

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

Lindsay J Collin,¹ Deirdre P Cronin-Fenton,² Thomas P Ahern,³ Peer M Christiansen,^{4,5,6} Per Damkier,^{7,8} Bent Ejlersen,^{6,9} Stephen Hamilton-Dutoit,¹⁰ Anders Kjærsgaard,² Rebecca A Silliman,^{2,11} Henrik Toft Sørensen,^{2,12} Timothy L Lash^{1,2}

To cite: Collin LJ, Cronin-Fenton DP, Ahern TP, *et al*. Cohort Profile: the Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark. *BMJ Open* 2018;**8**:e021805. doi:10.1136/bmjopen-2018-021805

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-021805>).

Received 19 January 2018
Revised 28 June 2018
Accepted 3 July 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Lindsay J Collin;
lindsay.jane.collin@emory.edu

ABSTRACT

Purpose The Predictors of Breast Cancer Recurrence (ProBe CaRE) study was established to evaluate modification of tamoxifen (TAM) effectiveness in premenopausal women through reduced activity of TAM-metabolising 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 5959 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 (ER α) expression and receipt of TAM therapy, 4600 are classified as ER α +/TAM+ and 1359 are classified as ER α -/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 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 predictive values (PPVs) and negative predictive values. We observed PPVs near 100% for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and TAM treatment. We found that the PPVs were 96% (95% CI 83% to 100%) for change in endocrine therapy and 61% (95% CI 42% to 77%) for menopausal transition.

Future plans The ProBeCaRE cohort study is well positioned to comprehensively examine pharmacogenetic variants. We will use a Bayesian pathway analysis to evaluate the complete TAM 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 Predictors of Breast Cancer Recurrence study is the homogeneity of the study sample, as almost all are of European descent.
- In addition to being the first large epidemiological 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 the incorrect classification of patients into specific categories of key analytical 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 (TAM) effectiveness in premenopausal women through reduced activity of TAM-metabolising enzymes. Candidates for adjuvant TAM therapy include patients with stage I–IV breast cancer with oestrogen receptor (ER) positive tumours, who constitute about two-thirds² of the approximately 1.7 million newly diagnosed patients with breast cancer each year worldwide.³ Current guidelines recommend that premenopausal patients with ER alpha positive (ER α +) cancers receive TAM for 5–10 years,^{4–6} which reduces recurrence risk by nearly half,⁶ and that TAM may be offered to postmenopausal women with ER α + cancers as an alternative to aromatase inhibitors. TAM metabolism is complex, but is principally catalysed by cytochrome P450 enzymes. Some metabolites bind with the ER with significantly greater affinity than TAM itself, especially endoxifen, which has the highest

ER-binding activity among TAM metabolites. Activity of the enzymes involved in TAM metabolism can vary between individuals due to inherited gene variants^{7–11} or use of comedications.^{7,8} Although many studies have explored the association between these gene variants or use of comedications and failure of TAM 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,^{1,5,14} but do recommend avoiding inhibiting comedications.¹⁵

To date, little available evidence on this topic is specific to premenopausal patients with breast cancer. The competition between oestrogen and TAM for ER binding is highly important for these patients because TAM is a first-line guideline-recommended therapy¹⁵ for premenopausal patients and because premenopausal women have higher concentrations of oestrogens to compete with TAM for ER binding. Oestradiol (E2), the most active oestrogen metabolite, binds with the ER with approximately the same affinity as endoxifen.¹⁶ Premenopausal women have 10-fold higher concentrations of E2 than postmenopausal women¹⁷ and E2 concentrations tend to increase during TAM therapy.^{17,18} This suggests that inhibition of TAM-metabolising 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 patients with breast cancer to fill this important evidence gap, with the following primary study aims:

Assess pharmacogenetics of TAM metabolism and risk of breast cancer recurrence

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

Assess the interaction between the pharmacogenetics of TAM metabolism and ER beta (ERβ) expression

We will assess the effect of interaction between the pharmacogenetics of TAM metabolism and ERβ expression on risk of breast cancer recurrence. Previous studies have shown that coexpression of ERβ is associated with improved survival among patients with ERα+ tumours who are treated with TAM.^{19,20} The ERβ receptor opposes ERα-mediated proliferation.¹⁹ Tumours that express

