancer Group (DBCG), including 5,959 premenopausal women diagnosed with stage I–III primary breast cancer between 2002 and 2010 in Denmark. Eligible participants were divided into two groups based on oestrogen receptor alpha (ERα) expression and receipt of tamoxifen therapy (TAM), 4,600 are classified as ERα+/TAM+ and 1,359 are classified as ERα–/TAM–. The ProBe CaRe study is a population-based cohort study nested in a nearly complete source population, clinical, tumour, and demographic data were abstracted from DBCG registry data. Linkage to Danish registries allows for abstraction of information regarding comorbid conditions, comedication use, and mortality. Formalin-fixed paraffin-embedded (FFPE) tissue samples have been prepared for DNA extraction and immunohistochemical assay.

Findings to date: To mitigate incorrect classification of patients into specific categories we conducted a validation substudy. We compared data acquired from registry and from medical record review to calculate positive and negative predictive values (PPV and NPV). We observed PPVs near 100% for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment. We found that the PPVs were

96% (95%CI 83, 100) for change in endocrine therapy, and 61% (95%CI 42,77) for menopausal transition.

Future plans: The ProBe CaRe cohort study is well-positioned to comprehensively examine pharmacogenetic variants. We will use a Bayesian pathway analysis to evaluate the complete tamoxifen metabolic path to allow for gene-gene interactions, incorporating information of other important patient characteristics.

Strengths and Limitations of this Study

- One potential limitation of the ProBe CaRe study is the homogeneity of the study sample, as almost all are of European descent.
- In addition to being the first large epidemiologic study to examine reduced activity of tamoxifen metabolism in premenopausal women, this study is strengthened by completeness of high quality data.
- Our study includes a validation substudy to mitigate errors from incorrect classification of patients into specific categories of key analytic variables.

INTRODUCTION

Endocrine therapy improves survival in patients with breast cancer regardless of axillary lymph node status. The Predictors of Breast Cancer Recurrence (ProBe CaRE) cohort study was established to evaluate modification of tamoxifen effectiveness in premenopausal women through reduced activity of tamoxifen-metabolizing enzymes. Candidates for adjuvant tamoxifen therapy include stage I–IV breast cancer patients with ER positive tumours, who constitute about two-thirds[1] of the approximately 1.7 million newly diagnosed breast cancer patients each year worldwide.[2] Current guidelines recommend that premenopausal patients with ER alpha positive (ERα+) cancers receive tamoxifen for five to ten years,[3-5] which reduces recurrence risk by nearly half,[5] and that tamoxifen may be offered to post-menopausal women with ERα+ cancers as an alternative to aromatase inhibitors. Tamoxifen metabolism is complex, but is principally catalysed by cytochrome P450 enzymes. Some metabolites bind with the ER with significantly greater affinity than tamoxifen itself, especially endoxifen, which has the highest ER-binding activity among tamoxifen metabolites. Activity of the enzymes involved in tamoxifen metabolism can vary between individuals due to inherited gene variants[6-10] or use of comedications.[7, 6] Although many studies have explored the association between these gene variants or use of comedications and failure of tamoxifen treatment,[11, 12] which manifests clinically as a recurrence, the interpretation of these studies remains controversial. Current clinical guidelines do not recommend genotyping these variant alleles to support treatment decisions,[13-15] but do recommend avoiding inhibiting comedications.[16]

To date, little available evidence on this topic is specific to premenopausal breast cancer patients. The competition between oestrogen and tamoxifen for ER binding is highly important for these patients because tamoxifen is a first-line guideline-recommended therapy[15, 13] for

premenopausal patients and because premenopausal women have higher concentrations of oestrogens to compete with tamoxifen for ER binding. Oestradiol, the most active oestrogen metabolite, binds with the ER with approximately the same affinity as endoxifen.[17] Premenopausal women have tenfold higher concentrations of oestradiol than postmenopausal women[18] and oestradiol concentrations tend to increase during tamoxifen therapy.[19, 18] This suggests that inhibition of tamoxifen-metabolizing enzymes is more likely to decrease effectiveness in premenopausal women, yet they have been seldom studied in this topic area.

Research questions

We established a premenopausal cohort of breast cancer patients to fill this important evidence gap, with the following primary study aims:

(1) Assess pharmacogenetics of tamoxifen metabolism and risk of breast cancer recurrence

We will assess the pharmacogenetics of tamoxifen metabolism by genotyping 32 variants in 15 enzymes (**Table 1**) thought to affect the concentration of the most active tamoxifen metabolites, and will evaluate the association between these variants and breast cancer recurrence in tamoxifen treated premenopausal breast cancer patients. Each of the selected enzymes is involved in at least one step in the tamoxifen metabolic pathway (**Figure 1**). Interactions with comedications that inhibit these metabolic enzymes also will be evaluated.

Table 1 Selected Functional variants and inhibitor comedications in genes whose enzymes metabolize tamoxifen

Gene	Number of selected functional variants	Inhibitor comedications
CYP2D6	5	bupropion, cinacalcet, fluoxetine, paroxetine, quinidine, duloxetine, sertraline, terbinafine, amiodarone, cimetidine, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, telithromycin, aprepitant, erythromycin, fluconazole, verapamil, diltiazem, cimetidine, voriconazole
CYP3A	1	
CYP3A5	1	
CYP2C9	2	fluconazole, amiodarone, voriconazole
CYP2C19	2	
CYP2B6	2	
CYP1A1	1	
SULT1A1	3	
SULT1E1	2	
UGT2B7	1	
UGT2B10	1	
ABCC2	3	
ABCG2	3	
ABCB1	4	
UGT2B15	1	

(2) Assess the interaction between the pharmacogenetics of tamoxifen metabolism and oestrogen receptor beta (ERβ) expression

We will assess the effect of interaction between the pharmacogenetics of tamoxifen metabolism and ERβ expression on risk of breast cancer recurrence. Previous studies have shown that co-expression of ERβ is associated with improved survival among patients with ERα+ tumours who are treated with tamoxifen.[20, 21] The ERβ receptor opposes ERα-mediated

proliferation.[20] Tumours that express both ERα and ERβ are less aggressive than tumours that express homodimers of ERα,[22, 20] due to the attenuated stimulation response from ERα/ERβ heterodimers. This suggests that metabolic inhibition may only affect ERβ– tumours. *In vitro* analyses have demonstrated that in ERα+/ERβ+ MCF7 cells, proliferation is inhibited by a wide range of endoxifen concentrations.[23] Still, in ERα+/ERβ– MCF7 cells, only physiologically high endoxifen concentrations inhibit proliferation,[23] indicating that metabolic inhibition affects risk of recurrence only when ERβ is absent.

(3) Assess interaction between inhibition of tamoxifen metabolism and oestrogen-regulating enzymes

Finally, we will assess the association between tumour expression of 17β-hydroxysteroid dehydrogenase 1 and 2 (17βHSD1 and 17βHSD2) and breast cancer recurrence. 17βHSD1 catalyses the conversion of oestrone (E1) to the most potent form of oestrogen, oestradiol (E2), and 17βHSD2 catalyses the reverse reaction.[24] E2 has the highest binding affinity for ER, and endoxifen acts through competitive inhibition at the receptor-binding site.[24] In breast tumour tissue, 17βHSD1 is more highly expressed than 17βHSD2. The opposite is usually observed in adjacent normal tissue.[25] Tumours with higher capacity to produce E2 endogenously through increased expression of the 17βHSD1 enzyme are more likely to overwhelm the tamoxifen metabolites in competition for ER binding, affecting tamoxifen effectiveness. These enzymes are ideal therapeutic targets to modulate E2 concentrations in tumour cells, and candidate inhibitors have been developed.[26, 24] We will evaluate whether disequilibrium of the 17βHSD1 and -2 enzymes (ratio >1) results in compromised tamoxifen effectiveness.

COHORT DESCRIPTION

The ProBe CaRe cohort was established using the resources of the Danish Breast Cancer Group (DBCG) registry. The DBCG registry was established in 1976 and began to register patients in 1977, with the goals of standardizing treatment, facilitating clinical trials, and monitoring outcomes among Danish breast cancer patients.[27] Since its inception, the DBCG has registered over 90% of women diagnosed with breast cancer in Denmark. Breast cancer patients are registered in the DBCG via standardized forms. The registry has a standard protocol to collect information on tumour, treatment, and patient characteristics. Using this information-rich resource, the ProBe CaRe cohort is nested in a nearly complete source population of premenopausal women diagnosed with stage I–III first primary breast cancer between 2002–2010 whose breast cancer was reported to the DBCG. In Denmark, all citizens and legal residents are assigned a Civil Personal Register (CPR) number, a unique 10-digit personal identifier assigned at birth or upon immigration that is used for identification across all national registries (**Electronic Supplementary Figure 1**).[28]

Of the 8047 premenopausal women diagnosed with breast cancer between 2002 and 2010 and recorded in the DBCG registry, 5959 cancers were identified as eligible based on being a stage I–III first primary breast cancer and untreated with neoadjuvant therapy; all others (n=2088) were excluded. The 5959 eligible patients then were divided into two cohorts based on ERα expression and receipt of tamoxifen therapy (**Figure 2**). To address competing explanations, (for instance if the biomarkers under study affect risk directly rather than mediating the tamoxifen effect), we will also evaluate the risk of recurrence in the subset of women with ERα– tumours who did not receive tamoxifen therapy (T–).

Our final ProBe CaRe study population consists of these 5,959 breast cancer patients divided into a cohort of ERα+/T+ (4600 patients) and a cohort of ERα–/T– (1359 patients). The socio-demographic and clinical characteristics of the two cohorts are described in **Table 2**. The distribution of the clinical and demographic characteristics between the two cohorts (ERα+/T+ vs. ERα–/T–) are relatively similar. They only differ meaningfully with respect to progesterone receptor (PR) status (58% vs. 1.4% PR+, respectively) and human epidermal growth factor receptor 2 (HER2) status (14% vs. 26% HER2+, respectively). With respect to outcomes, the ER–/TAM– cohort has a higher proportion of subjects who experienced recurrence (8.6% vs 16%, respectively) and who died by the end of follow up (7.8% vs. 18% respectively). This pattern is to be expected, as ER– breast cancers generally have a worse prognosis than ER+ cancers, especially within the first five years following diagnosis.[29, 30]

Table 2 Distribution of clinical and tumour characteristics by ER status and receipt of tamoxifen among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study.

Patient and tumour characteristics	**ER+/TAM+**		**ER–/TAM–**	
	N	%	N	%
Total	4600	100	1359	100
Age at diagnosis				
<35	222	4.8	182	23
35-49	487	11	229	27
40-44	1123	24	321	24
45-49	1668	36	385	28
50+	1100	24	242	18
Menopausal status at diagnosis				
Premenopausal	4600	100	1359	100
Stage at diagnosis				
Stage I	1184	26	402	29.6
Stage II	2476	54	702	51.7
Stage III	917	20	246	18.1

Unknown stage	23	0.5	9	0.7
Tumor size				
< 2mm	2646	58	677	50
2 - <5mm	1780	39	632	47
> 5mm	156	3.4	44	3.2
Unknown	18	0.4	6	0.4
Number of metastatic lymph nodes				
0	1704	37	695	51
1	1148	25	238	17
2	583	13	116	9
3+	1152	25	306	23
Unknown	13	0.3	4	0.3
Lymph node evaluation				
No	8	0.2	3	0.2
Yes	4592	100	1356	100
Histologic grade				
Unsuitable	10	0.2	13	1
I	955	21	21	1.5
II	2391	52	216	16
III	950	21	884	65
Unknown	294	6.4	225	17
Type of primary surgery				
Mastectomy	2033	44	627	46
Lumpectomy	2567	56	732	54
Progesterone receptor status				
PR-	383	8.3	1121	83
PR+	2680	58	19	1.4
Unknown/Not measured	1537	33	219	16
HER2 status				
HER2-	2887.00	63	692	51
HER2+	619	14	354	26
Unknown/Not measured	1094	24	313	23
Intention to treat with chemotherapy				
No	144	3	14	1
Yes	4456	97	1345	99
Chemotherapy				
No	437	9	109	8
Yes	4163	91	1250	92
Intention to treat with Tamoxifen				
No	70	1.5	1351	99
Yes	4530	98	8	0.6

Radiation therapy					
	No	655	14	267	20
	Yes	3945	86	1092	80
Anti-HER2 therapy					
	No	2887	63	692	51
	Yes	619	13	354	26
	Unknown	1094	24	313	23
Recurrence					
	No	4204	91	1143	84
	Yes	396	8.6	216	16
Another malignancy					
	No	4544	99	1341	99
	Yes	56	1.2	18	1.3
Dead at end-of-follow-up					
	No	4239	92	1115	82
	Yes	361	8	244	18
Charlson Comorbidity Score					
	0	4587	99	1344	99
	1	4	0.1	4	0.3
	2	2	0	3	0.2
	3+	7	0.2	8	0.6

Cohort follow-up

Women diagnosed with breast cancer and subsequently enrolled in the DBCG registry undergo semi-annual examinations during the first five years after diagnosis and annual examinations during years six to ten.[31] Women undergoing treatment for breast cancer receive endocrine therapy through the Danish government and obtain their medicine at the hospital. Members of both the ERα+/T+ and the ERα–/T– ProBe CaRe cohorts have been followed from breast cancer diagnosis to the first of (a) recurrence, (b) death, (c) ten years of follow-up, (d) loss to follow-up due to emigration, (e) another malignancy, or (f) the end of the study follow-up period. Breast cancer recurrence was identified using the DBCG registry. We adopted the DBCG definition of breast cancer recurrence as any type of breast cancer diagnosed subsequent to the initial course of therapy.[31] Recurrences are then further categorized as loco-regional (in the

scar or regional lymph nodes), contralateral (opposite breast), distant (all other sites), or unknown (site of recurrence not documented). The date of recurrence is recorded in the DBCG registry, including recurrences diagnosed between scheduled follow-up exams. Mortality and emigration were identified using the Danish Civil Registration System, which is updated daily.[28] Emigration is the only expected source of loss to follow-up and has impacted less than 1% of the study population.

Data Collection

Registry Data

Once participants eligible for ProBe CaRe were selected from the DBCG registry, we extracted clinical and demographic information from the DBCG registry. This information included date and place of diagnosis, tumour characteristics, treatment received, and patient characteristics, which are presented in **Electronic Supplementary Table 1**. We also extracted information on comorbid diseases at time of breast cancer diagnosis, summarized using the Charlson Comorbidity Index.[32, 33] The registry information allows us to update the participants' conditions during study follow-up and therefore to account for time-varying exposures and confounding factors. The CPR number for each patient was used to link cohort members to the Danish National Prescription Registry,[34] which provided information on filled prescriptions of drugs that inhibit tamoxifen-metabolizing enzymes. This allowed us to assess the drug-drug interactions discussed above.

Biobank

The CPR number for each patient was used to identify the hospital at which the surgery was performed and to locate and retrieve the corresponding formalin-fixed paraffin-embedded

(FFPE) tissue samples. The list of ProBe CaRe cohort members and their CPR numbers and hospitals of diagnosis were provided to a medical research technician at the Institute of Pathology, blinded to whether the CPR numbers corresponded to a patient with a recurrence. The technician reviewed a description of the available tissue blocks (routinely available in the Danish pathology registry[35]), identified the tumour-rich and non-neoplastic blocks for each patient, and specified which FFPE blocks should be requested from the hospitals. This list of blocks to be requested was then returned to the Department of Clinical Epidemiology, which prepared and mailed the request letters to the pathology archives at the respective Danish hospitals. Staff at the hospital pathology archives returned the blocks to the Department of Clinical Epidemiology, which assigned a project identification number to the block and then provided it to the Institute of Pathology. The project identification number maintained blinding of laboratory personnel to whether blocks corresponded to patients with a recurrence. The Department of Clinical Epidemiology maintains the key linking the project identification number for the blocks to the clinical data, including recurrence status.

Non-neoplastic tissue samples are taken routinely from normal adjacent tissue or cancer-free lymph nodes resected during breast cancer surgery and were used as controls in creation of tissue microarrays (TMAs). Of the 4,600 patients included in the ERα+/T+ cohort, 4,599 patients had samples evaluated by clinicians, and tumour samples were available for 3,959 (86%). Among the ER–/TAM– cohort, 1139 (84%) patients had tumour samples available. Distribution of clinical and demographic characteristics among patients with, and without, available tumour samples are described in **Table 3**. Of patients included in the ERα+/T+ cohort, 2746 (82%) had a non-neoplastic tissue sample available, while 1082 (80%) patients in the ER–/TAM– cohort had non-neoplastic tissue samples available. Distributions of demographic and clinical characteristics

among patients with and without non-neoplastic tissue samples are summarized in **Electronic Supplementary Table 2**.

Table 3 Summary of exposure, covariate, and outcome variables collected in the ProBe CaRe study

Exposures	**Outcome**
Genetic variants	Recurrence
ERα and ERβ	Mortality
17βHSD1 and 17βHSD2	
Biomarkers	
Clinical	**Demographic**
Tumour characteristics	Age
Treatment therapy	Region
Comorbidity	Hospital of Diagnosis
Medication History and Use	

Sections of collected tissue blocks have been prepared for DNA extraction and immunohistochemical (IHC) assay. In accordance with the study's primary aims, we will genotype 32 variants across 15 genes known to be involved in tamoxifen metabolism, in order to predict extent of metabolic inhibition. We will also re-assay ERα expression to ensure correct classification of the two cohorts, as original ERα expression was measured in different pathology laboratories using different methods. In our previous case-control study of post-menopausal patients diagnosed during 1985–2001, we reported 95% concordance of positive ERα expression between initial assays and reassays and 74% concordance of negative ERα expression between initial assays and reassays.[36] ERβ expression will be assayed using IHC in TMAs to assess its possible modification of tamoxifen metabolic inhibition. Expression of the enzymes 17βHSD1 and 17βHSD2 also will be assayed using IHC to address the study aim examining whether the

ratio of these two enzymes modulates tamoxifen's efficacy in preventing breast cancer recurrence. The aforementioned assays of biomarkers are the primary starting point. However, we anticipate that the study will yield a substantial tumour biobank and ultimately provide a valuable resource to researchers for further characterization of prognostic and predictive biomarkers in premenopausal breast cancer.

