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 Kjærgaard,2 Rebecca A Silliman,2,11 Henrik Toft Sørensen,2,12 Timothy L Lash1,2

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 (ERα) expression and receipt of TAM therapy, 4600 are classified as ERα+/TAM+ and 1359 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 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.

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 (ERα+) cancers receive TAM for 5–10 years,⁴⁵ which reduces recurrence risk by nearly half,⁶ and that TAM may be offered to postmenopausal women with ERα+ 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 or use of comedications. Although many studies have explored the association between these gene variants or use of comedications and failure of TAM treatment, 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, but do recommend avoiding inhibiting comedications.

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 therapy 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. Premenopausal women have 10-fold higher concentrations of E2 than postmenopausal women and E2 concentrations tend to increase during TAM therapy. 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. The ERβ receptor opposes ERα-mediated proliferation. 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                                  | rs10238420, rs1045642, rs1128503, rs2032582 |                                                             |
| UGT2B15| 1                                  | rs1902023                                  |                                                             |
both ERα and ERβ are less aggressive than tumours that express homodimers of ERα.19 21 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.22 Still, in ERα+/ ERβ− MCF7 cells, only physiologically high endoxifen concentrations inhibit proliferation,22 indicating that metabolic inhibition affects risk of recurrence only when ERβ is absent.

Assess interaction between inhibition of TAM 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 to the most potent form of oestrogen, E2 and 17βHSD2 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, 17βHSD1 is more highly expressed than 17βHSD2. The opposite is usually observed in adjacent normal tissue.24 Tumours with higher capacity to produce E2 endogenously through increased expression of the 17βHSD1 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 17βHSD1 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 ERα 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
subset of women with ERα− 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 ERα+/T+ (4600 patients) and a cohort of ERα−/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 (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 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 ERα+/T+ and the ERα−/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
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

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- |
|------------------------------------|---------|---------|
| Total                              | 4600    | 1359    |
| **Age at diagnosis**               |         |         |
| <35                                | 222     | 182     |
| 35–39                              | 487     | 229     |
| 40–44                              | 1123    | 321     |
| 45–49                              | 1668    | 385     |
| 50+                                | 1100    | 242     |
| **Menopausal status at diagnosis** |         |         |
| Premenopausal                      | 4600    | 1359    |
| **Stage at diagnosis**             |         |         |
| Stage I                            | 1184    | 402     |
| Stage II                           | 2476    | 702     |
| Stage III                          | 917     | 246     |
| Unknown stage                      | 23      | 9       |
| **Tumour size**                    |         |         |
| <2 mm                              | 2646    | 677     |
| 2≤5 mm                             | 1780    | 632     |
| >5 mm                              | 156     | 44      |
| Unknown                            | 18      | 6       |
| **No of metastatic lymph nodes**   |         |         |
| 0                                  | 1704    | 695     |
| 1                                  | 1148    | 238     |
| 2                                  | 583     | 116     |
| 3+                                 | 1152    | 306     |
| Unknown                            | 13      | 4       |
| **Lymph node evaluation**          |         |         |
| No                                 | 8       | 3       |
| Yes                                | 4592    | 1356    |
| **Histological grade**             |         |         |
| Unsuitable                         | 10      | 13      |
| I                                  | 955     | 21      |
| II                                 | 2391    | 52      |
| III                                | 950     | 884     |
| Unknown                            | 294     | 225     |
| **Type of primary surgery**        |         |         |
| Mastectomy                         | 2033    | 627     |
| Lumpectomy                         | 2567    | 732     |
| **Progesterone receptor status**   |         |         |
| PR–                                | 383     | 1121    |
| PR+                                | 2680    | 19      |
| Unknown/not measured               | 1537    | 219     |

Table 2 Continued

| Patient and tumour characteristics | ER+/TAM+ | ER-/TAM- |
|------------------------------------|---------|---------|
| **HER2 status**                    |         |         |
| HER2−                              | 2887.00 | 692     |
| HER2+                              | 619     | 354     |
| Unknown/not measured               | 1094    | 313     |
| **Intention to treat with chemotherapy** |         |         |
| No                                 | 144     | 14      |
| Yes                                | 4456    | 1345    |
| **Chemotherapy**                   |         |         |
| No                                 | 437     | 109     |
| Yes                                | 4163    | 1250    |
| **Intention to treat with tamoxifen** |         |         |
| No                                 | 70      | 1351    |
| Yes                                | 4530    | 8       |
| **Radiation therapy**              |         |         |
| No                                 | 655     | 267     |
| Yes                                | 3945    | 1092    |
| **Anti-HER2 therapy**              |         |         |
| No                                 | 2887    | 692     |
| Yes                                | 619     | 354     |
| Unknown                            | 1094    | 313     |
| **Recurrence**                     |         |         |
| No                                 | 4204    | 1143    |
| Yes                                | 396     | 216     |
| **Another malignancy**             |         |         |
| No                                 | 4544    | 1341    |
| Yes                                | 56      | 18      |
| **Dead at end of follow-up**       |         |         |
| No                                 | 4239    | 1115    |
| Yes                                | 361     | 244     |
| **Charlson Comorbidity Score**     |         |         |
| 0                                  | 4587    | 1344    |
| 1                                  | 4       | 4       |
| 2                                  | 2       | 3       |
| 3+                                 | 7       | 8       |