Table 1 Selected functional variants and inhibitor comedications in genes whose enzymes metabolise tamoxifen

Gene	No 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	

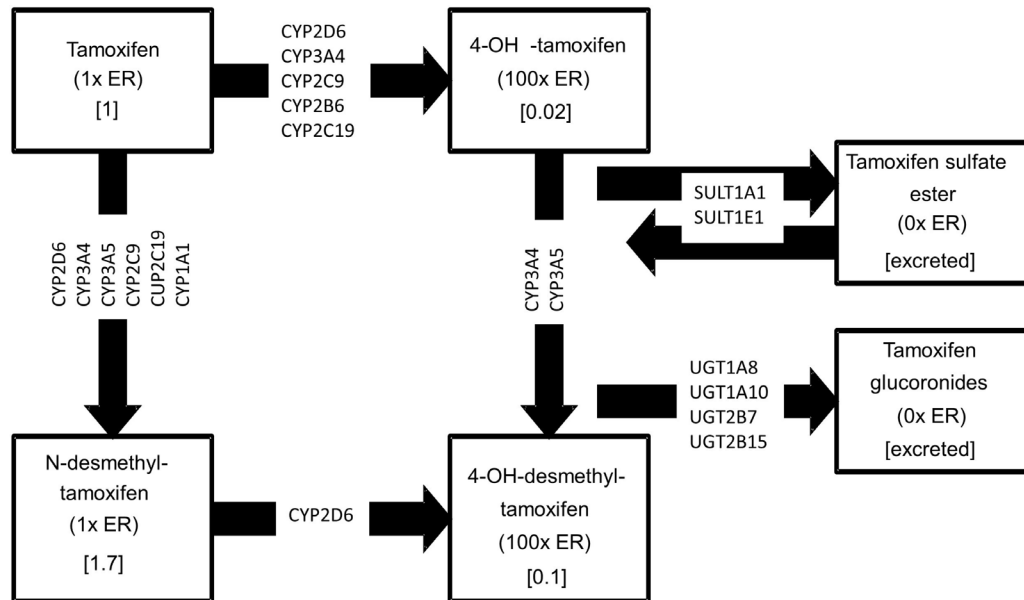


Figure 1 Metabolic pathway of tamoxifen and related metabolites including enzymes that have been genotyped. ER, oestrogen receptor.

both ER α and ER β are less aggressive than tumours that express homodimers of ER α ,^{19 21} due to the attenuated stimulation response from ER α /ER β heterodimers. This suggests that metabolic inhibition may only affect ER β -tumours. In vitro, analyses have demonstrated that in ER α +/ER β +MCF7 cells, proliferation is inhibited by a wide range of endoxifen concentrations.²² Still, in ER α +/ER β - MCF7 cells, only physiologically high endoxifen concentrations inhibit proliferation,²² indicating that metabolic inhibition affects risk of recurrence only when ER β is absent.

Assess interaction between inhibition of TAM metabolism and oestrogen-regulating enzymes

Finally, we will assess the association between tumour expression of 17 β -hydroxysteroid dehydrogenase 1 and 2 (17 β HSD1 and 17 β HSD2) and breast cancer recurrence. 17 β HSD1 catalyses the conversion of oestrone to the most potent form of oestrogen, E2 and 17 β HSD2 catalyses the reverse reaction.²³ E2 has the highest binding affinity for ER, and endoxifen acts through competitive inhibition at the receptor-binding site.²³ In breast tumour tissue, 17 β HSD1 is more highly expressed than 17 β HSD2. The opposite is usually observed in adjacent normal tissue.²⁴ Tumours with higher capacity to produce E2 endogenously through increased expression of the 17 β HSD1 enzyme are more likely to overwhelm the TAM metabolites in competition for ER binding, affecting TAM effectiveness. These enzymes are ideal therapeutic targets to modulate E2 concentrations in tumour cells, and candidate inhibitors have been developed.^{23 25} We will evaluate whether disequilibrium of the 17 β HSD1 and -2 enzymes (ratio >1) results in compromised TAM 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 standardising treatment, facilitating clinical trials and monitoring outcomes among Danish patients with breast cancer.²⁶ Since its inception, the DBCG has registered over 90% of women diagnosed with breast cancer in Denmark. Patients with breast cancer are registered in the DBCG via standardised 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 and 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 on immigration that is used for identification across all national registries electronic online supplementary figure 1.²⁷