Validation substudy

Registry data are not error-free.[37] To mitigate incorrect classification of patients into specific categories, we conducted a validation substudy.[38] By comparing data procured both from the registry and from medical record review, we calculated positive and negative predictive values (PPV and NPV, respectively) and their corresponding confidence intervals for key analytic variables. We observed near perfect PPVs for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy, and tamoxifen treatment. The PPVs were 96% (95% CI 83, 100) for change in endocrine therapy and 61% (95%CI 42, 77) for menopausal transition. While the PPV for DBCG-recorded recurrence was 100%, there were more recurrences reported in the medical records than reported in the DBCG database.[38] These parameters will allow us to adjust for measurement errors in our analyses, improving data quality and confidence in the resulting measures of association.[39]

FINDINGS TO DATE

In our preceding ProBe CaRe nested case-control study, where 94% of breast cancer patients were postmenopausal, we compared rates of breast cancer recurrence for women with a polymorphism that impairs the function of CYP2D6 (an enzyme involved in tamoxifen metabolism) to those in women without this polymorphism and found a null association

(adjusted OR = 0.99; 95% CI = 0.76, 1.3).[40] We further evaluated functional variants in the phase II UDP-glucuronosyl transferases, which contributes to deactivation tamoxifen, and again found near null associations.[41] We have assessed drug-drug interactions with concomitant use of selective serotonin reuptake inhibitor antidepressants using the Danish National Prescription Registry, and reported an adjusted odds ratio of 1.1 (95% CI = 0.7, 1.7).[42]

The current ProBe CaRe longitudinal cohort will build on our previous research to address gene-gene and gene-drug interactions that may compromise tamoxifen effectiveness by focusing on premenopausal women and by more comprehensively evaluating variants in the metabolic path. We will use a Bayesian pathway analysis, (the Algorithm for Learning Pathway Analysis (ALPS)),[43] to evaluate the complete tamoxifen metabolic pathway and to allow for identification of gene-gene interactions, while also estimating the net effect of the entire pathway.[44] This analytic approach will allow for incorporation of time-varying information on tamoxifen adherence, use of inhibiting co-medications, comorbidity, and transition to postmenopausal status, while modelling complex gene-gene interactions without issues of sparse data or a reduction in power.[43] ALPS also permits incorporation of prior biologic knowledge regarding the metabolic path of tamoxifen, so that the search space for the algorithm is constrained to pathways consistent with currently understood biology.

The DBCG has a long history of contributions to the scientific community, informing clinical and treatment guidelines for breast cancer.[45, 27, 31] It is thus an indispensable resource for addressing our study aims.

STRENGTHS AND LIMITATIONS

The current ProBe CaRe study is a large prospective cohort nested within a nearly complete source population. The cohort has many strengths, including the completeness of high quality data and a large representative study population from the Danish source population. Our study design allows for thorough assessment of competing explanations for our findings, both through inclusion of a cohort of ER–/TAM– participants and an internal validation study to address possible errors in classification of key variables. It is the first cohort to examine reduced activity of tamoxifen metabolism in premenopausal women with ample sample size. Moreover, all data (except for new laboratory data) were collected from standardized reports submitted to population-based prospective registries. In addition to DCBG data, we can link patient records to drug prescription, morbidity, and mortality data from independently maintained registries to ensure that relevant covariates are considered.

One potential limitation of the ProBe CaRe cohort is the homogeneity of the study sample, as almost all patients are of European descent. However, there is no comparable source with the same level of information quality to allow exploration of our aims in a more diverse study population. Lack of diversity is a potential limitation, but previous studies indicate that our findings may be extrapolated to external populations and can inform the future direction of research in more diverse populations.[46-49]

COLLABORATIONS

ProBe CaRe study data are held and managed by the Department of Clinical Epidemiology (KEA) in Aarhus, Denmark. We welcome collaborations to enhance the utility of the data and biobank and will respond to all inquiries (tlash@emory.edu).

FURTHER DETAILS

Contributors: LC prepared the original draft of the manuscript. AK conducted data analyses and put together the tables. TL, DCF, HTS, SHD, and TA were responsible for study development and planning. DCF and HTS were responsible for application for data access in Denmark. SHD led the collection and preparation of the tumour samples for genotyping and immunohistochemistry assays. PD provided methodological input surrounding the pharmacogenetic aspects of the study. PC, BE, and RS provided methodologic insight into study design, operationalization of the study aims, and clinical insights. All authors provided critical review of the manuscript and approved the final version.

Funding The ProBe CaRe cohort study was established with funding from the National Cancer Institute at the U.S. National Institutes of Health (R01 CA166825; PI: Lash).

Competing Interests: The authors declare no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committiee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The research is approved by the regional ethical board in Denmark and by the Institutional Review Boards in the US. The study does not contain any animal experiments performed by any of the authors.

Data sharing statement Once the initial data analyses are complete, we will be open to collaborations with outside investigtors as permitted by the IRBs of participating sites. In particular, we will encourage collaborations with researchers whose expertise is under-represented on our research team. To become a collaborator, a researcher will be required to

submit an application, which will undergo both a scientific and IRB review. In view of the complexity of the database and requirements of Danish Law, interested investigators will be asked to form collaborative arragnement with the ProBe CaRe investigators rather than sharing data directly.

Informed Consent Informed consent was not required. Registry-based research is exempt from informed consent requirements according to Danish law.

REFERENCES

1. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians*. 2016;66(1):31-42.

2. Stewart B, Wild CP. World cancer report 2014. 2014.

3. Burstein HJ, Griggs JJ, Prestrud AA, Temin S. American society of clinical oncology clinical practice guideline update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J of Oncol Practice*. 2010;6(5):243-6.

4. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, et al. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011;22(8):1736-47.

5. 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.

6. Stearns V, Johnson MD, Rae JM, Morocho A, Novielli A, Bhargava P, et al. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J of Natl Cancer Inst*. 2003;95(23):1758-64.

7. Jin Y, Desta Z, Stearns V, Ward B, Ho H, Lee KH, et al. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J of Natl Cancer Inst*. 2005;97(1):30-9.

8. Gjerde J, Hauglid M, Breilid H, Lundgren S, Varhaug JE, Kisanga ER, et al. Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. *Ann Oncol* : official journal of the European Society for Medical Oncology. 2008;19(1):56-61.

9. Lim JS, Chen XA, Singh O, Yap YS, Ng RC, Wong NS, et al. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP2C19 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *Br J Clin Pharmacol*. 2011;71(5):737-50.

10. Murdter TE, Schroth W, Bacchus-Gerybadze L, Winter S, Heinkele G, Simon W, 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. *Clin Pharmacol Ther*. 2011;89(5):708-17.

11. Cronin-Fenton DP, Damkier P, Lash TL. Metabolism and transport of tamoxifen in relation to its effectiveness: new perspectives on an ongoing controversy. *Future Oncol*. 2014;10(1):107-22.

12. Province MA, Goetz MP, Brauch H, Flockhart DA, Hebert JM, Whaley R, et al. CYP2D6 genotype and adjuvant tamoxifen: meta-analysis of heterogeneous study populations. *Clin Pharmacol Ther*. 2014;95(2):216-27.

13. Burstein HJ, Temin S, Anderson H, Buchholz TA, Davidson NE, Gelmon KE, 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):2255-69.

14. Network NCC. NCCN practice guidelines in oncology—breast cancer v. 2.2008. 2009.

15. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* : official journal of the European Society for Medical Oncology. 2011;22(8):1736-47.

16. Sideras K, Ingle JN, Ames MM, Loprinzi CL, Mrazek DP, Black JL, et al. Coprescription of tamoxifen and medications that inhibit CYP2D6. *J Clin Oncol*. 2010;28(16):2768-76.

17. Johnson MD, Zuo H, Lee K-H, Trebley JP, Rae JM, Weatherman RV, et al. Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res Treat*. 2004;85(2):151-9.

18. Lash TL, Lien EA, Sørensen HT, Hamilton-Dutoit S. Genotype-guided tamoxifen therapy: time to pause for reflection? *Lancet Oncol*. 2009;10(8):825-33.

19. Sherman BM, Chapler FK, Crickard K, Wycoff D. Endocrine consequences of continuous antiestrogen therapy with tamoxifen in premenopausal women. *J Clin Investig*. 1979;64(2):398.

20. Lindberg K, Helguero LA, Omoto Y, Gustafsson JA, Haldosen LA. Estrogen receptor beta represses Akt signaling in breast cancer cells via downregulation of HER2/HER3 and upregulation of PTEN: implications for tamoxifen sensitivity. *Breast Cancer Res* : BCR. 2011;13(2):R43.

21. Murphy LC, Watson PH. Is oestrogen receptor-beta a predictor of endocrine therapy responsiveness in human breast cancer? *Endocr-Relat Cancer*. 2006;13(2):327-34.

22. Fox EM, Davis RJ, Shupnik MA. ERbeta in breast cancer--onlooker, passive player, or active protector? *Steroids*. 2008;73(11):1039-51.

23. Wu X, Subramaniam M, Grygo SB, Sun Z, Negron V, Lingle WL, et al. Estrogen receptor-beta sensitizes breast cancer cells to the anti-estrogenic actions of endoxifen. *Breast Cancer Res: BCR*. 2011;13(2):R27.

24. Marchais-Oberwinkler S, Henn C, Moller G, Klein T, Negri M, Oster A, et al. 17beta-Hydroxysteroid dehydrogenases (17beta-HSDs) as therapeutic targets: protein structures, functions, and recent progress in inhibitor development. *J Steroid Biochem Mol Biol*. 2011;125(1-2):66-82.

25. Speirs V, Green AR, Atkin SL. Activity and gene expression of 17beta-hydroxysteroid dehydrogenase type I in primary cultures of epithelial and stromal cells derived from normal and tumourous human breast tissue: the role of IL-8. *J Steroid Biochem Mol Biol*. 1998;67(3):267-74.

26. Poirier D. Contribution to the development of inhibitors of 17beta-hydroxysteroid dehydrogenase types 1 and 7: key tools for studying and treating estrogen-dependent diseases. *J Steroid Biochem Mol Biol*. 2011;125(1-2):83-94.

27. Blichert-Toft M, Christiansen P, Mouridsen HT. Danish Breast Cancer Cooperative Group--DBCG: History, organization, and status of scientific achievements at 30-year anniversary. *Acta Oncol*. 2008;47(4):497-505.

28. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541-9.

29. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492-502.

30. Blows FM, Driver KE, Schmidt MK, Broeks A, Van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med*. 2010;7(5):e1000279.

31. Moller S, Jensen MB, Ejlertsen B, Bjerre KD, Larsen M, Hansen HB, et al. The clinical database and the treatment guidelines of the Danish Breast Cancer Cooperative Group (DBCG); its 30-years experience and future promise. *Acta Oncol*. 2008;47(4):506-24.

32. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83.

33. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol*. 2011;11(1):83.

34. Pottegard A, Schmidt SA, Wallach-Kildemoes H, Sorensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol*. 2016.

35. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clin Epidemiol*. 2010;2:51.

36. Cronin-Fenton DP, Hellberg Y, Lauridsen KL, Ahern TP, Garne JP, Rosenberg C, et al. Factors associated with concordant estrogen receptor expression at diagnosis and centralized re-assay in a Danish population-based breast cancer study. *Acta Oncol*. 2012;51(2):254-61.

37. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol*. 2010;2:273.

38. Cronin-Fenton DP, Kjaersgaard A, Ahern TP, Mele M, Ewertz M, Hamilton-Dutoit S, et al. Validity of Danish Breast Cancer Group (DBCG) registry data used in the predictors of breast cancer recurrence (ProBeCaRe) premenopausal breast cancer cohort study. *Acta Oncol*. 2017:1-6.

39. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-85.

40. Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA, et al. CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. *J of Natl Cancer Inst*. 2011;103(6):489-500.

41. Ahern TP, Christensen M, Cronin-Fenton DP, Lunetta KL, Soiland H, Gjerde J, et al. Functional polymorphisms in UDP-glucuronosyl transferases and recurrence in tamoxifen-treated breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20(9):1937-43.

42. Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA, et al. Breast cancer recurrence risk related to concurrent use of SSRI antidepressants and tamoxifen. *Acta Oncol*. 2010;49(3):305-12.

43. Baurley JW, Conti DV, Gauderman WJ, Thomas DC. Discovery of complex pathways from observational data. *Stat Med*. 2010;29(19):1998-2011.

44. Cronin-Fenton DP, Lash TL. Clinical epidemiology and pharmacology of CYP2D6 inhibition related to breast cancer outcomes. *Expert Rev Clin Pharmacol*. 2011;4(3):363-77.

45. Jensen AR, Storm HH, Moller S, Overgaard J. Validity and representativity in the Danish Breast Cancer Cooperative Group--a study on protocol allocation and data validity from one county to a multi-centre database. *Acta Oncol.* 2003;42(3):179-85.

46. Kraft P. Population stratification bias: more widespread than previously thought. *Epidemiology*. 2011;22(3):408-9.

47. Jaja C, Burke W, Thummel K, Edwards K, Veenstra DL. Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet*. 2008;11(3):141-9.

48. Cai WM, Nikoloff DM, Pan RM, de Leon J, Fanti P, Fairchild M, et al. CYP2D6 genetic variation in healthy adults and psychiatric African-American subjects: implications for clinical practice and genetic testing. *Pharmacogenomics J*. 2006;6(5):343-50.

49. Gaedigk A, Isidoro-Garcia M, Pearce RE, Sanchez S, Garcia-Solaesa V, Lorenzo-Romo C, et al. Discovery of the nonfunctional CYP2D6 31 allele in Spanish, Puerto Rican, and US Hispanic populations. *Eur J Clin Pharmacol*. 2010;66(9):859-64.

Figure Legends

Figure 1: Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped.

Figure 2 Selection of study sample and group based on the inclusion criteria. The source population consisted of 8,047 premenopausal women diagnosed with a first primary stage I-III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2011. After exclusions (n=2,088), the study population consists of 5,959 patients in the ProBe CaRe study.

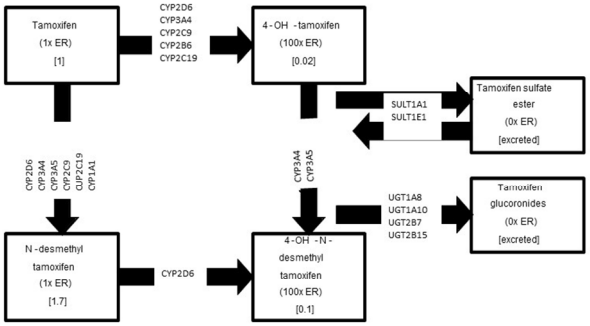


Figure 1: Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped.

108x60mm (300 x 300 DPI)

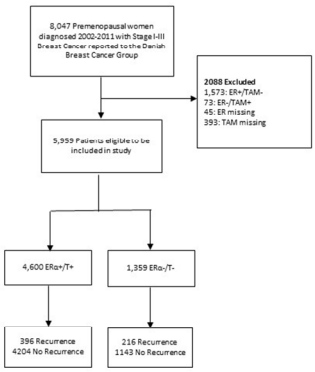


Figure 2 Selection of study sample and group based on the inclusion criteria. The source population consisted of 8,047 premenopausal women diagnosed with a first primary stage I-III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2011. After exclusions (n=2,088), the study population consists of 5,959 patients in the ProBe CaRe study.

108x60mm (300 x 300 DPI)

Electronic Supplementary Tables and Figures

Title: Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Author List: Lindsay J Collin, Deirdre P Cronin-Fenton, Thomas P Ahern, Peer M Christiansen, Per Damkier, Bent Ejlertsen, Stephen Hamilton-Dutoit, Anders Kjærsgaard, Rebecca A Silliman, Henrik Toft Sørensen, Timothy L Lash

To whom correspondence should be addressed: Lindsay J Collin, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE Atlanta GA 30322; Phone: 1-530-386-3341; Email: lindsay.jane.collin@emory.edu, ORCID: 0000-0003-0170-2592

Electronic Supplementary Table 1 Comparison of clinical and tumour characteristics by availability of tumour specimen for each ERα+/T+ and ERα–/T– group, among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study.