ER, oestrogen receptor; HER, human epidermal growth factor receptor 2; ProBe CaRe, Predictors of Breast Cancer Recurrence; PR, progesterone receptor; TAM, tamoxifen therapy.
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, 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 registry), 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+/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 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 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 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 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 postmenopausal 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. ERβ expression will be assayed using IHC in TMAs to assess its possible modification of TAM 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 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. To mitigate incorrect classification of patients into specific categories, we conducted a validation substudy. 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

| 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 | |
| ERα, oestrogen receptor alpha; ERβ, oestrogen receptor beta; 17β-hydroxysteroid dehydrogenase 1 and 2; ProBe CaRe, Predictors of Breast Cancer Recurrence. |
medical records than reported in the DBCG database. These parameters will allow us to adjust for measurement errors in our analyses, improving data quality and confidence in the resulting measures of association.

**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). We further evaluated functional variants in the phase II UDP-glucuronosyl transferases, which contributes to deactivation TAM, and again found near null associations. 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).

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)), 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. 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 modeling complex gene–gene interactions without issues of sparse data or a reduction in power. 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.

**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 standardized reports submitted to population-based prospective registries. In addition to DBCG 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.

**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
9Department of Pathology, Aarhus University Hospital, Aarhus, Denmark
10Boston University School of Medicine, Boston University, Boston, Massachusetts, USA
11Department 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 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 ProBo Café 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
1. Burstein HJ, Temin S, Anderson H, et al. Adjunctive 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.
2. 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.
3. Stewart S, Wild CP, et al. Breast cancer incidence and mortality in Denmark, 2004–2008. Br J Cancer 2010;103:698–706.
4. 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.
5. Goldhirsh 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.
6. 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;365:1867–71.
7. 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.
8. 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.
9. Gjerde J, Hauglid M, Bredl H, et al. Effects of CYP2D6 and SULT1A1 genotypes including SULT1A1 gene copy number on tamoxifen metabolism. Ann Oncol 2008;19:56–61.
10. Lim JS, Chen VA, Singh O, et al. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP1A2 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. Br J Clin Pharmacol 2011;71:737–50.
11. Mürder TE, Schrott 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.
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;10:1099–112.
13. 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.
14. Network NCC. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis, 2009.
15. Sideras K, Ingle JN, Ames MM, et al. Coprescription of tamoxifen and medications that inhibit CYP2D6. J Clin Oncol 2010;28:2768–76.
16. 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.
17. Lash TL, Lien EA, Sørensen HT, et al. Genotype-guided tamoxifen therapy: time to pause for reflection? Lancet Oncol 2009;10:825–33.
18. 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.
19. Lindberg K, Helguero LA, Omoyo 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.
20. Murphy EC, Watson PH. Is oestrogen receptor-beta a predictor of endocrine therapy responsiveness in human breast cancer? Endocr Relat Cancer 2006;13:327–34.
21. Fox EM, Davis RJ, Shupnik MA. ERbeta in breast cancer—onlooker, passive player, or active protector? Steroids 2008;73:1039–51.
22. 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 Treat 2011;13:27.
23. Marchais-Oberwinkler S, Henn C, Möller G, et al. 17β-Hydroxysteroid dehydrogenases (17β-SD5) as therapeutic targets: protein structures, functions, and recent progress in inhibitor development. J Steroid Biochem Mol Biol 2011;125:66–82.
24. Spears V, Green AR, Atkin SL, et al. Activity and gene expression of 17-beta-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.
25. Poirier D. Contribution to the development of inhibitors of 17β-hydroxysteroid dehydrogenase types 1 and 7: key tools for studying and treating estrogen-dependent diseases. J Steroid Biochem Mol Biol 2011;125:83–94.
26. Blichert-Toft M, Christiansen P, Mørchidsen HT. Danish Breast Cancer Cooperative Group–DBCG: History, organization, and status of scientific achievements at 30-year anniversary. Acta Oncol 2008;47:497–505.
27. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 2014;29:541–9.
28. 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.
29. 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.
30. Møller S, Jensen MB, Ejfertsen 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;47:508–24.
31. Charlson ME, Pompei P, MacLehose RF, et al. Good practices for the use of the Charlson Comorbidity Index: a review of studies from 1987 to 2003. J Clin Epidemiol 2005;58:1007–13.
32. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, et al. Data Resource Profile: The Danish National Prescription Registry. Int J Epidemiol 2010;39:e12.
33. 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.
34. Cronin-Fenton DP, Helberg 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.
35. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. Clin Epidemiol 2010;2:273.
36. 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 and 15-year survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med 2010;7:e1000279.
37. Cronin-Fenton DP, Helberg 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.
38. Ehrenstein V, Antonsen S, Pedersen L. Existing data sources for clinical epidemiology: Aarhus University Prescription Database. Clin Epidemiol 2010;2:273.
39. 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.
40. 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. Cal 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-Garcia 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.