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 ER α expression and receipt of TAM therapy (figure 2). To address competing explanations (eg, if the biomarkers under study affect risk directly rather than mediating the TAM effect), we will also evaluate the risk of recurrence in the

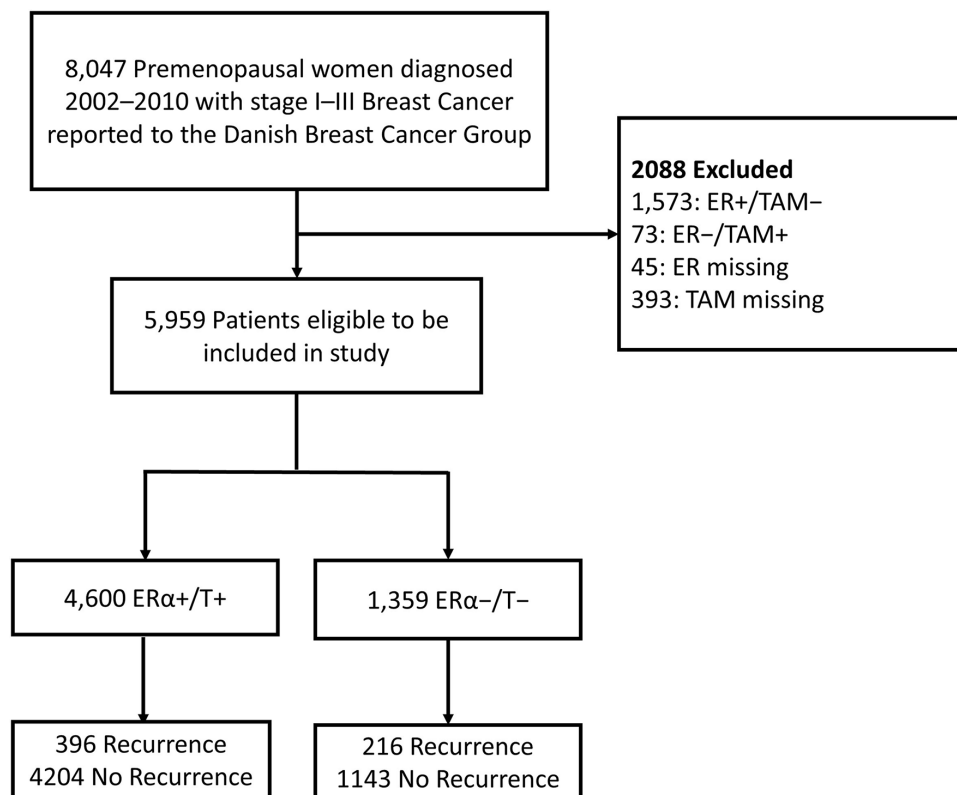


Figure 2 Selection of study sample and group based on the inclusion criteria. The source population consisted of 8047 premenopausal women diagnosed with a first primary stage I–III breast cancer and reported to the Danish Breast Cancer Group between 2002 and 2010. After exclusions (n=2088), the study population consists of 5959 patients in the ProBe CaRe study. ER, oestrogen receptor; ER α , oestrogen receptor alpha; ProBe CaRe, Predictors of Breast Cancer Recurrence; TAM, tamoxifen therapy.

subset of women with ER α - tumours who did not receive TAM therapy (T-).

Our final ProBe CaRe study population consists of these 5959 patients with breast cancer divided into a cohort of ER α +/T+ (4600 patients) and a cohort of ER α -/T- (1359 patients). The sociodemographic and clinical characteristics of the two cohorts are described in [table 2](#). The distribution of the clinical and demographic characteristics between the two cohorts (ER α +/T+ vs ER α -/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 5 years following diagnosis.^{28 29}

Cohort follow-up

Women diagnosed with breast cancer and subsequently enrolled in the DBCG registry undergo semiannual examinations during the first 5 years after diagnosis and annual examinations during years 6–10.³⁰ 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 TAM adherence. Members of both the ER α +/T+ and the ER α -/T- ProBe CaRe cohorts have been followed from breast cancer diagnosis to the first of (1) recurrence, (2) death, (3) 10 years of follow-up, (4) loss to follow-up due to emigration, (5) another malignancy or (6) 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.³⁰ Recurrences are then further categorised as locoregional (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.²⁷ 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

Table 2 Distribution of clinical and tumour characteristics by ER status and receipt of tamoxifen among the 5959 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
Tumour 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
No 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
Histological 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

Continued

Table 2 Continued

Patient and tumour characteristics	ER+/TAM+		ER-/TAM-	
	N	%	N	%
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

ER, oestrogen receptor; HER, human epidermal growth factor receptor 2; ProBe CaRe, Predictors of Breast Cancer Recurrence; PR, progesterone receptor; TAM, tamoxifen therapy.