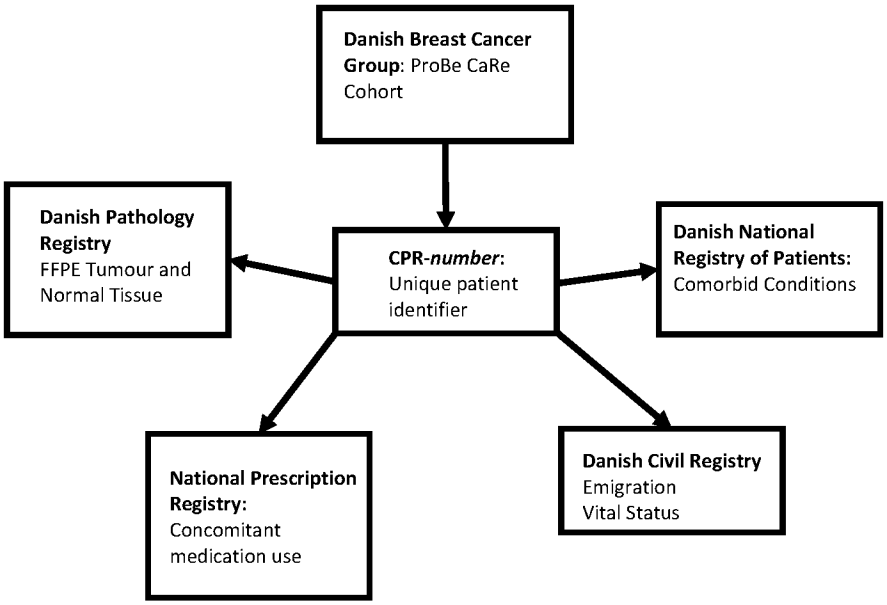
Patient and tumour characteristics	ER+/TAM+				ER–/TAM–			
	Tumour Specimen Available							
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
Total	3959	86	640	14.0	1139	84.0	220	16.0
Age at diagnosis								
<35	197	5	25	3.9	156	14	26	12
35-49	408	10	79	12	178	16	51	23
40-44	973	25	149	23	270	24	51	23
45-49	1455	37	213	33	324	28	61	28
50+	926	23	174	27	211	19	31	14
Stage at diagnosis								
Stage I	1020	26	164	26	334	29	68	31
Stage II	2125	54	350	55	590	52	112	51
Stage III	796	20	121	19	207	18	39	18
Unknown stage	18	0.5	5	0.8	8	0.7	1	0.5
Tumor size								
< 2mm	2267	57	379	59	553	49	124	56
2 - <5mm	1544	39	235	37	540	47	92	42
> 5mm	133	3.4	23	3.6	41	3.6	3	1.4
Unknown	15	0.4	3	0.5	5	0.4	1	0.5
Number of metastatic lymph nodes								
0	1469	37	235	37	582	51	113	51
1	982	25	166	26	194	17	44	20
2	493	13	90	14	101	8.3	15	6.8
3+	1004	25	147	23	259	23	47	21
Unknown	11	0.3	2	0.3	3	0.3	1	0.5
Lymph node evaluation								
No	7	0.2	1	0.2	2	0.2	1	0.5
Yes	3952	100	639	100	1137	99.8	219	99.5
Histologic grade								
Unsuitable	7	0.2	3	0.5	11	1	2	0.9
I	839	21	116	18	21	1.8	0	0
II	2024	51	366	57	182	16	34	15
III	822	21	128	20	737	65	147	67
Unknown	267	6.7	27	4.2	188	17	37	17
Type of primary surgery								
Mastectomy	1759	44	273	43	538	47	89	40

Lumpectomy	2200	56	367	57	601	53	131	60
Progesterone receptor status								
PR-	308	7.8	75	12	919	80.7	202	91.8
PR+	2245	57	435	68	17	1.5	2	0.9
Unknown/Not measured	1406	35	130	20	203	17.8	16	7.3
HER2 status								
HER2-	2489	63	397	62	579	51	113	51
HER2+	532	13	87	14	300	26	54	25
Unknown/Not measured	938	24	156	24	260	23	53	24
Intention to treat with chemotherapy								
No	126	3.2	18	3	11	1	3	1.4
Yes	3833	97	622	97	1128	99	217	99
Chemotherapy								
No	386	9.7	51	8	85	7.5	24	11
Yes	3573	90	589	92	1054	93	196	89
Intention to treat with Tamoxifen								
No	62	1.6	8	1.3	1131	99	220	100
Yes	3897	98	632	99	8	0.7	0	0
Radiation therapy								
No	565	14	90	14	230	20	37	17
Yes	3394	86	550	86	909	80	183	83
Anti-HER2 therapy								
No	2489	63	397	62	579	51	113	51
Yes	532	13	87	14	300	26	54	25
Unknown	938	24	156	24	260	23	53	24
Recurrence								
No	3625	92	579	90	962	84	181	82
Yes	334	8.4	61	9.5	177	16	39	18
Another malignancy								
No	3911	99	632	99	1126	99	215	98
Yes	48	1.2	8	1.3	13	1.1	5	2.3
Dead at end-of-follow-up								
No	3648	92	591	92	932	82	183	83
Yes	311	7.9	49	7.7	207	18	37	17
Charlson Comorbidity Score								
0	3949	100	637	100	1124	99	220	100
1	3	0.1	1	0.2	4	0.4	0	0
2	0	0	2	0.3	3	0.3	0	0
3+	7	0.2	0	0	8	0.7	0	0

Electronic Supplementary Table 2 Comparison of clinical and tumour characteristics by availability of non-neoplastic tissue specimen for each ERα+/T+ and ERα–/T– group, among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study

Patient and tumour characteristics	ER+/TAM+				ER–/TAM–			
	Non-neoplastic Tissue Sample Available							
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
Total	3746	92.0	853	19.0	1082	80.0	277	20.0
Age at diagnosis								
<35	188	5	34	4	157	15	25	9
35-49	392	11	95	11	161	15	68	25
40-44	916	24	206	24	267	25	54	20
45-49	1364	36	304	36	309	29	76	27
50+	886	24	214	25	188	17	54	20
Stage at diagnosis								
Stage I	966	26	218	26	317	29	85	31
Stage II	2005	54	470	55	560	52	142	51
Stage III	759	20	158	19	199	18	47	17
Unknown stage	16	0.4	7	0.8	6	0.6	3	1.1
Tumor size								
< 2mm	2150	57	496	58	522	48	155	56
2 - <5mm	1450	39	329	39	515	48	117	42
> 5mm	131	3.5	25	2.9	41	3.8	3	1.1
Unknown	15	0.4	3	0.4	4	0.4	2	0.7
Number of metastatic lymph nodes								
0	1385	37	319	37	556	51	139	50
1	929	25	219	26	177	16	61	22
2	482	13	101	12	97	9	19	6.9
3+	941	25	210	25	250	23	56	20
Unknown	9	0.2	4	0.5	2	0.2	2	0.7
Lymph node evaluation								
No	6	0.2	2	0.2	1	0.1	2	0.7
Yes	3740	100	851	100	1081	100	275	99
Histologic grade								
Unsuitable	5	0.1	5	0.6	10	0.9	3	1.1
I	791	21	164	19	17	1.6	4	1.4
II	1882	50	508	60	166	15	50	18
III	806	22	144	17	708	65	176	64
Unknown	262	7	32	3.8	181	17	44	16
Type of primary surgery								

Mastectomy	1696	45	336	39	524	48	103	37
Lumpectomy	2050	55	517	61	558	52	174	63
Progesterone receptor status								
PR-	281	7.5	102	12	860	80	261	94
PR+	2019	54	661	77	15	1.4	4	1.4
Unknown/Not measured	1446	39	90	11	207	19	12	4.3
HER2 status								
HER2-	2304	62	582	68	543	50	149	54
HER2+	525	14	94	11	282	26	72	26
Unknown/Not measured	917	25	177	21	257	24	56	20
Intention to treat with chemotherapy								
No	122	3.3	22	2.6	8	0.7	6	2.2
Yes	3624	97	831	97	1074	99	271	98
Chemotherapy								
No	381	10	56	6.6	87	8	22	8
Yes	3365	90	797	93	995	92	255	92
Intention to treat with Tamoxifen								
No	59	1.6	11	1.3	1075	99	276	100
Yes	3687	98	842	99	7	0.6	1	0.4
Radiation therapy								
No	547	15	108	13	218	20	49	18
Yes	3199	85	745	87	864	80	228	82
Anti-HER2 therapy								
No	2304	62	582	68	543	50	149	54
Yes	525	14	94	11	282	26	72	26
Unknown	917	25	177	21	257	24	56	20
Recurrence								
No	3433	92	771	90	918	85	225	81
Yes	313	8.4	82	9.6	164	15	52	19
Another malignancy								
No	3695	99	848	99	1069	99	272	98
Yes	51	1.4	5	0.6	13	1.2	5	1.8
Dead at end-of-follow-up								
No	3449	92	790	93	890	82	225	81
Yes	297	7.9	63	7.4	192	18	52	19
Charlson Comorbidity Score								
0	3733	100	853	100	1067	99	277	100
1	4	0.1	0	0	4	0.4	0	0
2	2	0.1	0	0	3	0.3	0	0
3+	7	0.2	0	0	8	0.7	0	0



BMJ Open

Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021805.R1
Article Type:	Cohort profile
Date Submitted by the Author:	08-Mar-2018
Complete List of Authors:	Collin, Lindsay; Emory University School of Public Health, Epidemiology Cronin-Fenton, Deirdre; Aarhus Universitetshospital, Ahern, Thomas; The Robert Larner M.D. College of Medicine at The University of Vermont Christiansen, P; Aarhus Universitet Damkier, Per; University of Southern Denmark, Dept. Clinical Biochemistry and Pharmacology Ejlertsen, Bent; Copenhagen University Hospital, Oncology Hamilton Dutoit, Stephen; Aarhus Universitetshospital, Institute of Pathology Kjærsgaard, Anders; Aarhus Universitetshospital, Department of Clinical Epidemiology Silliman, Rebecca; Boston University School of Medicine; Aarhus Universitetshospital, Department of Clinical Epidemiology Sørensen, Henrik T.; Aarhus University Hospital, Department of Clinical Epidemiology Lash, Timothy; Emory University Rollins School of Public Health, Department of Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Pharmacology and therapeutics
Keywords:	Cohort Study, Breast tumours < ONCOLOGY, Pharmacogenetics, EPIDEMIOLOGY

Title: Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Author List: Lindsay J Collin[a], Deirdre P Cronin-Fenton[b], Thomas P Ahern[c], Peer M Christiansen[d,e,f], Per Damkier[g,h], Bent Ejlertsen[f,i], Stephen Hamilton-Dutoit[j], Anders Kjærsgaard[b], Rebecca A Silliman[b,k], Henrik Toft Sørensen[b,l], Timothy L Lash[a,b]

[a]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [b]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; [c]The Robert Larner, M.D. College of Medicine at The University of Vermont, Burlington, VT, USA; [d]Breast Unit, Surgical Department, Randers Regional Hospital, Randers, Denmark; [e]Department of Oncology, Odense University Hospital, Odense, Denmark; [f]Danish Breast Cancer Group, Copenhagen, Denmark; [g]Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark; [h]Department of Clinical Research, University of Southern Denmark, Denmark; [i]Rigshospitalet, Copenhagen, Denmark; [j]Institute of Pathology, Aarhus University Hospital, Aarhus, Denmark; [k]Boston University School of Medicine, Boston University, Boston, MA, USA; [l]Department of Health Research & Policy (Epidemiology), Stanford University, Stanford, CA, USA;

To whom correspondence should be addressed: Lindsay J Collin, Department of Epidemiology, 1518 Clifton Rd NE Atlanta GA 30322; Phone: 1-530-386-3341; Email: lindsay.jane.collin@emory.edu, ORCID: 0000-0003-0170-2592

Word Count: 3242

Keywords: Cohort study, breast cancer, pharmacogenetics, epidemiology

ABSTRACT

Purpose: The Predictors of Breast Cancer Recurrence (ProBe CaRe) study was established to evaluate modification of tamoxifen effectiveness in premenopausal women through reduced activity of tamoxifen-metabolizing enzymes. It comprehensively evaluates the effects of pharmacogenetic variants, use of concomitant medications, and biomarkers involved in oestrogen metabolism on breast cancer recurrence risk.

Participants: The ProBe CaRe study was established using resources from the Danish Breast Cancer Group (DBCG), including 5,959 premenopausal women diagnosed with stage I–III primary breast cancer between 2002 and 2010 in Denmark. Eligible participants were divided into two groups based on oestrogen receptor alpha (ERα) expression and receipt of tamoxifen therapy (TAM), 4,600 are classified as ERα+/TAM+ and 1,359 are classified as ERα–/TAM–. The ProBe CaRe study is a population-based cohort study nested in a nearly complete source population, clinical, tumour, and demographic data were abstracted from DBCG registry data. Linkage to Danish registries allows for abstraction of information regarding comorbid conditions, comedication use, and mortality. Formalin-fixed paraffin-embedded (FFPE) tissue samples have been prepared for DNA extraction and immunohistochemical assay.

Findings to date: To mitigate incorrect classification of patients into specific categories we conducted a validation substudy. We compared data acquired from registry and from medical record review to calculate positive and negative predictive values (PPV and NPV). We observed PPVs near 100% for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment. We found that the PPVs were

96% (95%CI 83, 100) for change in endocrine therapy, and 61% (95%CI 42, 77) for menopausal transition.

Future plans: The ProBe CaRe cohort study is well-positioned to comprehensively examine pharmacogenetic variants. We will use a Bayesian pathway analysis to evaluate the complete tamoxifen metabolic path to allow for gene-gene interactions, incorporating information of other important patient characteristics.

Strengths and Limitations of this Study

- One potential limitation of the ProBe CaRe study is the homogeneity of the study sample, as almost all are of European descent.
- In addition to being the first large epidemiologic study to examine reduced activity of tamoxifen metabolism in premenopausal women, this study is strengthened by completeness of high quality data.
- Our study includes a validation substudy to mitigate errors from incorrect classification of patients into specific categories of key analytic variables.

INTRODUCTION

Endocrine therapy improves survival in patients with breast cancer regardless of axillary lymph node status.[1] The Predictors of Breast Cancer Recurrence (ProBe CaRE) cohort study was established to evaluate modification of tamoxifen effectiveness in premenopausal women through reduced activity of tamoxifen-metabolizing enzymes. Candidates for adjuvant tamoxifen therapy include stage I–IV breast cancer patients with ER positive tumours, who constitute about two-thirds[2] of the approximately 1.7 million newly diagnosed breast cancer patients each year worldwide.[3] Current guidelines recommend that premenopausal patients with ER alpha positive (ERα+) cancers receive tamoxifen for five to ten years,[4-6] which reduces recurrence risk by nearly half,[6] and that tamoxifen may be offered to post-menopausal women with ERα+ cancers as an alternative to aromatase inhibitors. Tamoxifen metabolism is complex, but is principally catalysed by cytochrome P450 enzymes. Some metabolites bind with the ER with significantly greater affinity than tamoxifen itself, especially endoxifen, which has the highest ER-binding activity among tamoxifen metabolites. Activity of the enzymes involved in tamoxifen metabolism can vary between individuals due to inherited gene variants[7-11] or use of comedications.[8, 7] Although many studies have explored the association between these gene variants or use of comedications and failure of tamoxifen treatment,[12, 13] which manifests clinically as a recurrence, the interpretation of these studies remains controversial. Current clinical guidelines do not recommend genotyping these variant alleles to support treatment decisions,[1, 14, 15] but do recommend avoiding inhibiting comedications.[16]

To date, little available evidence on this topic is specific to premenopausal breast cancer patients. The competition between oestrogen and tamoxifen for ER binding is highly important for these patients because tamoxifen is a first-line guideline-recommended therapy[15, 1] for

premenopausal patients and because premenopausal women have higher concentrations of oestrogens to compete with tamoxifen for ER binding. Oestradiol, the most active oestrogen metabolite, binds with the ER with approximately the same affinity as endoxifen.[17] Premenopausal women have tenfold higher concentrations of oestradiol than postmenopausal women[18] and oestradiol concentrations tend to increase during tamoxifen therapy.[19, 18] This suggests that inhibition of tamoxifen-metabolizing enzymes is more likely to decrease effectiveness in premenopausal women, yet they have been seldom studied in this topic area.

Research questions

We established a premenopausal cohort of breast cancer patients to fill this important evidence gap, with the following primary study aims:

(1) Assess pharmacogenetics of tamoxifen metabolism and risk of breast cancer recurrence

We will assess the pharmacogenetics of tamoxifen metabolism by genotyping 32 variants in 15 enzymes (**Table 1**) thought to affect the concentration of the most active tamoxifen metabolites, and will evaluate the association between these variants and breast cancer recurrence in tamoxifen treated premenopausal breast cancer patients. Each of the selected enzymes is involved in at least one step in the tamoxifen metabolic pathway (**Figure 1**). Interactions with comedications that inhibit these metabolic enzymes also will be evaluated.

TABLE 1: Selected Functional variants and inhibitor comedications in genes whose enzymes metabolize tamoxifen

Gene	Number of selected functional variants	SNPs	Inhibitor comedications
CYP2D6	5	rs1065852, rs16947, rs3892097, rs28371706, rs28371725	bupropion, cinacalcet, fluoxetine, paroxetine, quinidine, duloxetine, sertraline, terbinafine, amiodarone, cimetidine, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, telithromycin, aprepitant, erythromycin, fluconazole, verapamil, diltiazem, cimetidine, voriconazole
CYP3A	1	rs10273424	
CYP3A5	1	rs776746	
CYP2C9	2	rs1057910, rs1799853	fluconazole, amiodarone, voriconazole
CYP2C19	2	rs12248560, rs4244285	
CYP2B6	2	rs3745274, rs8192709	
CYP1A1	1	rs1048943	
SULT1A1	3	rs1042157, rs1801030, rs9282861	
SULT1E1	2	rs3775775, rs3775778	
UGT2B7	1	rs7434332	
UGT2B10	1	rs294769	
ABCC2	3	rs3740065, rs717620, rs8187710	
ABCG2	3	rs1564481, rs2231164, rs2622604	
ABCB1	4	rs10248420, rs1045642, rs1128503, rs2032582	
UGT2B15	1	rs1902023	

(2) Assess the interaction between the pharmacogenetics of tamoxifen metabolism and oestrogen receptor beta (ERβ) expression

We will assess the effect of interaction between the pharmacogenetics of tamoxifen metabolism and ERβ expression on risk of breast cancer recurrence. Previous studies have shown that co-expression of ERβ is associated with improved survival among patients with ERα+ tumours who are treated with tamoxifen.[20, 21] The ERβ receptor opposes ERα-mediated proliferation.[20] Tumours that express both ERα and ERβ are less aggressive than tumours that express homodimers of ERα,[22, 20] due to the attenuated stimulation response from ERα/ERβ heterodimers. This suggests that metabolic inhibition may only affect ERβ– tumours. *In vitro* analyses have demonstrated that in ERα+/ERβ+ MCF7 cells, proliferation is inhibited by a wide range of endoxifen concentrations.[23] Still, in ERα+/ERβ– MCF7 cells, only physiologically high endoxifen concentrations inhibit proliferation,[23] indicating that metabolic inhibition affects risk of recurrence only when ERβ is absent.

(3) Assess interaction between inhibition of tamoxifen metabolism and oestrogen-regulating enzymes

Finally, we will assess the association between tumour expression of 17β-hydroxysteroid dehydrogenase 1 and 2 (17βHSD1 and 17βHSD2) and breast cancer recurrence. 17βHSD1 catalyses the conversion of oestrone (E1) to the most potent form of oestrogen, oestradiol (E2), and 17βHSD2 catalyses the reverse reaction.[24] E2 has the highest binding affinity for ER, and endoxifen acts through competitive inhibition at the receptor-binding site.[24] In breast tumour tissue, 17βHSD1 is more highly expressed than 17βHSD2. The opposite is usually observed in adjacent normal tissue.[25] Tumours with higher capacity to produce E2 endogenously through increased expression of the 17βHSD1 enzyme are more likely to overwhelm the tamoxifen

metabolites in competition for ER binding, affecting tamoxifen effectiveness. These enzymes are ideal therapeutic targets to modulate E2 concentrations in tumour cells, and candidate inhibitors have been developed.[26, 24] We will evaluate whether disequilibrium of the 17βHSD1 and -2 enzymes (ratio >1) results in compromised tamoxifen effectiveness.

COHORT DESCRIPTION

The ProBe CaRe cohort was established using the resources of the Danish Breast Cancer Group (DBCG) registry. The DBCG registry was established in 1976 and began to register patients in 1977, with the goals of standardizing treatment, facilitating clinical trials, and monitoring outcomes among Danish breast cancer patients.[27] Since its inception, the DBCG has registered over 90% of women diagnosed with breast cancer in Denmark. Breast cancer patients are registered in the DBCG via standardized forms. The registry has a standard protocol to collect information on tumour, treatment, and patient characteristics. Using this information-rich resource, the ProBe CaRe cohort is nested in a nearly complete source population of premenopausal women diagnosed with stage I–III first primary breast cancer between 2002–2010 whose breast cancer was reported to the DBCG. In Denmark, all citizens and legal residents are assigned a Civil Personal Register (CPR) number, a unique 10-digit personal identifier assigned at birth or upon immigration that is used for identification across all national registries (**Electronic Supplementary Figure 1**).[28]

Of the 8047 premenopausal women diagnosed with breast cancer between 2002 and 2010 and recorded in the DBCG registry, 5959 cancers were identified as eligible based on being a stage I–III first primary breast cancer and untreated with neoadjuvant therapy; all others

(n=2088) were excluded. The 5959 eligible patients then were divided into two cohorts based on ERα expression and receipt of tamoxifen therapy (**Figure 2**). To address competing explanations, (for instance if the biomarkers under study affect risk directly rather than mediating the tamoxifen effect), we will also evaluate the risk of recurrence in the subset of women with ERα– tumours who did not receive tamoxifen therapy (T–).

Our final ProBe CaRe study population consists of these 5,959 breast cancer patients divided into a cohort of ERα+/T+ (4600 patients) and a cohort of ERα–/T– (1359 patients). The socio-demographic and clinical characteristics of the two cohorts are described in **Table 2**. The distribution of the clinical and demographic characteristics between the two cohorts (ERα+/T+ vs. ERα–/T–) are relatively similar. They only differ meaningfully with respect to progesterone receptor (PR) status (58% vs. 1.4% PR+, respectively) and human epidermal growth factor receptor 2 (HER2) status (14% vs. 26% HER2+, respectively). With respect to outcomes, the ER–/TAM– cohort has a higher proportion of subjects who experienced recurrence (8.6% vs 16%, respectively) and who died by the end of follow up (7.8% vs. 18% respectively). This pattern is to be expected, as ER– breast cancers generally have a worse prognosis than ER+ cancers, especially within the first five years following diagnosis.[29, 30]

Table 2 Distribution of clinical and tumour characteristics by ER status and receipt of tamoxifen among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study.

Patient and tumour characteristics		**ER+/TAM+**		**ER–/TAM–**	
		N	%	N	%
Total		4600	100	1359	100
Age at diagnosis					
	<35	222	4.8	182	23
	35-39	487	11	229	27
	40-44	1123	24	321	24

45-49	1668	36	385	28
50+	1100	24	242	18
Menopausal status at diagnosis				
Premenopausal	4600	100	1359	100
Stage at diagnosis				
Stage I	1184	26	402	29.6
Stage II	2476	54	702	51.7
Stage III	917	20	246	18.1
Unknown stage	23	0.5	9	0.7
Tumor size				
< 2mm	2646	58	677	50
2 - <5mm	1780	39	632	47
> 5mm	156	3.4	44	3.2
Unknown	18	0.4	6	0.4
Number of metastatic lymph nodes				
0	1704	37	695	51
1	1148	25	238	17
2	583	13	116	9
3+	1152	25	306	23
Unknown	13	0.3	4	0.3
Lymph node evaluation				
No	8	0.2	3	0.2
Yes	4592	100	1356	100
Histologic grade				
Unsuitable	10	0.2	13	1
I	955	21	21	1.5
II	2391	52	216	16
III	950	21	884	65
Unknown	294	6.4	225	17
Type of primary surgery				
Mastectomy	2033	44	627	46
Lumpectomy	2567	56	732	54
Progesterone receptor status				
PR-	383	8.3	1121	83
PR+	2680	58	19	1.4
Unknown/Not measured	1537	33	219	16
HER2 status				
HER2-	2887.00	63	692	51
HER2+	619	14	354	26

Unknown/Not measured	1094	24	313	23
Intention to treat with chemotherapy				
No	144	3	14	1
Yes	4456	97	1345	99
Chemotherapy				
No	437	9	109	8
Yes	4163	91	1250	92
Intention to treat with Tamoxifen				
No	70	1.5	1351	99
Yes	4530	98	8	0.6
Radiation therapy				
No	655	14	267	20
Yes	3945	86	1092	80
Anti-HER2 therapy				
No	2887	63	692	51
Yes	619	13	354	26
Unknown	1094	24	313	23
Recurrence				
No	4204	91	1143	84
Yes	396	8.6	216	16
Another malignancy				
No	4544	99	1341	99
Yes	56	1.2	18	1.3
Dead at end-of-follow-up				
No	4239	92	1115	82
Yes	361	8	244	18
Charlson Comorbidity Score				
0	4587	99	1344	99
1	4	0.1	4	0.3
2	2	0	3	0.2
3+	7	0.2	8	0.6

Cohort follow-up

Women diagnosed with breast cancer and subsequently enrolled in the DBCG registry undergo semi-annual examinations during the first five years after diagnosis and annual examinations during years six to ten.[31] Women undergoing treatment for breast cancer receive endocrine therapy through the Danish government and obtain their medicine at the hospital, which will be used to estimate tamoxifen adherence. Members of both the ERα+/T+ and the

ERα–/T– ProBe CaRe cohorts have been followed from breast cancer diagnosis to the first of (a) recurrence, (b) death, (c) ten years of follow-up, (d) loss to follow-up due to emigration, (e) another malignancy, or (f) the end of the study follow-up period. Breast cancer recurrence was identified using the DBCG registry. We adopted the DBCG definition of breast cancer recurrence as any type of breast cancer diagnosed subsequent to the initial course of therapy.[31] Recurrences are then further categorized as loco-regional (in the scar or regional lymph nodes), contralateral (opposite breast), distant (all other sites), or unknown (site of recurrence not documented). The date of recurrence is recorded in the DBCG registry, including recurrences diagnosed between scheduled follow-up exams. Mortality and emigration were identified using the Danish Civil Registration System, which is updated daily.[28] Emigration is the only expected source of loss to follow-up and has impacted less than 1% of the study population.

Data Collection

Registry Data

Once participants eligible for ProBe CaRe were selected from the DBCG registry, we extracted clinical and demographic information from the DBCG registry. This information included date and place of diagnosis, tumour characteristics, treatment received, and patient characteristics, which are presented in **Electronic Supplementary Table 1**. We also extracted information on comorbid diseases at time of breast cancer diagnosis, summarized using the Charlson Comorbidity Index.[32, 33] The registry information allows us to update the participants' conditions during study follow-up and therefore to account for time-varying exposures and confounding factors. The CPR number for each patient was used to link cohort members to the Danish National Prescription Registry,[34] which provided information on filled

prescriptions of drugs that inhibit tamoxifen-metabolizing enzymes. This allowed us to assess the drug-drug interactions discussed above.

Biobank

The CPR number for each patient was used to identify the hospital at which the surgery was performed and to locate and retrieve the corresponding formalin-fixed paraffin-embedded (FFPE) tissue samples. The list of ProBe CaRe cohort members and their CPR numbers and hospitals of diagnosis were provided to a medical research technician at the Institute of Pathology, blinded to whether the CPR numbers corresponded to a patient with a recurrence. The technician reviewed a description of the available tissue blocks (routinely available in the Danish pathology registry[35]), identified the tumour-rich and non-neoplastic blocks for each patient, and specified which FFPE blocks should be requested from the hospitals. This list of blocks to be requested was then returned to the Department of Clinical Epidemiology, which prepared and mailed the request letters to the pathology archives at the respective Danish hospitals. Staff at the hospital pathology archives returned the blocks to the Department of Clinical Epidemiology, which assigned a project identification number to the block and then provided it to the Institute of Pathology. The project identification number maintained blinding of laboratory personnel to whether blocks corresponded to patients with a recurrence. The Department of Clinical Epidemiology maintains the key linking the project identification number for the blocks to the clinical data, including recurrence status.

Non-neoplastic tissue samples are taken routinely from normal adjacent tissue or cancer-free lymph nodes resected during breast cancer surgery and were used as controls in creation of tissue microarrays (TMAs). Of the 4,600 patients included in the ERα+/T+ cohort, 4,599 patients had samples evaluated by clinicians, and tumour samples were available for 3,959 (86%).

Among the ER–/TAM– cohort, 1139 (84%) patients had tumour samples available. Distribution of clinical and demographic characteristics among patients with, and without, available tumour samples are described in **Table 3**. Of patients included in the ERα+/T+ cohort, 2746 (82%) had a non-neoplastic tissue sample available, while 1082 (80%) patients in the ER–/TAM– cohort had non-neoplastic tissue samples available. Distributions of demographic and clinical characteristics among patients with and without non-neoplastic tissue samples are summarized in **Electronic Supplementary Table 2**.

Table 3 Summary of exposure, covariate, and outcome variables collected in the ProBe CaRe study

Exposures	**Outcome**
Genetic variants	Recurrence
ERα and ERβ	Mortality
17βHSD1 and 17βHSD2	
Biomarkers	
Clinical	**Demographic**
Tumour characteristics	Age
Treatment therapy	Region
Comorbidity	Hospital of Diagnosis
Medication History and Use	

Sections of collected tissue blocks have been prepared for DNA extraction and immunohistochemical (IHC) assay. In accordance with the study's primary aims, we will genotype 32 variants across 15 genes known to be involved in tamoxifen metabolism, in order to predict extent of metabolic inhibition. We will also re-assay ERα expression to ensure correct classification of the two cohorts, as original ERα expression was measured in different pathology laboratories using different methods. In our previous case-control study of post-menopausal

patients diagnosed during 1985–2001, we reported 95% concordance of positive ERα expression between initial assays and reassays and 74% concordance of negative ERα expression between initial assays and reassays.[36] ERβ expression will be assayed using IHC in TMAs to assess its possible modification of tamoxifen metabolic inhibition. Expression of the enzymes 17βHSD1 and 17βHSD2 also will be assayed using IHC to address the study aim examining whether the ratio of these two enzymes modulates tamoxifen's efficacy in preventing breast cancer recurrence. The aforementioned assays of biomarkers are the primary starting point. However, we anticipate that the study will yield a substantial tumour biobank and ultimately provide a valuable resource to researchers for further characterization of prognostic and predictive biomarkers in premenopausal breast cancer.

Validation substudy

Registry data are not error-free.[37] To mitigate incorrect classification of patients into specific categories, we conducted a validation substudy.[38] By comparing data procured both from the registry and from medical record review, we calculated positive and negative predictive values (PPV and NPV, respectively) and their corresponding confidence intervals for key analytic variables. We observed near perfect PPVs for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy, and tamoxifen treatment. The PPVs were 96% (95% CI 83, 100) for change in endocrine therapy and 61% (95%CI 42, 77) for menopausal transition. While the PPV for DBCG-recorded recurrence was 100%, there were more recurrences reported in the medical records than reported in the DBCG database.[38] These parameters will allow us to adjust for measurement errors in our analyses, improving data quality and confidence in the resulting measures of association.[39]

FINDINGS TO DATE

In our preceding ProBe CaRe nested case-control study, where 94% of breast cancer patients were postmenopausal, we compared rates of breast cancer recurrence for women with a polymorphism that impairs the function of CYP2D6 (an enzyme involved in tamoxifen metabolism) to those in women without this polymorphism and found a null association (adjusted OR = 0.99; 95% CI = 0.76, 1.3).[40] We further evaluated functional variants in the phase II UDP-glucuronosyl transferases, which contributes to deactivation tamoxifen, and again found near null associations.[41] We have assessed drug-drug interactions with concomitant use of selective serotonin reuptake inhibitor antidepressants using the Danish National Prescription Registry, and reported an adjusted odds ratio of 1.1 (95% CI = 0.7, 1.7).[42]

The current ProBe CaRe longitudinal cohort will build on our previous research to address gene-gene and gene-drug interactions that may compromise tamoxifen effectiveness by focusing on premenopausal women and by more comprehensively evaluating variants in the metabolic path. We will use a Bayesian pathway analysis, (the Algorithm for Learning Pathway Analysis (ALPS)),[43] to evaluate the complete tamoxifen metabolic pathway and to allow for identification of gene-gene interactions, while also estimating the net effect of the entire pathway.[44] This analytic approach will allow for incorporation of time-varying information on tamoxifen adherence, use of inhibiting co-medications, comorbidity, and transition to postmenopausal status, while modelling complex gene-gene interactions without issues of sparse data or a reduction in power.[43] ALPS also permits incorporation of prior biologic knowledge regarding the metabolic path of tamoxifen, so that the search space for the algorithm is constrained to pathways consistent with currently understood biology.

The DBCG has a long history of contributions to the scientific community, informing clinical and treatment guidelines for breast cancer.[45, 27, 31] It is thus an indispensable resource for addressing our study aims.

STRENGTHS AND LIMITATIONS

The current ProBe CaRe study is a large prospective cohort nested within a nearly complete source population. The cohort has many strengths, including the completeness of high quality data and a large representative study population from the Danish source population. Our study design allows for thorough assessment of competing explanations for our findings, both through inclusion of a cohort of ER–/TAM– participants and an internal validation study to address possible errors in classification of key variables. It is the first cohort to examine reduced activity of tamoxifen metabolism in premenopausal women with ample sample size. Moreover, all data (except for new laboratory data) were collected from standardized reports submitted to population-based prospective registries. In addition to DCBG data, we can link patient records to drug prescription, morbidity, and mortality data from independently maintained registries to ensure that relevant covariates are considered.

One potential limitation of the ProBe CaRe cohort is the homogeneity of the study sample, as almost all patients are of European descent. However, there is no comparable source with the same level of information quality to allow exploration of our aims in a more diverse study population. Lack of diversity is a potential limitation, but previous studies indicate that our

findings may be extrapolated to external populations and can inform the future direction of research in more diverse populations.[46-49]

COLLABORATIONS

ProBe CaRe study data are held and managed by the Department of Clinical Epidemiology (KEA) in Aarhus, Denmark. We welcome collaborations to enhance the utility of the data and biobank and will respond to all inquiries (tlash@emory.edu).

FURTHER DETAILS

Contributors: LC prepared the original draft of the manuscript. AK conducted data analyses and put together the tables. TL, DCF, HTS, SHD, and TA were responsible for study development and planning. DCF and HTS were responsible for application for data access in Denmark. SHD led the collection and preparation of the tumour samples for genotyping and immunohistochemistry assays. PD provided methodological input surrounding the pharmacogenetic aspects of the study. PC, BE, and RS provided methodologic insight into study design, operationalization of the study aims, and clinical insights. All authors provided critical review of the manuscript and approved the final version.

Funding The ProBe CaRe cohort study was established with funding from the National Cancer Institute at the U.S. National Institutes of Health (R01 CA166825; PI: Lash).