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 online supplementary table 1. We also extracted information on comorbid diseases at time of breast cancer diagnosis, summarised using the Charlson Comorbidity Index.^{31 32} 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,³³ which provided information on filled prescriptions of drugs that inhibit TAM-metabolising 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 registry³⁴), 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 4600 patients included in the ER α +/T+ cohort, 4599 patients had samples evaluated by clinicians, and tumour samples were available for 3959 (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 ER α +/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 summarised in electronic online supplementary table 2.

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

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

Exposures	Outcome
Genetic variants	Recurrence
ER α and ER β	Mortality
17 β HSD1 and 17 β HSD2	
Biomarkers	
Clinical	Demographic
Tumour characteristics	Age
Treatment therapy	Region
Comorbidity	Hospital of diagnosis
Medication history and use	

ER α , oestrogen receptor alpha; ER β , oestrogen receptor beta; 17 β -hydroxysteroid dehydrogenase 1 and 2; ProBe CaRe, Predictors of Breast Cancer Recurrence.

genotype 32 variants across 15 genes known to be involved in TAM metabolism, in order to predict extent of metabolic inhibition. We will also reassay ER α expression to ensure correct classification of the two cohorts, as original ER α expression was measured in different pathology laboratories using different methods. In our previous case–control study of postmenopausal patients diagnosed during 1985–2001, we reported 95% concordance of positive ER α expression between initial assays and re-assays and 74% concordance of negative ER α expression between initial assays and re-assays.³⁵ ER β expression will be assayed using IHC in TMAs to assess its possible modification of TAM metabolic inhibition. Expression of the enzymes 17 β HSD1 and 17 β HSD2 also will be assayed using IHC to address the study aim examining whether the ratio of these two enzymes modulates TAM'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 characterisation of prognostic and predictive biomarkers in premenopausal breast cancer.

Validation substudy

Registry data are not error-free.³⁶ To mitigate incorrect classification of patients into specific categories, we conducted a validation substudy.³⁷ By comparing data procured both from the registry and from medical record review, we calculated positive predictive values (PPVs) and negative predictive values and their corresponding CIs for key analytical variables. We observed near perfect PPVs for tumour size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and TAM treatment. The PPVs were 96% (95% CI 83% to 100%) for change in endocrine therapy and 61% (95% CI 42% to 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.³⁷ These parameters will allow us to adjust for measurement errors in our analyses, improving data quality and confidence in the resulting measures of association.³⁸

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% patients with breast cancer were postmenopausal, we compared rates of breast cancer recurrence for women with a polymorphism that impairs the function of CYP2D6 (an enzyme involved in TAM metabolism) to those in women without this polymorphism and found a null association (adjusted OR 0.99; 95% CI 0.76 to 1.3).³⁹ We further evaluated functional variants in the phase II UDP-glucuronosyl transferases, which contributes to deactivation TAM, and again found near null associations.⁴⁰ 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 OR of 1.1 (95% CI 0.7 to 1.7).⁴¹

The current ProBe CaRe longitudinal cohort will build on our previous research to address gene–gene and gene–drug interactions that may compromise TAM 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 Structure (ALPS)),⁴² to evaluate the complete TAM metabolic pathway and to allow for identification of gene–gene interactions, while also estimating the net effect of the entire pathway.⁴³ This analytical approach will allow for incorporation of time-varying information on TAM adherence, use of inhibiting comedications, comorbidity and transition to postmenopausal status, while modelling complex gene–gene interactions without issues of sparse data or a reduction in power.⁴² ALPS also permits incorporation of prior biological knowledge regarding the metabolic path of TAM, 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.^{26 30 44} 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 TAM metabolism in premenopausal women with ample sample size. Moreover, all data (except for new laboratory data) were collected from standardised 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.^{45–48}