Competing Interests: The authors declare no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

The research is approved by the regional ethical board in Denmark and by the Institutional Review Boards in the US. The study does not contain any animal experiments performed by any of the authors.

Data sharing statement Once the initial data analyses are complete, we will be open to collaborations with outside investigators as permitted by the IRBs of participating sites. In particular, we will encourage collaborations with researchers whose expertise is under-represented on our research team. To become a collaborator, a researcher will be required to submit an application, which will undergo both a scientific and IRB review. In view of the complexity of the database and requirements of Danish Law, interested investigators will be asked to form collaborative arrangement with the ProBe CaRe investigators rather than sharing data directly.

Informed Consent Informed consent was not required. Registry-based research is exempt from informed consent requirements according to Danish law.

Patient and Public Involvement Patients and public were not involved.

REFERENCES

1. Burstein HJ, Temin S, Anderson H, Buchholz TA, Davidson NE, Gelmon KE 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):2255-69.

2. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians* . 2016;66(1):31-42.

3. Stewart B, Wild CP. World cancer report 2014. 2014.

4. Burstein HJ, Griggs JJ, Prestrud AA, Temin S. American society of clinical oncology clinical practice guideline update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J of Oncol Practice* . 2010;6(5):243-6. doi:10.1200/jop.000082.

5. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ et al. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011;22(8):1736-47. doi:10.1093/annonc/mdr304.

6. 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. doi:10.1016/s0140-6736(05)66544-0.

7. Stearns V, Johnson MD, Rae JM, Morocho A, Novielli A, Bhargava P et al. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J of Natl Cancer Inst* . 2003;95(23):1758-64.

8. Jin Y, Desta Z, Stearns V, Ward B, Ho H, Lee KH et al. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J of Natl Cancer Inst* . 2005;97(1):30-9. doi:10.1093/jnci/dji005.

9. Gjerde J, Hauglid M, Breilid H, Lundgren S, Varhaug JE, Kisanga ER et al. Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. *Ann Oncol* : official journal of the European Society for Medical Oncology. 2008;19(1):56-61. doi:10.1093/annonc/mdm434.

10. Lim JS, Chen XA, Singh O, Yap YS, Ng RC, Wong NS et al. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP2C19 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *Br J Clin Pharmacol* . 2011;71(5):737-50. doi:10.1111/j.1365-2125.2011.03905.x.

11. Murdter TE, Schroth W, Bacchus-Gerybadze L, Winter S, Heinkele G, Simon W 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. *Clin Pharmacol Ther*. 2011;89(5):708-17. doi:10.1038/clpt.2011.27.

12. Cronin-Fenton DP, Damkier P, Lash TL. Metabolism and transport of tamoxifen in relation to its effectiveness: new perspectives on an ongoing controversy. *Future Oncol* . 2014;10(1):107-22. doi:10.2217/fon.13.168.

13. Province MA, Goetz MP, Brauch H, Flockhart DA, Hebert JM, Whaley R et al. CYP2D6 genotype and adjuvant tamoxifen: meta-analysis of heterogeneous study populations. *Clin Pharmacol Ther*. 2014;95(2):216-27. doi:10.1038/clpt.2013.186.

14. Network NCC. NCCN practice guidelines in oncology—breast cancer v. 2.2008. 2009.

15. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* : official journal of the European Society for Medical Oncology. 2011;22(8):1736-47. doi:10.1093/annonc/mdr304.

16. Sideras K, Ingle JN, Ames MM, Loprinzi CL, Mrazek DP, Black JL et al. Coprescription of tamoxifen and medications that inhibit CYP2D6. *J Clin Oncol* . 2010;28(16):2768-76.

17. Johnson MD, Zuo H, Lee K-H, Trebley JP, Rae JM, Weatherman RV et al. Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res Treat*. 2004;85(2):151-9.

18. Lash TL, Lien EA, Sørensen HT, Hamilton-Dutoit S. Genotype-guided tamoxifen therapy: time to pause for reflection? *Lancet Oncol* . 2009;10(8):825-33.

19. Sherman BM, Chapler FK, Crickard K, Wycoff D. Endocrine consequences of continuous antiestrogen therapy with tamoxifen in premenopausal women. *J Clin Investig* . 1979;64(2):398.

20. Lindberg K, Helguero LA, Omoto Y, Gustafsson JA, Haldosen LA. Estrogen receptor beta represses Akt signaling in breast cancer cells via downregulation of HER2/HER3 and upregulation of PTEN: implications for tamoxifen sensitivity. *Breast Cancer Res* : BCR. 2011;13(2):R43. doi:10.1186/bcr2865.

21. Murphy LC, Watson PH. Is oestrogen receptor-beta a predictor of endocrine therapy responsiveness in human breast cancer? *Endocr-Relat Cancer*. 2006;13(2):327-34. doi:10.1677/erc.1.01141.

22. Fox EM, Davis RJ, Shupnik MA. ERbeta in breast cancer--onlooker, passive player, or active protector? *Steroids*. 2008;73(11):1039-51. doi:10.1016/j.steroids.2008.04.006.

23. Wu X, Subramaniam M, Grygo SB, Sun Z, Negron V, Lingle WL et al. Estrogen receptor-beta sensitizes breast cancer cells to the anti-estrogenic actions of endoxifen. *Breast Cancer Res: BCR* . 2011;13(2):R27. doi:10.1186/bcr2844.

24. Marchais-Oberwinkler S, Henn C, Moller G, Klein T, Negri M, Oster A et al. 17beta-Hydroxysteroid dehydrogenases (17beta-HSDs) as therapeutic targets: protein structures, functions, and recent progress in inhibitor development. *J Steroid Biochem Mol Biol*. 2011;125(1-2):66-82. doi:10.1016/j.jsbmb.2010.12.013.

25. Speirs V, Green AR, Atkin SL. Activity and gene expression of 17beta-hydroxysteroid dehydrogenase type I in primary cultures of epithelial and stromal cells derived from normal and tumourous human breast tissue: the role of IL-8. *J Steroid Biochem Mol Biol*. 1998;67(3):267-74.

26. Poirier D. Contribution to the development of inhibitors of 17beta-hydroxysteroid dehydrogenase types 1 and 7: key tools for studying and treating estrogen-dependent diseases. *J Steroid Biochem Mol Biol*. . 2011;125(1-2):83-94. doi:10.1016/j.jsbmb.2010.12.007.

27. Blichert-Toft M, Christiansen P, Mouridsen HT. Danish Breast Cancer Cooperative Group--DBCG: History, organization, and status of scientific achievements at 30-year anniversary. *Acta Oncol*. 2008;47(4):497-505. doi:10.1080/02841860802068615.

28. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541-9.

29. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492-502.

30. Blows FM, Driver KE, Schmidt MK, Broeks A, Van Leeuwen FE, Wesseling J et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med*. 2010;7(5):e1000279.

31. Moller S, Jensen MB, Ejlertsen B, Bjerre KD, Larsen M, Hansen HB et al. The clinical database and the treatment guidelines of the Danish Breast Cancer Cooperative Group (DBCG); its 30-years experience and future promise. *Acta Oncol*. 2008;47(4):506-24. doi:10.1080/02841860802059259.

32. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83.

33. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol*. 2011;11(1):83.

34. Pottegard A, Schmidt SA, Wallach-Kildemoes H, Sorensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol*. 2016. doi:10.1093/ije/dyw213.

35. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clin Epidemiol*. 2010;2:51.

36. Cronin-Fenton DP, Hellberg Y, Lauridsen KL, Ahern TP, Garne JP, Rosenberg C et al. Factors associated with concordant estrogen receptor expression at diagnosis and centralized re-assay in a Danish population-based breast cancer study. *Acta Oncol*. 2012;51(2):254-61. doi:10.3109/0284186x.2011.633556.

37. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol*. 2010;2:273.

38. Cronin-Fenton DP, Kjaersgaard A, Ahern TP, Mele M, Ewertz M, Hamilton-Dutoit S et al. Validity of Danish Breast Cancer Group (DBCG) registry data used in the predictors of breast cancer recurrence (ProBeCaRe) premenopausal breast cancer cohort study. *Acta Oncol*. 2017:1-6. doi:10.1080/0284186x.2017.1327720.

39. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-85. doi:10.1093/ije/dyu149.

40. Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA et al. CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. *J of Natl Cancer Inst*. 2011;103(6):489-500. doi:10.1093/jnci/djr010.

41. Ahern TP, Christensen M, Cronin-Fenton DP, Lunetta KL, Soiland H, Gjerde J et al. Functional polymorphisms in UDP-glucuronosyl transferases and recurrence in tamoxifen-treated breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20(9):1937-43. doi:10.1158/1055-9965.epi-11-0419.

42. Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA et al. Breast cancer recurrence risk related to concurrent use of SSRI antidepressants and tamoxifen. *Acta Oncol*. 2010;49(3):305-12. doi:10.3109/02841860903575273.

43. Baurley JW, Conti DV, Gauderman WJ, Thomas DC. Discovery of complex pathways from observational data. *Stat Med*. 2010;29(19):1998-2011.

44. Cronin-Fenton DP, Lash TL. Clinical epidemiology and pharmacology of CYP2D6 inhibition related to breast cancer outcomes. *Expert Rev Clin Pharmacol*. 2011;4(3):363-77. doi:10.1586/ecp.11.18.

45. Jensen AR, Storm HH, Moller S, Overgaard J. Validity and representativity in the Danish Breast Cancer Cooperative Group--a study on protocol allocation and data validity from one county to a multi-centre database. *Acta Oncol*. 2003;42(3):179-85.

46. Kraft P. Population stratification bias: more widespread than previously thought. *Epidemiology*. 2011;22(3):408-9. doi:10.1097/EDE.0b013e3182137e03.

47. Jaja C, Burke W, Thummel K, Edwards K, Veenstra DL. Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet*. 2008;11(3):141-9. doi:10.1159/000113876.

48. Cai WM, Nikoloff DM, Pan RM, de Leon J, Fanti P, Fairchild M et al. CYP2D6 genetic variation in healthy adults and psychiatric African-American subjects: implications for clinical practice and genetic testing. *Pharmacogenomics J*. 2006;6(5):343-50. doi:10.1038/sj.tpj.6500378.

49. Gaedigk A, Isidoro-Garcia M, Pearce RE, Sanchez S, Garcia-Solaesa V, Lorenzo-Romo C et al. Discovery of the nonfunctional CYP2D6 31 allele in Spanish, Puerto Rican, and US Hispanic populations. *Eur J Clin Pharmacol*. 2010;66(9):859-64. doi:10.1007/s00228-010-0831-4.

Figure Legends

Figure 1: Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped.

Figure 2 Selection of study sample and group based on the inclusion criteria. The source population consisted of 8,047 premenopausal women diagnosed with a first primary stage I-III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2011. After exclusions (n=2,088), the study population consists of 5,959 patients in the ProBe CaRe study.

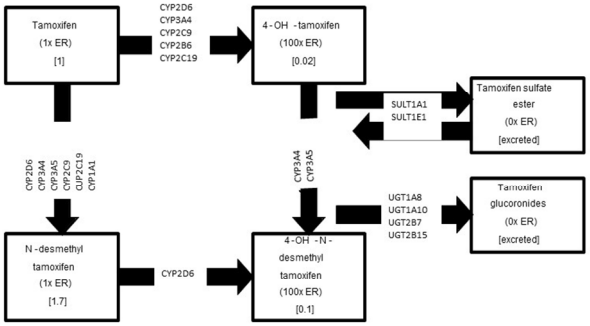


Figure 1: Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped.

108x60mm (300 x 300 DPI)

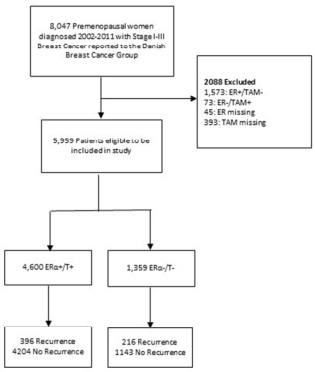


Figure 2 Selection of study sample and group based on the inclusion criteria. The source population consisted of 8,047 premenopausal women diagnosed with a first primary stage I-III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2011. After exclusions (n=2,088), the study population consists of 5,959 patients in the ProBe CaRe study.

108x60mm (300 x 300 DPI)

Electronic Supplementary Tables and Figures

Title: Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Author List: Lindsay J Collin, Deirdre P Cronin-Fenton, Thomas P Ahern, Peer M Christiansen, Per Damkier, Bent Ejlertsen, Stephen Hamilton-Dutoit, Anders Kjærsgaard, Rebecca A Silliman, Henrik Toft Sørensen, Timothy L Lash

To whom correspondence should be addressed: Lindsay J Collin, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE Atlanta GA 30322; Phone: 1-530-386-3341; Email: lindsay.jane.collin@emory.edu, ORCID: 0000-0003-0170-2592

Electronic Supplementary Table 1 Comparison of clinical and tumour characteristics by availability of tumour specimen for each ERα+/T+ and ERα–/T– group, among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study.

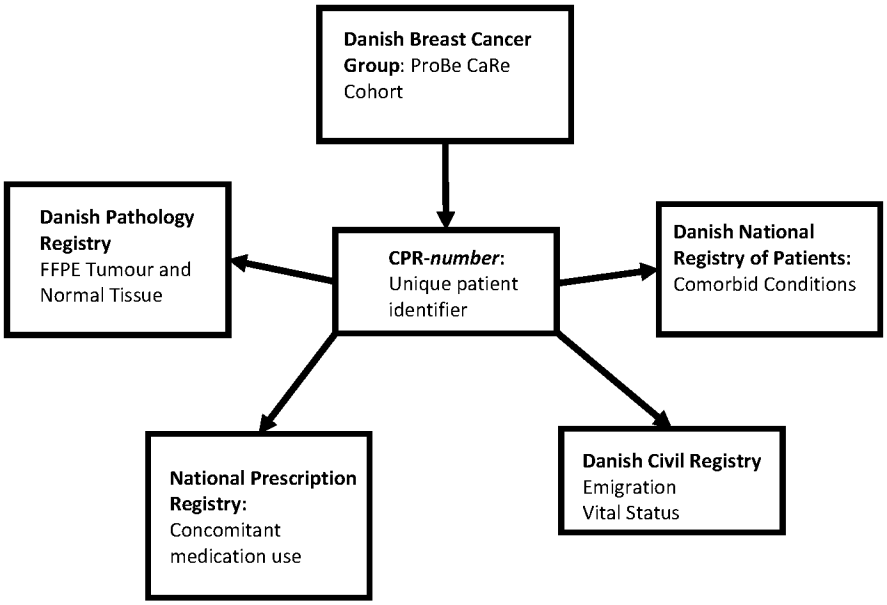
Patient and tumour characteristics	ER+/TAM+				ER–/TAM–			
	Tumour Specimen Available							
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
Total	3959	86	640	14.0	1139	84.0	220	16.0
Age at diagnosis								
<35	197	5	25	3.9	156	14	26	12
35-49	408	10	79	12	178	16	51	23
40-44	973	25	149	23	270	24	51	23
45-49	1455	37	213	33	324	28	61	28
50+	926	23	174	27	211	19	31	14
Stage at diagnosis								
Stage I	1020	26	164	26	334	29	68	31
Stage II	2125	54	350	55	590	52	112	51
Stage III	796	20	121	19	207	18	39	18
Unknown stage	18	0.5	5	0.8	8	0.7	1	0.5
Tumor size								
< 2mm	2267	57	379	59	553	49	124	56
2 - <5mm	1544	39	235	37	540	47	92	42
> 5mm	133	3.4	23	3.6	41	3.6	3	1.4
Unknown	15	0.4	3	0.5	5	0.4	1	0.5
Number of metastatic lymph nodes								
0	1469	37	235	37	582	51	113	51
1	982	25	166	26	194	17	44	20
2	493	13	90	14	101	8.3	15	6.8
3+	1004	25	147	23	259	23	47	21
Unknown	11	0.3	2	0.3	3	0.3	1	0.5
Lymph node evaluation								
No	7	0.2	1	0.2	2	0.2	1	0.5
Yes	3952	100	639	100	1137	99.8	219	99.5
Histologic grade								
Unsuitable	7	0.2	3	0.5	11	1	2	0.9
I	839	21	116	18	21	1.8	0	0
II	2024	51	366	57	182	16	34	15
III	822	21	128	20	737	65	147	67
Unknown	267	6.7	27	4.2	188	17	37	17
Type of primary surgery								
Mastectomy	1759	44	273	43	538	47	89	40

Lumpectomy	2200	56	367	57	601	53	131	60
Progesterone receptor status								
PR-	308	7.8	75	12	919	80.7	202	91.8
PR+	2245	57	435	68	17	1.5	2	0.9
Unknown/Not measured	1406	35	130	20	203	17.8	16	7.3
HER2 status								
HER2-	2489	63	397	62	579	51	113	51
HER2+	532	13	87	14	300	26	54	25
Unknown/Not measured	938	24	156	24	260	23	53	24
Intention to treat with chemotherapy								
No	126	3.2	18	3	11	1	3	1.4
Yes	3833	97	622	97	1128	99	217	99
Chemotherapy								
No	386	9.7	51	8	85	7.5	24	11
Yes	3573	90	589	92	1054	93	196	89
Intention to treat with Tamoxifen								
No	62	1.6	8	1.3	1131	99	220	100
Yes	3897	98	632	99	8	0.7	0	0
Radiation therapy								
No	565	14	90	14	230	20	37	17
Yes	3394	86	550	86	909	80	183	83
Anti-HER2 therapy								
No	2489	63	397	62	579	51	113	51
Yes	532	13	87	14	300	26	54	25
Unknown	938	24	156	24	260	23	53	24
Recurrence								
No	3625	92	579	90	962	84	181	82
Yes	334	8.4	61	9.5	177	16	39	18
Another malignancy								
No	3911	99	632	99	1126	99	215	98
Yes	48	1.2	8	1.3	13	1.1	5	2.3
Dead at end-of-follow-up								
No	3648	92	591	92	932	82	183	83
Yes	311	7.9	49	7.7	207	18	37	17
Charlson Comorbidity Score								
0	3949	100	637	100	1124	99	220	100
1	3	0.1	1	0.2	4	0.4	0	0
2	0	0	2	0.3	3	0.3	0	0
3+	7	0.2	0	0	8	0.7	0	0