COLLABORATIONS

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

Author affiliations

¹Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

²Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark

³Department of Surgery, The Robert Larner, M.D. College of Medicine at The University of Vermont, Burlington, Vermont, USA

⁴Breast Unit, Surgical Department, Randers Regional Hospital, Randers, Denmark

⁵Department of Oncology, Odense University Hospital, Odense, Denmark

⁶Danish Breast Cancer Group, Copenhagen, Denmark

⁷Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark

⁸Department of Clinical Research, University of Southern Denmark, Odense, Denmark

⁹Rigshospitalet, Copenhagen, Denmark

¹⁰Institute of Pathology, Aarhus University Hospital, Aarhus, Denmark

¹¹Boston University School of Medicine, Boston University, Boston, Massachusetts, USA

¹²Department of Health Research & Policy (Epidemiology), Stanford University, Stanford, California, USA

Contributors Contributors: LJC prepared the original draft of the manuscript. AK conducted data analyses and put together the tables. TLL, DPC-F, HTS, SH-D and TPA were responsible for study development and planning. DPC-F and HTS were responsible for application for data access in Denmark. SH-D 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. PMC, BE and RAS provided methodological insight into study design, operationalisation 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 US National Institutes of Health (R01 CA166825; PI: Lash).

Competing interests None declared.

Patient consent Not required.

Ethics 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 Declaration of Helsinki 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 USA. The study does not contain any animal experiments performed by any of the authors.

Provenance and peer review Not commissioned; externally peer reviewed.

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.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Burstein HJ, Temin S, Anderson H, *et al*. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: american society of clinical oncology clinical practice guideline focused update. *J Clin Oncol* 2014;32:2255–69.
- DeSantis CE, Fedewa SA, Goding Sauer A, *et al*. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA Cancer J Clin* 2016;66:31–42.
- Stewart B, Wild CP. *World cancer report 2014*, 2014.
- Burstein HJ, Griggs JJ, Prestrud AA, *et al*. American society of clinical oncology clinical practice guideline update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J Oncol Pract* 2010;6:243–6.
- Goldhirsch A, Wood WC, Coates AS, *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;22:1736–47.
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;365:1687–717.
- Stearns V, Johnson MD, Rae JM, *et al*. Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J Natl Cancer Inst* 2003;95:1758–64.
- Jin Y, Desta Z, Stearns V, *et al*. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J Natl Cancer Inst* 2005;97:30–9.
- Gjerde J, Hauglid M, Breilid H, *et al*. Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. *Ann Oncol* 2008;19:56–61.
- Lim JS, Chen XA, Singh O, *et al*. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP2C19 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *Br J Clin Pharmacol* 2011;71:737–50.
- Mürdter TE, Schroth W, Bacchus-Gerybadze L, *et al*. Activity levels of tamoxifen metabolites at the estrogen receptor and the impact of genetic polymorphisms of phase I and II enzymes on their concentration levels in plasma. *Clin Pharmacol Ther* 2011;89:708–17.
- 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;10:107–22.
- Province MA, Goetz MP, Brauch H, *et al*. CYP2D6 genotype and adjuvant tamoxifen: meta-analysis of heterogeneous study populations. *Clin Pharmacol Ther* 2014;95:216–27.
- Network NCC. *NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis*, 2009.
- Sideras K, Ingle JN, Ames MM, *et al*. Coprescription of tamoxifen and medications that inhibit CYP2D6. *J Clin Oncol* 2010;28:2768–76.
- Johnson MD, Zuo H, Lee KH, *et al*. Pharmacological characterization of 4-hydroxy-N-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res Treat* 2004;85:151–9.
- Lash TL, Lien EA, Sørensen HT, *et al*. Genotype-guided tamoxifen therapy: time to pause for reflection? *Lancet Oncol* 2009;10:825–33.
- Sherman BM, Chapler FK, Crickard K, *et al*. Endocrine consequences of continuous antiestrogen therapy with tamoxifen in premenopausal women. *J Clin Invest* 1979;64:398–404.
- Lindberg K, Helguero LA, Omoto Y, *et al*. Estrogen receptor β represses Akt signaling in breast cancer cells via downregulation of HER2/HER3 and upregulation of PTEN: implications for tamoxifen sensitivity. *Breast Cancer Res* 2011;13:R43.
- Murphy LC, Watson PH. Is oestrogen receptor-beta a predictor of endocrine therapy responsiveness in human breast cancer? *Endocr Relat Cancer* 2006;13:327–34.
- Fox EM, Davis RJ, Shupnik MA. ERbeta in breast cancer—looker, passive player, or active protector? *Steroids* 2008;73:1039–51.
- Wu X, Subramaniam M, Grygo SB, *et al*. Estrogen receptor-beta sensitizes breast cancer cells to the anti-estrogenic actions of endoxifen. *Breast Cancer Res* 2011;13:R27.
- Marchais-Oberwinkler S, Henn C, Möller G, *et al*. 17 β -Hydroxysteroid dehydrogenases (17 β -HSDs) as therapeutic targets: protein structures, functions, and recent progress in inhibitor development. *J Steroid Biochem Mol Biol* 2011;125:66–82.
- Speirs V, Green AR, Atkin SL. Activity and gene expression of 17 β -hydroxysteroid dehydrogenase type I in primary cultures of epithelial and stromal cells derived from normal and tumorous human breast tissue: the role of IL-8. *J Steroid Biochem Mol Biol* 1998;67:267–74.
- Poirier D. Contribution to the development of inhibitors of 17 β -hydroxysteroid dehydrogenase types 1 and 7: key tools for studying and treating estrogen-dependent diseases. *J Steroid Biochem Mol Biol* 2011;125:83–94.
- 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;47:497–505.
- Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–9.
- Carey LA, Perou CM, Livasy CA, *et al*. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006;295:2492–502.
- Blows FM, Driver KE, Schmidt MK, *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:e1000279.
- Møller S, Jensen MB, Ejlersen B, *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;47:506–24.
- Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
- Thygesen SK, Christiansen CF, Christensen S, *et al*. 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;11:83.
- Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, *et al*. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2017;46:dyw213.
- Erichsen R, Lash TL, Hamilton-Dutoit SJ, *et al*. Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank. *Clin Epidemiol* 2010;2:51.
- Cronin-Fenton DP, Hellberg Y, Lauridsen KL, *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;51:254–61.
- Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. *Clin Epidemiol* 2010;2:273.
- Cronin-Fenton DP, Kjærsgaard A, Ahern TP, *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;56:1155–60.
- Lash TL, Fox MP, MacLehose RF, *et al*. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;43:1969–85.
- Lash TL, Cronin-Fenton D, Ahern TP, *et al*. CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. *J Natl Cancer Inst* 2011;103:489–500.
- Ahern TP, Christensen M, Cronin-Fenton DP, *et al*. Functional polymorphisms in UDP-glucuronosyl transferases and recurrence

- in tamoxifen-treated breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* 2011;20:1937–43.
41. Lash TL, Cronin-Fenton D, Ahern TP, *et al.* Breast cancer recurrence risk related to concurrent use of SSRI antidepressants and tamoxifen. *Acta Oncol* 2010;49:305–12.
 42. Baurley JW, Conti DV, Gauderman WJ, *et al.* Discovery of complex pathways from observational data. *Stat Med* 2010;29:1998–2011.
 43. Cronin-Fenton DP, Lash TL. Clinical epidemiology and pharmacology of CYP2D6 inhibition related to breast cancer outcomes. *Expert Rev Clin Pharmacol* 2011;4:363–77.
 44. Jensen AR, Storm HH, Møller S, *et al.* 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;42:179–85.
 45. Kraft P. Population stratification bias: more widespread than previously thought. *Epidemiology* 2011;22:408–9.
 46. Jaja C, Burke W, Thummel K, *et al.* Cytochrome p450 enzyme polymorphism frequency in indigenous and native american populations: a systematic review. *Community Genet* 2008;11:141–9.
 47. Cai WM, Nikoloff DM, Pan RM, *et al.* CYP2D6 genetic variation in healthy adults and psychiatric African-American subjects: implications for clinical practice and genetic testing. *Pharmacogenomics J* 2006;6:343–50.
 48. Gaedigk A, Isidoro-García M, Pearce RE, *et al.* Discovery of the nonfunctional CYP2D6 31 allele in Spanish, Puerto Rican, and US Hispanic populations. *Eur J Clin Pharmacol* 2010;66:859–64.