Electronic Supplementary Table 2 Comparison of clinical and tumour characteristics by availability of non-neoplastic tissue specimen for each ERα+/T+ and ERα–/T– group, among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study

Patient and tumour characteristics	ER+/TAM+				ER–/TAM–			
	Non-neoplastic Tissue Sample Available							
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
Total	3746	92.0	853	19.0	1082	80.0	277	20.0
Age at diagnosis								
<35	188	5	34	4	157	15	25	9
35-49	392	11	95	11	161	15	68	25
40-44	916	24	206	24	267	25	54	20
45-49	1364	36	304	36	309	29	76	27
50+	886	24	214	25	188	17	54	20
Stage at diagnosis								
Stage I	966	26	218	26	317	29	85	31
Stage II	2005	54	470	55	560	52	142	51
Stage III	759	20	158	19	199	18	47	17
Unknown stage	16	0.4	7	0.8	6	0.6	3	1.1
Tumor size								
< 2mm	2150	57	496	58	522	48	155	56
2 - <5mm	1450	39	329	39	515	48	117	42
> 5mm	131	3.5	25	2.9	41	3.8	3	1.1
Unknown	15	0.4	3	0.4	4	0.4	2	0.7
Number of metastatic lymph nodes								
0	1385	37	319	37	556	51	139	50
1	929	25	219	26	177	16	61	22
2	482	13	101	12	97	9	19	6.9
3+	941	25	210	25	250	23	56	20
Unknown	9	0.2	4	0.5	2	0.2	2	0.7
Lymph node evaluation								
No	6	0.2	2	0.2	1	0.1	2	0.7
Yes	3740	100	851	100	1081	100	275	99
Histologic grade								
Unsuitable	5	0.1	5	0.6	10	0.9	3	1.1
I	791	21	164	19	17	1.6	4	1.4
II	1882	50	508	60	166	15	50	18
III	806	22	144	17	708	65	176	64
Unknown	262	7	32	3.8	181	17	44	16
Type of primary surgery								

Mastectomy	1696	45	336	39	524	48	103	37
Lumpectomy	2050	55	517	61	558	52	174	63
Progesterone receptor status								
PR-	281	7.5	102	12	860	80	261	94
PR+	2019	54	661	77	15	1.4	4	1.4
Unknown/Not measured	1446	39	90	11	207	19	12	4.3
HER2 status								
HER2-	2304	62	582	68	543	50	149	54
HER2+	525	14	94	11	282	26	72	26
Unknown/Not measured	917	25	177	21	257	24	56	20
Intention to treat with chemotherapy								
No	122	3.3	22	2.6	8	0.7	6	2.2
Yes	3624	97	831	97	1074	99	271	98
Chemotherapy								
No	381	10	56	6.6	87	8	22	8
Yes	3365	90	797	93	995	92	255	92
Intention to treat with Tamoxifen								
No	59	1.6	11	1.3	1075	99	276	100
Yes	3687	98	842	99	7	0.6	1	0.4
Radiation therapy								
No	547	15	108	13	218	20	49	18
Yes	3199	85	745	87	864	80	228	82
Anti-HER2 therapy								
No	2304	62	582	68	543	50	149	54
Yes	525	14	94	11	282	26	72	26
Unknown	917	25	177	21	257	24	56	20
Recurrence								
No	3433	92	771	90	918	85	225	81
Yes	313	8.4	82	9.6	164	15	52	19
Another malignancy								
No	3695	99	848	99	1069	99	272	98
Yes	51	1.4	5	0.6	13	1.2	5	1.8
Dead at end-of-follow-up								
No	3449	92	790	93	890	82	225	81
Yes	297	7.9	63	7.4	192	18	52	19
Charlson Comorbidity Score								
0	3733	100	853	100	1067	99	277	100
1	4	0.1	0	0	4	0.4	0	0
2	2	0.1	0	0	3	0.3	0	0
3+	7	0.2	0	0	8	0.7	0	0



BMJ Open

Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021805.R2
Article Type:	Cohort profile
Date Submitted by the Author:	28-Jun-2018
Complete List of Authors:	Collin, Lindsay; Emory University School of Public Health, Epidemiology Cronin-Fenton, Deirdre; Aarhus Universitetshospital, Ahern, Thomas; The Robert Larner M.D. College of Medicine at The University of Vermont Christiansen, P; Aarhus Universitet Damkier, Per; University of Southern Denmark, Dept. Clinical Biochemistry and Pharmacology Ejlertsen, Bent; Copenhagen University Hospital, Oncology Hamilton Dutoit, Stephen; Aarhus Universitetshospital, Institute of Pathology Kjærsgaard, Anders; Aarhus Universitetshospital, Department of Clinical Epidemiology Silliman, Rebecca; Boston University School of Medicine; Aarhus Universitetshospital, Department of Clinical Epidemiology Sørensen, Henrik T.; Aarhus University Hospital, Department of Clinical Epidemiology Lash, Timothy; Emory University Rollins School of Public Health, Department of Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Pharmacology and therapeutics
Keywords:	Cohort Study, Breast tumours < ONCOLOGY, Pharmacogenetics, EPIDEMIOLOGY

Title: Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Author List: Lindsay J Collin[a], Deirdre P Cronin-Fenton[b], Thomas P Ahern[c], Peer M Christiansen[d,e,f], Per Damkier[g,h], Bent Ejlertsen[f,i], Stephen Hamilton-Dutoit[j], Anders Kjærsgaard[b], Rebecca A Silliman[b,k], Henrik Toft Sørensen[b,l], Timothy L Lash[a,b]

[a]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; [b]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; [c]The Robert Larner, M.D. College of Medicine at The University of Vermont, Burlington, VT, USA; [d]Breast Unit, Surgical Department, Randers Regional Hospital, Randers, Denmark; [e]Department of Oncology, Odense University Hospital, Odense, Denmark; [f]Danish Breast Cancer Group, Copenhagen, Denmark; [g]Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark; [h]Department of Clinical Research, University of Southern Denmark, Denmark; [i]Rigshospitalet, Copenhagen, Denmark; [j]Institute of Pathology, Aarhus University Hospital, Aarhus, Denmark; [k]Boston University School of Medicine, Boston University, Boston, MA, USA; [l]Department of Health Research & Policy (Epidemiology), Stanford University, Stanford, CA, USA;

To whom correspondence should be addressed: Lindsay J Collin, Department of Epidemiology, 1518 Clifton Rd NE Atlanta GA 30322; Phone: 1-530-386-3341; Email: lindsay.jane.collin@emory.edu, ORCID: 0000-0003-0170-2592

Word Count: 3242

Keywords: Cohort study, breast cancer, pharmacogenetics, epidemiology

ABSTRACT

Purpose: The Predictors of Breast Cancer Recurrence (ProBe CaRe) study was established to evaluate modification of tamoxifen effectiveness in premenopausal women through reduced activity of tamoxifen-metabolizing enzymes. It comprehensively evaluates the effects of pharmacogenetic variants, use of concomitant medications, and biomarkers involved in oestrogen metabolism on breast cancer recurrence risk.

Participants: The ProBe CaRe study was established using resources from the Danish Breast Cancer Group (DBCG), including 5,959 premenopausal women diagnosed with stage I–III primary breast cancer between 2002 and 2010 in Denmark. Eligible participants were divided into two groups based on oestrogen receptor alpha (ERα) expression and receipt of tamoxifen therapy (TAM), 4,600 are classified as ERα+/TAM+ and 1,359 are classified as ERα–/TAM–. The ProBe CaRe study is a population-based cohort study nested in a nearly complete source population, clinical, tumour, and demographic data were abstracted from DBCG registry data. Linkage to Danish registries allows for abstraction of information regarding comorbid conditions, comedication use, and mortality. Formalin-fixed paraffin-embedded (FFPE) tissue samples have been prepared for DNA extraction and immunohistochemical assay.

Findings to date: To mitigate incorrect classification of patients into specific categories we conducted a validation substudy. We compared data acquired from registry and from medical record review to calculate positive and negative predictive values (PPV and NPV). We observed PPVs near 100% for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment. We found that the PPVs were

96% (95%CI 83, 100) for change in endocrine therapy, and 61% (95%CI 42, 77) for menopausal transition.

Future plans: The ProBe CaRe cohort study is well-positioned to comprehensively examine pharmacogenetic variants. We will use a Bayesian pathway analysis to evaluate the complete tamoxifen metabolic path to allow for gene-gene interactions, incorporating information of other important patient characteristics.

Strengths and Limitations of this Study

- One potential limitation of the ProBe CaRe study is the homogeneity of the study sample, as almost all are of European descent.
- In addition to being the first large epidemiologic study to examine reduced activity of tamoxifen metabolism in premenopausal women, this study is strengthened by completeness of high quality data.
- Our study includes a validation substudy to mitigate errors from incorrect classification of patients into specific categories of key analytic variables.

INTRODUCTION

Endocrine therapy improves survival in patients with breast cancer regardless of axillary lymph node status.[1] The Predictors of Breast Cancer Recurrence (ProBe CaRE) cohort study was established to evaluate modification of tamoxifen effectiveness in premenopausal women through reduced activity of tamoxifen-metabolizing enzymes. Candidates for adjuvant tamoxifen therapy include stage I–IV breast cancer patients with ER positive tumours, who constitute about two-thirds[2] of the approximately 1.7 million newly diagnosed breast cancer patients each year worldwide.[3] Current guidelines recommend that premenopausal patients with ER alpha positive (ERα+) cancers receive tamoxifen for five to ten years,[4-6] which reduces recurrence risk by nearly half,[6] and that tamoxifen may be offered to post-menopausal women with ERα+ cancers as an alternative to aromatase inhibitors. Tamoxifen metabolism is complex, but is principally catalysed by cytochrome P450 enzymes. Some metabolites bind with the ER with significantly greater affinity than tamoxifen itself, especially endoxifen, which has the highest ER-binding activity among tamoxifen metabolites. Activity of the enzymes involved in tamoxifen metabolism can vary between individuals due to inherited gene variants[7-11] or use of comedications.[8, 7] Although many studies have explored the association between these gene variants or use of comedications and failure of tamoxifen treatment,[12, 13] which manifests clinically as a recurrence, the interpretation of these studies remains controversial. Current clinical guidelines do not recommend genotyping these variant alleles to support treatment decisions,[1, 14, 15] but do recommend avoiding inhibiting comedications.[16]

To date, little available evidence on this topic is specific to premenopausal breast cancer patients. The competition between oestrogen and tamoxifen for ER binding is highly important for these patients because tamoxifen is a first-line guideline-recommended therapy[15, 1] for

premenopausal patients and because premenopausal women have higher concentrations of oestrogens to compete with tamoxifen for ER binding. Oestradiol, the most active oestrogen metabolite, binds with the ER with approximately the same affinity as endoxifen.[17] Premenopausal women have tenfold higher concentrations of oestradiol than postmenopausal women[18] and oestradiol concentrations tend to increase during tamoxifen therapy.[19, 18] This suggests that inhibition of tamoxifen-metabolizing enzymes is more likely to decrease effectiveness in premenopausal women, yet they have been seldom studied in this topic area.

Research questions

We established a premenopausal cohort of breast cancer patients to fill this important evidence gap, with the following primary study aims:

(1) Assess pharmacogenetics of tamoxifen metabolism and risk of breast cancer recurrence

We will assess the pharmacogenetics of tamoxifen metabolism by genotyping 32 variants in 15 enzymes (**Table 1**) thought to affect the concentration of the most active tamoxifen metabolites, and will evaluate the association between these variants and breast cancer recurrence in tamoxifen treated premenopausal breast cancer patients. Each of the selected enzymes is involved in at least one step in the tamoxifen metabolic pathway (**Figure 1**). Interactions with comedications that inhibit these metabolic enzymes also will be evaluated.

TABLE 1: Selected Functional variants and inhibitor comedications in genes whose enzymes metabolize tamoxifen

Gene	Number of selected functional variants	SNPs	Inhibitor comedications
CYP2D6	5	rs1065852, rs16947, rs3892097, rs28371706, rs28371725	bupropion, cinacalcet, fluoxetine, paroxetine, quinidine, duloxetine, sertraline, terbinafine, amiodarone, cimetidine, indinavir, nelfinavir, ritonavir, clarithromycin, itraconazole, ketoconazole, nefazodone, saquinavir, telithromycin, aprepitant, erythromycin, fluconazole, verapamil, diltiazem, cimetidine, voriconazole
CYP3A	1	rs10273424	
CYP3A5	1	rs776746	
CYP2C9	2	rs1057910, rs1799853	fluconazole, amiodarone, voriconazole
CYP2C19	2	rs12248560, rs4244285	
CYP2B6	2	rs3745274, rs8192709	
CYP1A1	1	rs1048943	
SULT1A1	3	rs1042157, rs1801030, rs9282861	
SULT1E1	2	rs3775775, rs3775778	
UGT2B7	1	rs7434332	
UGT2B10	1	rs294769	
ABCC2	3	rs3740065, rs717620, rs8187710	
ABCG2	3	rs1564481, rs2231164, rs2622604	
ABCB1	4	rs10248420, rs1045642, rs1128503, rs2032582	
UGT2B15	1	rs1902023	

(2) Assess the interaction between the pharmacogenetics of tamoxifen metabolism and oestrogen receptor beta (ERβ) expression

We will assess the effect of interaction between the pharmacogenetics of tamoxifen metabolism and ERβ expression on risk of breast cancer recurrence. Previous studies have shown that co-expression of ERβ is associated with improved survival among patients with ERα+ tumours who are treated with tamoxifen.[20, 21] The ERβ receptor opposes ERα-mediated proliferation.[20] Tumours that express both ERα and ERβ are less aggressive than tumours that express homodimers of ERα,[22, 20] due to the attenuated stimulation response from ERα/ERβ heterodimers. This suggests that metabolic inhibition may only affect ERβ– tumours. *In vitro* analyses have demonstrated that in ERα+/ERβ+ MCF7 cells, proliferation is inhibited by a wide range of endoxifen concentrations.[23] Still, in ERα+/ERβ– MCF7 cells, only physiologically high endoxifen concentrations inhibit proliferation,[23] indicating that metabolic inhibition affects risk of recurrence only when ERβ is absent.

(3) Assess interaction between inhibition of tamoxifen metabolism and oestrogen-regulating enzymes

Finally, we will assess the association between tumour expression of 17β-hydroxysteroid dehydrogenase 1 and 2 (17βHSD1 and 17βHSD2) and breast cancer recurrence. 17βHSD1 catalyses the conversion of oestrone (E1) to the most potent form of oestrogen, oestradiol (E2), and 17βHSD2 catalyses the reverse reaction.[24] E2 has the highest binding affinity for ER, and endoxifen acts through competitive inhibition at the receptor-binding site.[24] In breast tumour tissue, 17βHSD1 is more highly expressed than 17βHSD2. The opposite is usually observed in adjacent normal tissue.[25] Tumours with higher capacity to produce E2 endogenously through increased expression of the 17βHSD1 enzyme are more likely to overwhelm the tamoxifen

metabolites in competition for ER binding, affecting tamoxifen effectiveness. These enzymes are ideal therapeutic targets to modulate E2 concentrations in tumour cells, and candidate inhibitors have been developed.[26, 24] We will evaluate whether disequilibrium of the 17βHSD1 and -2 enzymes (ratio >1) results in compromised tamoxifen effectiveness.

COHORT DESCRIPTION

The ProBe CaRe cohort was established using the resources of the Danish Breast Cancer Group (DBCG) registry. The DBCG registry was established in 1976 and began to register patients in 1977, with the goals of standardizing treatment, facilitating clinical trials, and monitoring outcomes among Danish breast cancer patients.[27] Since its inception, the DBCG has registered over 90% of women diagnosed with breast cancer in Denmark. Breast cancer patients are registered in the DBCG via standardized forms. The registry has a standard protocol to collect information on tumour, treatment, and patient characteristics. Using this information-rich resource, the ProBe CaRe cohort is nested in a nearly complete source population of premenopausal women diagnosed with stage I–III first primary breast cancer between 2002–2010 whose breast cancer was reported to the DBCG. In Denmark, all citizens and legal residents are assigned a Civil Personal Register (CPR) number, a unique 10-digit personal identifier assigned at birth or upon immigration that is used for identification across all national registries (**Electronic Supplementary Figure 1**).[28]

Of the 8047 premenopausal women diagnosed with breast cancer between 2002 and 2010 and recorded in the DBCG registry, 5959 cancers were identified as eligible based on being a stage I–III first primary breast cancer and untreated with neoadjuvant therapy; all others

(n=2088) were excluded. The 5959 eligible patients then were divided into two cohorts based on ERα expression and receipt of tamoxifen therapy (**Figure 2**). To address competing explanations, (for instance if the biomarkers under study affect risk directly rather than mediating the tamoxifen effect), we will also evaluate the risk of recurrence in the subset of women with ERα– tumours who did not receive tamoxifen therapy (T–).

Our final ProBe CaRe study population consists of these 5,959 breast cancer patients divided into a cohort of ERα+/T+ (4600 patients) and a cohort of ERα–/T– (1359 patients). The socio-demographic and clinical characteristics of the two cohorts are described in **Table 2**. The distribution of the clinical and demographic characteristics between the two cohorts (ERα+/T+ vs. ERα–/T–) are relatively similar. They only differ meaningfully with respect to progesterone receptor (PR) status (58% vs. 1.4% PR+, respectively) and human epidermal growth factor receptor 2 (HER2) status (14% vs. 26% HER2+, respectively). With respect to outcomes, the ER–/TAM– cohort has a higher proportion of subjects who experienced recurrence (8.6% vs 16%, respectively) and who died by the end of follow up (7.8% vs. 18% respectively). This pattern is to be expected, as ER– breast cancers generally have a worse prognosis than ER+ cancers, especially within the first five years following diagnosis.[29, 30]

Table 2 Distribution of clinical and tumour characteristics by ER status and receipt of tamoxifen among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study.

Patient and tumour characteristics		**ER+/TAM+**		**ER–/TAM–**	
		N	%	N	%
Total		4600	100	1359	100
Age at diagnosis					
	<35	222	4.8	182	23
	35-39	487	11	229	27
	40-44	1123	24	321	24

45-49	1668	36	385	28
50+	1100	24	242	18
Menopausal status at diagnosis				
Premenopausal	4600	100	1359	100
Stage at diagnosis				
Stage I	1184	26	402	29.6
Stage II	2476	54	702	51.7
Stage III	917	20	246	18.1
Unknown stage	23	0.5	9	0.7
Tumor size				
< 2mm	2646	58	677	50
2 - <5mm	1780	39	632	47
> 5mm	156	3.4	44	3.2
Unknown	18	0.4	6	0.4
Number of metastatic lymph nodes				
0	1704	37	695	51
1	1148	25	238	17
2	583	13	116	9
3+	1152	25	306	23
Unknown	13	0.3	4	0.3
Lymph node evaluation				
No	8	0.2	3	0.2
Yes	4592	100	1356	100
Histologic grade				
Unsuitable	10	0.2	13	1
I	955	21	21	1.5
II	2391	52	216	16
III	950	21	884	65
Unknown	294	6.4	225	17
Type of primary surgery				
Mastectomy	2033	44	627	46
Lumpectomy	2567	56	732	54
Progesterone receptor status				
PR-	383	8.3	1121	83
PR+	2680	58	19	1.4
Unknown/Not measured	1537	33	219	16
HER2 status				
HER2-	2887.00	63	692	51
HER2+	619	14	354	26

Unknown/Not measured	1094	24	313	23
Intention to treat with chemotherapy				
No	144	3	14	1
Yes	4456	97	1345	99
Chemotherapy				
No	437	9	109	8
Yes	4163	91	1250	92
Intention to treat with Tamoxifen				
No	70	1.5	1351	99
Yes	4530	98	8	0.6
Radiation therapy				
No	655	14	267	20
Yes	3945	86	1092	80
Anti-HER2 therapy				
No	2887	63	692	51
Yes	619	13	354	26
Unknown	1094	24	313	23
Recurrence				
No	4204	91	1143	84
Yes	396	8.6	216	16
Another malignancy				
No	4544	99	1341	99
Yes	56	1.2	18	1.3
Dead at end-of-follow-up				
No	4239	92	1115	82
Yes	361	8	244	18
Charlson Comorbidity Score				
0	4587	99	1344	99
1	4	0.1	4	0.3
2	2	0	3	0.2
3+	7	0.2	8	0.6

Cohort follow-up

Women diagnosed with breast cancer and subsequently enrolled in the DBCG registry undergo semi-annual examinations during the first five years after diagnosis and annual examinations during years six to ten.[31] Women undergoing treatment for breast cancer receive endocrine therapy through the Danish government and obtain their medicine at the hospital, which will be used to estimate tamoxifen adherence. Members of both the ERα+/T+ and the

ERα–/T– ProBe CaRe cohorts have been followed from breast cancer diagnosis to the first of (a) recurrence, (b) death, (c) ten years of follow-up, (d) loss to follow-up due to emigration, (e) another malignancy, or (f) the end of the study follow-up period. Breast cancer recurrence was identified using the DBCG registry. We adopted the DBCG definition of breast cancer recurrence as any type of breast cancer diagnosed subsequent to the initial course of therapy.[31] Recurrences are then further categorized as loco-regional (in the scar or regional lymph nodes), contralateral (opposite breast), distant (all other sites), or unknown (site of recurrence not documented). The date of recurrence is recorded in the DBCG registry, including recurrences diagnosed between scheduled follow-up exams. Mortality and emigration were identified using the Danish Civil Registration System, which is updated daily.[28] Emigration is the only expected source of loss to follow-up and has impacted less than 1% of the study population.

Data Collection

Registry Data

Once participants eligible for ProBe CaRe were selected from the DBCG registry, we extracted clinical and demographic information from the DBCG registry. This information included date and place of diagnosis, tumour characteristics, treatment received, and patient characteristics, which are presented in **Electronic Supplementary Table 1**. We also extracted information on comorbid diseases at time of breast cancer diagnosis, summarized using the Charlson Comorbidity Index.[32, 33] The registry information allows us to update the participants' conditions during study follow-up and therefore to account for time-varying exposures and confounding factors. The CPR number for each patient was used to link cohort members to the Danish National Prescription Registry,[34] which provided information on filled

prescriptions of drugs that inhibit tamoxifen-metabolizing enzymes. This allowed us to assess the drug-drug interactions discussed above.

Biobank

The CPR number for each patient was used to identify the hospital at which the surgery was performed and to locate and retrieve the corresponding formalin-fixed paraffin-embedded (FFPE) tissue samples. The list of ProBe CaRe cohort members and their CPR numbers and hospitals of diagnosis were provided to a medical research technician at the Institute of Pathology, blinded to whether the CPR numbers corresponded to a patient with a recurrence. The technician reviewed a description of the available tissue blocks (routinely available in the Danish pathology registry[35]), identified the tumour-rich and non-neoplastic blocks for each patient, and specified which FFPE blocks should be requested from the hospitals. This list of blocks to be requested was then returned to the Department of Clinical Epidemiology, which prepared and mailed the request letters to the pathology archives at the respective Danish hospitals. Staff at the hospital pathology archives returned the blocks to the Department of Clinical Epidemiology, which assigned a project identification number to the block and then provided it to the Institute of Pathology. The project identification number maintained blinding of laboratory personnel to whether blocks corresponded to patients with a recurrence. The Department of Clinical Epidemiology maintains the key linking the project identification number for the blocks to the clinical data, including recurrence status.

Non-neoplastic tissue samples are taken routinely from normal adjacent tissue or cancer-free lymph nodes resected during breast cancer surgery and were used as controls in creation of tissue microarrays (TMAs). Of the 4,600 patients included in the ERα+/T+ cohort, 4,599 patients had samples evaluated by clinicians, and tumour samples were available for 3,959 (86%).

Among the ER–/TAM– cohort, 1139 (84%) patients had tumour samples available. Distribution of clinical and demographic characteristics among patients with, and without, available tumour samples are described in **Table 3**. Of patients included in the ERα+/T+ cohort, 2746 (82%) had a non-neoplastic tissue sample available, while 1082 (80%) patients in the ER–/TAM– cohort had non-neoplastic tissue samples available. Distributions of demographic and clinical characteristics among patients with and without non-neoplastic tissue samples are summarized in **Electronic Supplementary Table 2**.

Table 3 Summary of exposure, covariate, and outcome variables collected in the ProBe CaRe study

Exposures	**Outcome**
Genetic variants	Recurrence
ERα and ERβ	Mortality
17βHSD1 and 17βHSD2	
Biomarkers	
Clinical	**Demographic**
Tumour characteristics	Age
Treatment therapy	Region
Comorbidity	Hospital of Diagnosis
Medication History and Use	

Sections of collected tissue blocks have been prepared for DNA extraction and immunohistochemical (IHC) assay. In accordance with the study's primary aims, we will genotype 32 variants across 15 genes known to be involved in tamoxifen metabolism, in order to predict extent of metabolic inhibition. We will also re-assay ERα expression to ensure correct classification of the two cohorts, as original ERα expression was measured in different pathology laboratories using different methods. In our previous case-control study of post-menopausal

patients diagnosed during 1985–2001, we reported 95% concordance of positive ERα expression between initial assays and reassays and 74% concordance of negative ERα expression between initial assays and reassays.[36] ERβ expression will be assayed using IHC in TMAs to assess its possible modification of tamoxifen metabolic inhibition. Expression of the enzymes 17βHSD1 and 17βHSD2 also will be assayed using IHC to address the study aim examining whether the ratio of these two enzymes modulates tamoxifen's efficacy in preventing breast cancer recurrence. The aforementioned assays of biomarkers are the primary starting point. However, we anticipate that the study will yield a substantial tumour biobank and ultimately provide a valuable resource to researchers for further characterization of prognostic and predictive biomarkers in premenopausal breast cancer.

Validation substudy

Registry data are not error-free.[37] To mitigate incorrect classification of patients into specific categories, we conducted a validation substudy.[38] By comparing data procured both from the registry and from medical record review, we calculated positive and negative predictive values (PPV and NPV, respectively) and their corresponding confidence intervals for key analytic variables. We observed near perfect PPVs for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy, and tamoxifen treatment. The PPVs were 96% (95% CI 83, 100) for change in endocrine therapy and 61% (95%CI 42, 77) for menopausal transition. While the PPV for DBCG-recorded recurrence was 100%, there were more recurrences reported in the medical records than reported in the DBCG database.[38] These parameters will allow us to adjust for measurement errors in our analyses, improving data quality and confidence in the resulting measures of association.[39]

Patient and Public Involvement

Patients and public were not involved in the development of this study.

FINDINGS TO DATE

In our preceding ProBe CaRe nested case-control study, where 94% of breast cancer patients were postmenopausal, we compared rates of breast cancer recurrence for women with a polymorphism that impairs the function of CYP2D6 (an enzyme involved in tamoxifen metabolism) to those in women without this polymorphism and found a null association (adjusted OR = 0.99; 95% CI = 0.76, 1.3).[40] We further evaluated functional variants in the phase II UDP-glucuronosyl transferases, which contributes to deactivation tamoxifen, and again found near null associations.[41] We have assessed drug-drug interactions with concomitant use of selective serotonin reuptake inhibitor antidepressants using the Danish National Prescription Registry, and reported an adjusted odds ratio of 1.1 (95% CI = 0.7, 1.7).[42]

The current ProBe CaRe longitudinal cohort will build on our previous research to address gene-gene and gene-drug interactions that may compromise tamoxifen effectiveness by focusing on premenopausal women and by more comprehensively evaluating variants in the metabolic path. We will use a Bayesian pathway analysis, (the Algorithm for Learning Pathway Analysis (ALPS)),[43] to evaluate the complete tamoxifen metabolic pathway and to allow for identification of gene-gene interactions, while also estimating the net effect of the entire pathway.[44] This analytic approach will allow for incorporation of time-varying information on tamoxifen adherence, use of inhibiting co-medications, comorbidity, and transition to postmenopausal status, while modelling complex gene-gene interactions without issues of sparse

data or a reduction in power.[43] ALPS also permits incorporation of prior biologic knowledge regarding the metabolic path of tamoxifen, so that the search space for the algorithm is constrained to pathways consistent with currently understood biology.

The DBCG has a long history of contributions to the scientific community, informing clinical and treatment guidelines for breast cancer.[45, 27, 31] It is thus an indispensable resource for addressing our study aims.

STRENGTHS AND LIMITATIONS

The current ProBe CaRe study is a large prospective cohort nested within a nearly complete source population. The cohort has many strengths, including the completeness of high quality data and a large representative study population from the Danish source population. Our study design allows for thorough assessment of competing explanations for our findings, both through inclusion of a cohort of ER–/TAM– participants and an internal validation study to address possible errors in classification of key variables. It is the first cohort to examine reduced activity of tamoxifen metabolism in premenopausal women with ample sample size. Moreover, all data (except for new laboratory data) were collected from standardized reports submitted to population-based prospective registries. In addition to DCBG data, we can link patient records to drug prescription, morbidity, and mortality data from independently maintained registries to ensure that relevant covariates are considered.

One potential limitation of the ProBe CaRe cohort is the homogeneity of the study sample, as almost all patients are of European descent. However, there is no comparable source

with the same level of information quality to allow exploration of our aims in a more diverse study population. Lack of diversity is a potential limitation, but previous studies indicate that our findings may be extrapolated to external populations and can inform the future direction of research in more diverse populations.[46-49]

COLLABORATIONS

ProBe CaRe study data are held and managed by the Department of Clinical Epidemiology (KEA) in Aarhus, Denmark. We welcome collaborations to enhance the utility of the data and biobank and will respond to all inquiries (tlash@emory.edu).

FURTHER DETAILS

Contributors: LC prepared the original draft of the manuscript. AK conducted data analyses and put together the tables. TL, DCF, HTS, SHD, and TA were responsible for study development and planning. DCF and HTS were responsible for application for data access in Denmark. SHD led the collection and preparation of the tumour samples for genotyping and immunohistochemistry assays. PD provided methodological input surrounding the pharmacogenetic aspects of the study. PC, BE, and RS provided methodologic insight into study design, operationalization of the study aims, and clinical insights. All authors provided critical review of the manuscript and approved the final version.

Funding The ProBe CaRe cohort study was established with funding from the National Cancer Institute at the U.S. National Institutes of Health (R01 CA166825; PI: Lash).

Competing Interests: The authors declare no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The research is approved by the regional ethical board in Denmark and by the Institutional Review Boards in the US. The study does not contain any animal experiments performed by any of the authors.

Data sharing statement Once the initial data analyses are complete, we will be open to collaborations with outside investigators as permitted by the IRBs of participating sites. In particular, we will encourage collaborations with researchers whose expertise is under-represented on our research team. To become a collaborator, a researcher will be required to submit an application, which will undergo both a scientific and IRB review. In view of the complexity of the database and requirements of Danish Law, interested investigators will be asked to form collaborative arrangement with the ProBe CaRe investigators rather than sharing data directly.

Informed Consent Informed consent was not required. Registry-based research is exempt from informed consent requirements according to Danish law.

REFERENCES

1. Burstein HJ, Temin S, Anderson H, Buchholz TA, Davidson NE, Gelmon KE 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):2255-69.

2. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians* . 2016;66(1):31-42.

3. Stewart B, Wild CP. World cancer report 2014. 2014.

4. Burstein HJ, Griggs JJ, Prestrud AA, Temin S. American society of clinical oncology clinical practice guideline update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J of Oncol Practice* . 2010;6(5):243-6. doi:10.1200/jop.000082.

5. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ et al. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol*. 2011;22(8):1736-47. doi:10.1093/annonc/mdr304.

6. 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. doi:10.1016/s0140-6736(05)66544-0.

7. Stearns V, Johnson MD, Rae JM, Morocho A, Novielli A, Bhargava P et al. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J of Natl Cancer Inst* . 2003;95(23):1758-64.

8. Jin Y, Desta Z, Stearns V, Ward B, Ho H, Lee KH et al. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J of Natl Cancer Inst* . 2005;97(1):30-9. doi:10.1093/jnci/dji005.

9. Gjerde J, Hauglid M, Breilid H, Lundgren S, Varhaug JE, Kisanga ER et al. Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. *Ann Oncol* : official journal of the European Society for Medical Oncology. 2008;19(1):56-61. doi:10.1093/annonc/mdm434.

10. Lim JS, Chen XA, Singh O, Yap YS, Ng RC, Wong NS et al. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP2C19 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *Br J Clin Pharmacol* . 2011;71(5):737-50. doi:10.1111/j.1365-2125.2011.03905.x.

11. Murdter TE, Schroth W, Bacchus-Gerybadze L, Winter S, Heinkele G, Simon W 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. *Clin Pharmacol Ther*. 2011;89(5):708-17. doi:10.1038/clpt.2011.27.

12. Cronin-Fenton DP, Damkier P, Lash TL. Metabolism and transport of tamoxifen in relation to its effectiveness: new perspectives on an ongoing controversy. *Future Oncol* . 2014;10(1):107-22. doi:10.2217/fon.13.168.

13. Province MA, Goetz MP, Brauch H, Flockhart DA, Hebert JM, Whaley R et al. CYP2D6 genotype and adjuvant tamoxifen: meta-analysis of heterogeneous study populations. *Clin Pharmacol Ther*. 2014;95(2):216-27. doi:10.1038/clpt.2013.186.

14. Network NCC. NCCN practice guidelines in oncology—breast cancer v. 2.2008. 2009.

15. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* : official journal of the European Society for Medical Oncology. 2011;22(8):1736-47. doi:10.1093/annonc/mdr304.

16. Sideras K, Ingle JN, Ames MM, Loprinzi CL, Mrazek DP, Black JL et al. Coprescription of tamoxifen and medications that inhibit CYP2D6. *J Clin Oncol* . 2010;28(16):2768-76.

17. Johnson MD, Zuo H, Lee K-H, Trebley JP, Rae JM, Weatherman RV et al. Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res Treat*. 2004;85(2):151-9.

18. Lash TL, Lien EA, Sørensen HT, Hamilton-Dutoit S. Genotype-guided tamoxifen therapy: time to pause for reflection? *Lancet Oncol* . 2009;10(8):825-33.

19. Sherman BM, Chapler FK, Crickard K, Wycoff D. Endocrine consequences of continuous antiestrogen therapy with tamoxifen in premenopausal women. *J Clin Investig* . 1979;64(2):398.

20. Lindberg K, Helguero LA, Omoto Y, Gustafsson JA, Haldosen LA. Estrogen receptor beta represses Akt signaling in breast cancer cells via downregulation of HER2/HER3 and upregulation of PTEN: implications for tamoxifen sensitivity. *Breast Cancer Res* : BCR. 2011;13(2):R43. doi:10.1186/bcr2865.

21. Murphy LC, Watson PH. Is oestrogen receptor-beta a predictor of endocrine therapy responsiveness in human breast cancer? *Endocr-Relat Cancer*. 2006;13(2):327-34. doi:10.1677/erc.1.01141.

22. Fox EM, Davis RJ, Shupnik MA. ERbeta in breast cancer--onlooker, passive player, or active protector? *Steroids*. 2008;73(11):1039-51. doi:10.1016/j.steroids.2008.04.006.

23. Wu X, Subramaniam M, Grygo SB, Sun Z, Negron V, Lingle WL et al. Estrogen receptor-beta sensitizes breast cancer cells to the anti-estrogenic actions of endoxifen. *Breast Cancer Res: BCR* . 2011;13(2):R27. doi:10.1186/bcr2844.

24. Marchais-Oberwinkler S, Henn C, Moller G, Klein T, Negri M, Oster A et al. 17beta-Hydroxysteroid dehydrogenases (17beta-HSDs) as therapeutic targets: protein structures, functions, and recent progress in inhibitor development. *J Steroid Biochem Mol Biol*. 2011;125(1-2):66-82. doi:10.1016/j.jsbmb.2010.12.013.

25. Speirs V, Green AR, Atkin SL. Activity and gene expression of 17beta-hydroxysteroid dehydrogenase type I in primary cultures of epithelial and stromal cells derived from normal and tumourous human breast tissue: the role of IL-8. *J Steroid Biochem Mol Biol*. 1998;67(3):267-74.

26. Poirier D. Contribution to the development of inhibitors of 17beta-hydroxysteroid dehydrogenase types 1 and 7: key tools for studying and treating estrogen-dependent diseases. *J Steroid Biochem Mol Biol*. . 2011;125(1-2):83-94. doi:10.1016/j.jsbmb.2010.12.007.

27. Blichert-Toft M, Christiansen P, Mouridsen HT. Danish Breast Cancer Cooperative Group--DBCG: History, organization, and status of scientific achievements at 30-year anniversary. *Acta Oncol*. 2008;47(4):497-505. doi:10.1080/02841860802068615.

28. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541-9.

29. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492-502.

30. Blows FM, Driver KE, Schmidt MK, Broeks A, Van Leeuwen FE, Wesseling J et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med*. 2010;7(5):e1000279.

31. Moller S, Jensen MB, Ejlertsen B, Bjerre KD, Larsen M, Hansen HB et al. The clinical database and the treatment guidelines of the Danish Breast Cancer Cooperative Group (DBCG); its 30-years experience and future promise. *Acta Oncol*. 2008;47(4):506-24. doi:10.1080/02841860802059259.

32. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83.

33. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol*. 2011;11(1):83.

34. Pottegard A, Schmidt SA, Wallach-Kildemoes H, Sorensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol*. 2016. doi:10.1093/ije/dyw213.

35. Erichsen R, Lash TL, Hamilton-Dutoit SJ, Bjerregaard B, Vyberg M, Pedersen L. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clin Epidemiol*. 2010;2:51.

36. Cronin-Fenton DP, Hellberg Y, Lauridsen KL, Ahern TP, Garne JP, Rosenberg C et al. Factors associated with concordant estrogen receptor expression at diagnosis and centralized re-assay in a Danish population-based breast cancer study. *Acta Oncol*. 2012;51(2):254-61. doi:10.3109/0284186x.2011.633556.

37. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol*. 2010;2:273.

38. Cronin-Fenton DP, Kjaersgaard A, Ahern TP, Mele M, Ewertz M, Hamilton-Dutoit S et al. Validity of Danish Breast Cancer Group (DBCG) registry data used in the predictors of breast cancer recurrence (ProBeCaRe) premenopausal breast cancer cohort study. *Acta Oncol*. 2017:1-6. doi:10.1080/0284186x.2017.1327720.

39. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-85. doi:10.1093/ije/dyu149.

40. Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA et al. CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. *J of Natl Cancer Inst*. 2011;103(6):489-500. doi:10.1093/jnci/djr010.

41. Ahern TP, Christensen M, Cronin-Fenton DP, Lunetta KL, Soiland H, Gjerde J et al. Functional polymorphisms in UDP-glucuronosyl transferases and recurrence in tamoxifen-treated breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20(9):1937-43. doi:10.1158/1055-9965.epi-11-0419.

42. Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA et al. Breast cancer recurrence risk related to concurrent use of SSRI antidepressants and tamoxifen. *Acta Oncol*. 2010;49(3):305-12. doi:10.3109/02841860903575273.

43. Baurley JW, Conti DV, Gauderman WJ, Thomas DC. Discovery of complex pathways from observational data. *Stat Med*. 2010;29(19):1998-2011.

44. Cronin-Fenton DP, Lash TL. Clinical epidemiology and pharmacology of CYP2D6 inhibition related to breast cancer outcomes. *Expert Rev Clin Pharmacol*. 2011;4(3):363-77. doi:10.1586/ecp.11.18.

45. Jensen AR, Storm HH, Moller S, Overgaard J. Validity and representativity in the Danish Breast Cancer Cooperative Group--a study on protocol allocation and data validity from one county to a multi-centre database. *Acta Oncol.* 2003;42(3):179-85.

46. Kraft P. Population stratification bias: more widespread than previously thought. *Epidemiology*. 2011;22(3):408-9. doi:10.1097/EDE.0b013e3182137e03.

47. Jaja C, Burke W, Thummel K, Edwards K, Veenstra DL. Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet*. 2008;11(3):141-9. doi:10.1159/000113876.

48. Cai WM, Nikoloff DM, Pan RM, de Leon J, Fanti P, Fairchild M et al. CYP2D6 genetic variation in healthy adults and psychiatric African-American subjects: implications for clinical practice and genetic testing. *Pharmacogenomics J.* 2006;6(5):343-50. doi:10.1038/sj.tpj.6500378.

49. Gaedigk A, Isidoro-Garcia M, Pearce RE, Sanchez S, Garcia-Solaesa V, Lorenzo-Romo C et al. Discovery of the nonfunctional CYP2D6 31 allele in Spanish, Puerto Rican, and US Hispanic populations. *Eur J Clin Pharmacol*. 2010;66(9):859-64. doi:10.1007/s00228-010-0831-4.

Figure Legends

Figure 1: Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped.

Figure 2 Selection of study sample and group based on the inclusion criteria. The source population consisted of 8,047 premenopausal women diagnosed with a first primary stage I-III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2011. After exclusions (n=2,088), the study population consists of 5,959 patients in the ProBe CaRe study.

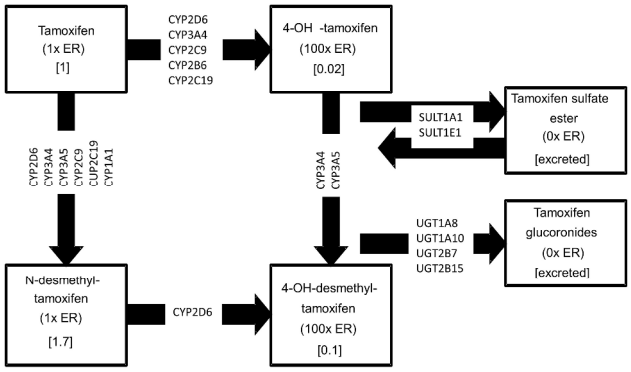


Figure 1: Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped.

338x190mm (300 x 300 DPI)

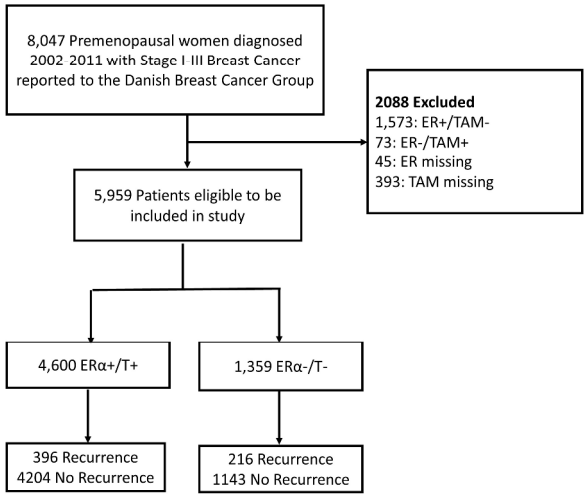


Figure 2: Selection of study sample and group based on the inclusion criteria. The source population consisted of 8,047 premenopausal women diagnosed with a first primary stage I–III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2011. After exclusions (n=2,088), the study population consists of 5,959 patients in the ProBe CaRe study.

338x190mm (300 x 300 DPI)

Electronic Supplementary Tables and Figures

Title: Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark

Author List: Lindsay J Collin, Deirdre P Cronin-Fenton, Thomas P Ahern, Peer M Christiansen, Per Damkier, Bent Ejlertsen, Stephen Hamilton-Dutoit, Anders Kjærsgaard, Rebecca A Silliman, Henrik Toft Sørensen, Timothy L Lash

To whom correspondence should be addressed: Lindsay J Collin, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE Atlanta GA 30322; Phone: 1-530-386-3341; Email: lindsay.jane.collin@emory.edu, ORCID: 0000-0003-0170-2592

Electronic Supplementary Table 1 Comparison of clinical and tumour characteristics by availability of tumour specimen for each ERα+/T+ and ERα–/T– group, among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study.

Patient and tumour characteristics	ER+/TAM+				ER–/TAM–			
	Tumour Specimen Available							
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
Total	3959	86	640	14.0	1139	84.0	220	16.0
Age at diagnosis								
<35	197	5	25	3.9	156	14	26	12
35-49	408	10	79	12	178	16	51	23
40-44	973	25	149	23	270	24	51	23
45-49	1455	37	213	33	324	28	61	28
50+	926	23	174	27	211	19	31	14
Stage at diagnosis								
Stage I	1020	26	164	26	334	29	68	31
Stage II	2125	54	350	55	590	52	112	51
Stage III	796	20	121	19	207	18	39	18
Unknown stage	18	0.5	5	0.8	8	0.7	1	0.5
Tumor size								
< 2mm	2267	57	379	59	553	49	124	56
2 - <5mm	1544	39	235	37	540	47	92	42
> 5mm	133	3.4	23	3.6	41	3.6	3	1.4
Unknown	15	0.4	3	0.5	5	0.4	1	0.5
Number of metastatic lymph nodes								
0	1469	37	235	37	582	51	113	51
1	982	25	166	26	194	17	44	20
2	493	13	90	14	101	8.3	15	6.8
3+	1004	25	147	23	259	23	47	21
Unknown	11	0.3	2	0.3	3	0.3	1	0.5
Lymph node evaluation								
No	7	0.2	1	0.2	2	0.2	1	0.5
Yes	3952	100	639	100	1137	99.8	219	99.5
Histologic grade								
Unsuitable	7	0.2	3	0.5	11	1	2	0.9
I	839	21	116	18	21	1.8	0	0
II	2024	51	366	57	182	16	34	15
III	822	21	128	20	737	65	147	67
Unknown	267	6.7	27	4.2	188	17	37	17
Type of primary surgery								
Mastectomy	1759	44	273	43	538	47	89	40

Lumpectomy	2200	56	367	57	601	53	131	60
Progesterone receptor status								
PR-	308	7.8	75	12	919	80.7	202	91.8
PR+	2245	57	435	68	17	1.5	2	0.9
Unknown/Not measured	1406	35	130	20	203	17.8	16	7.3
HER2 status								
HER2-	2489	63	397	62	579	51	113	51
HER2+	532	13	87	14	300	26	54	25
Unknown/Not measured	938	24	156	24	260	23	53	24
Intention to treat with chemotherapy								
No	126	3.2	18	3	11	1	3	1.4
Yes	3833	97	622	97	1128	99	217	99
Chemotherapy								
No	386	9.7	51	8	85	7.5	24	11
Yes	3573	90	589	92	1054	93	196	89
Intention to treat with Tamoxifen								
No	62	1.6	8	1.3	1131	99	220	100
Yes	3897	98	632	99	8	0.7	0	0
Radiation therapy								
No	565	14	90	14	230	20	37	17
Yes	3394	86	550	86	909	80	183	83
Anti-HER2 therapy								
No	2489	63	397	62	579	51	113	51
Yes	532	13	87	14	300	26	54	25
Unknown	938	24	156	24	260	23	53	24
Recurrence								
No	3625	92	579	90	962	84	181	82
Yes	334	8.4	61	9.5	177	16	39	18
Another malignancy								
No	3911	99	632	99	1126	99	215	98
Yes	48	1.2	8	1.3	13	1.1	5	2.3
Dead at end-of-follow-up								
No	3648	92	591	92	932	82	183	83
Yes	311	7.9	49	7.7	207	18	37	17
Charlson Comorbidity Score								
0	3949	100	637	100	1124	99	220	100
1	3	0.1	1	0.2	4	0.4	0	0
2	0	0	2	0.3	3	0.3	0	0
3+	7	0.2	0	0	8	0.7	0	0

Electronic Supplementary Table 2 Comparison of clinical and tumour characteristics by availability of non-neoplastic tissue specimen for each ERα+/T+ and ERα–/T– group, among the 5,959 participants in a population-based cohort of premenopausal women diagnosed with first primary breast cancer, ProBe CaRe study

Patient and tumour characteristics	ER+/TAM+				ER–/TAM–			
	Non-neoplastic Tissue Sample Available							
	Yes		No		Yes		No	
	N	%	N	%	N	%	N	%
Total	3746	92.0	853	19.0	1082	80.0	277	20.0
Age at diagnosis								
<35	188	5	34	4	157	15	25	9
35-49	392	11	95	11	161	15	68	25
40-44	916	24	206	24	267	25	54	20
45-49	1364	36	304	36	309	29	76	27
50+	886	24	214	25	188	17	54	20
Stage at diagnosis								
Stage I	966	26	218	26	317	29	85	31
Stage II	2005	54	470	55	560	52	142	51
Stage III	759	20	158	19	199	18	47	17
Unknown stage	16	0.4	7	0.8	6	0.6	3	1.1
Tumor size								
< 2mm	2150	57	496	58	522	48	155	56
2 - <5mm	1450	39	329	39	515	48	117	42
> 5mm	131	3.5	25	2.9	41	3.8	3	1.1
Unknown	15	0.4	3	0.4	4	0.4	2	0.7
Number of metastatic lymph nodes								
0	1385	37	319	37	556	51	139	50
1	929	25	219	26	177	16	61	22
2	482	13	101	12	97	9	19	6.9
3+	941	25	210	25	250	23	56	20
Unknown	9	0.2	4	0.5	2	0.2	2	0.7
Lymph node evaluation								
No	6	0.2	2	0.2	1	0.1	2	0.7
Yes	3740	100	851	100	1081	100	275	99
Histologic grade								
Unsuitable	5	0.1	5	0.6	10	0.9	3	1.1
I	791	21	164	19	17	1.6	4	1.4
II	1882	50	508	60	166	15	50	18
III	806	22	144	17	708	65	176	64
Unknown	262	7	32	3.8	181	17	44	16
Type of primary surgery								

Mastectomy	1696	45	336	39	524	48	103	37
Lumpectomy	2050	55	517	61	558	52	174	63
Progesterone receptor status								
PR-	281	7.5	102	12	860	80	261	94
PR+	2019	54	661	77	15	1.4	4	1.4
Unknown/Not measured	1446	39	90	11	207	19	12	4.3
HER2 status								
HER2-	2304	62	582	68	543	50	149	54
HER2+	525	14	94	11	282	26	72	26
Unknown/Not measured	917	25	177	21	257	24	56	20
Intention to treat with chemotherapy								
No	122	3.3	22	2.6	8	0.7	6	2.2
Yes	3624	97	831	97	1074	99	271	98
Chemotherapy								
No	381	10	56	6.6	87	8	22	8
Yes	3365	90	797	93	995	92	255	92
Intention to treat with Tamoxifen								
No	59	1.6	11	1.3	1075	99	276	100
Yes	3687	98	842	99	7	0.6	1	0.4
Radiation therapy								
No	547	15	108	13	218	20	49	18
Yes	3199	85	745	87	864	80	228	82
Anti-HER2 therapy								
No	2304	62	582	68	543	50	149	54
Yes	525	14	94	11	282	26	72	26
Unknown	917	25	177	21	257	24	56	20
Recurrence								
No	3433	92	771	90	918	85	225	81
Yes	313	8.4	82	9.6	164	15	52	19
Another malignancy								
No	3695	99	848	99	1069	99	272	98
Yes	51	1.4	5	0.6	13	1.2	5	1.8
Dead at end-of-follow-up								
No	3449	92	790	93	890	82	225	81
Yes	297	7.9	63	7.4	192	18	52	19
Charlson Comorbidity Score								
0	3733	100	853	100	1067	99	277	100
1	4	0.1	0	0	4	0.4	0	0
2	2	0.1	0	0	3	0.3	0	0
3+	7	0.2	0	0	8	0.7	0	0