PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cohort Profile: The Predictors of Breast Cancer Recurrence (ProBe CaRE) Premenopausal Breast Cancer Cohort Study in Denmark
AUTHORS	Collin, Lindsay; Cronin-Fenton, Deirdre; Ahern, Thomas; Christiansen, P; Damkier, Per; Ejlertsen, Bent; Hamilton Dutoit, Stephen; Kjærsgaard, Anders; Silliman, Rebecca; Sørensen, Henrik T.; Lash, Timothy

VERSION 1 – REVIEW

REVIEWER	Werner Schroth Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Germany
REVIEW RETURNED	01-Feb-2018

GENERAL COMMENTS	This is about the description of a large population-based premenopausal breast cancer cohort, collected between 2002 and 2010, to be used for future biomarker-based investigations to predict clinical outcome. The short-term goal is to test for associations between gene variants possibly affecting tamoxifen metabolism and outcome, with additional stratification by tumor expression biomarkers. To demonstrate the usefulness of the database/biobank, the authors show first data suggesting that patients can be correctly classified for main prognostic parameters based on registry databases. Comments While there is nothing yet presented regarding their main goal, i.e. pharmacogenetics of tamoxifen metabolism, the size of the cohort (4600 patients classified as ER+ treated with tamoxifen, and 1359 patients classified as ER negative, tamoxifen untreated) appears useful to expect well-powered pharmacogenetic analyses. I can only recommend regarding their future plan: that is to include as many functional variants of CYP2D6 as possible, because this is key to predict variable pharmacokinetics of active metabolites, compared to the rest of their listed enzymes/genes. Stratification by ER beta and steroid dehydrogenase tumor expression to control for receptor and ligand-receptor interactions holds potential to resolve some of the current ambiguities. Specifics/minors -In the discussion of Bayesian analysis (ALPS), the authors mentioned the possibility to control for tamoxifen adherence; is there any information regarding this critical parameter from their database? -Limitations of the study can only be addressed when actual pharmacogenetics data are presented; as with the homogeneity regarding ethnic representation, this is likely less important

	compared to other confounders -Table 1: This table makes no sense given the variants are not specified; for example if the CYP3A variant does not refer to CYP3A4*22 (rs35599367) it must be criticized, yet this would be better placed in the actual pharmacogenetics study -Table 2: correct age group 35-39 -Suppl Table 2: Fix number: Total number of pts with non-neoplastic tissue available should be 81% if the rest is 19%

REVIEWER	Wakako Tsuji Shiga General Hospital
REVIEW RETURNED	21-Feb-2018

GENERAL COMMENTS	Summary: This cohort study included 5959 premenopausal breast cancer patients and found positive predictive values were tumor size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment. Specific comments and suggestions: 1. A reference is needed for the first sentence in the introduction section. 2. Authors raised 3 research questions, and described precisely. However, these results are not clear so far. Once readers read this article, they will want to know the results. However, there are no results for these questions. Research question section is too long. 3. Authors found positive predictive values for premenopausal women as tumor size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment. These factors have been well studied as a positive prognostic factors to date. What's the new finding?

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Werner Schroth
Institution and Country: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Germany
Please state any competing interests or state 'None declared': None declared

This is about the description of a large population-based premenopausal breast cancer cohort, collected between 2002 and 2010, to be used for future biomarker-based investigations to predict clinical outcome. The short-term goal is to test for associations between gene variants possibly affecting tamoxifen metabolism and outcome, with additional stratification by tumor expression biomarkers. To demonstrate the usefulness of the database/biobank, the authors show first data suggesting that patients can be correctly classified for main prognostic parameters based on registry databases.

While there is nothing yet presented regarding their main goal, i.e. pharmacogenetics of tamoxifen metabolism, the size of the cohort (4600 patients classified as ER+ treated with tamoxifen, and 1359 patients classified as ER negative, tamoxifen untreated) appears useful to expect well-powered pharmacogenetic analyses. I can only recommend regarding their future plan: that is to include as many functional variants of CYP2D6 as possible, because this is key to predict variable pharmacokinetics of active metabolites, compared to the rest of their listed enzymes/genes.

Stratification by ER beta and steroid dehydrogenase tumor expression to control for receptor and ligand-receptor interactions holds potential to resolve some of the current ambiguities.

Specifics/minors
Comment#1:
-In the discussion of Bayesian analysis (ALPS), the authors mentioned the possibility to control for tamoxifen adherence; is there any information regarding this critical parameter from their database?

Response: We appreciate the reviewer's perspective and agree of the importance of the ability to control for tamoxifen adherence. On **page 11**, we have included further information regarding patient follow-up while undergoing treatment. Patients are required to pick up endocrine therapy from the treating hospital, and that information will be used to estimate tamoxifen adherence.

Comment#2:
-Limitations of the study can only be addressed when actual pharmacogenetics data are presented; as with the homogeneity regarding ethnic representation, this is likely less important compared to other confounders -Table 1: This table makes no sense given the variants are not specified; for example if the CYP3A variant does not refer to CYP3A4*22 (rs35599367) it must be criticized, yet this would be better placed in the actual pharmacogenetics study

Response: We agree with the reviewer that **Table 1** would be strengthened by the inclusion of the SNPs included in the pharmacogenetic analysis, **Table 1** now includes the SNPs used in the genotyping assays.

Comment#3:
-Table 2: correct age group 35-39

Response: We thank the reviewer for noticing this typo, the age group has been corrected in **Table 2**.

Comment#4:
-Suppl Table 2: Fix number: Total number of pts with non-neoplastic tissue available should be 81% if the rest is 19%

Response: We thank the reviewer for noticing this detail, the reported percentages have been corrected in **Supplemental Table 2**.

Reviewer: 2
Reviewer Name: Wakako Tsuji
Institution and Country: Shiga General Hospital Please state any competing interests or state 'None declared': None declared

Summary:
This cohort study included 5959 premenopausal breast cancer patients and found positive predictive values were tumor size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment.

Specific comments and suggestions:

Comment#1
A reference is needed for the first sentence in the introduction section.

Response: We thank the reviewer for catching this oversight, a reference has been added to the first sentence of the introduction section.

Comment#2:
Authors raised 3 research questions, and described precisely. However, these results are not clear so far. Once readers read this article, they will want to know the results. However, there are no results for these questions. Research question section is too long.

Response: We appreciate the author's comments and perspective that we have precisely described the three research questions. In line with BMJ Open submission categories, this manuscript follows the guidelines outlined for a **Cohort Profile** with "the goal to describe in detail the objectives to be carried out in our Cohort study and its future plans". As the ProBe CaRe cohort is newly established, we do not have results from our research aims to report and have only published results from our validation substudy.

Comment#3:
Authors found positive predictive values for premenopausal women as tumor size, lymph node involvement, receptor status, surgery type, receipt of radiotherapy, receipt of chemotherapy and tamoxifen treatment. These factors have been well studied as a positive prognostic factors to date. What's the new finding?

Response: Where we appreciate the reviewer's perspective that these findings may not be novel, we have calculated positive and negative predictive values comparing the registry information to that of the medical records, which will provide valuable classification parameters for future use in bias analyses that will be conducted in conjunction with our analyses in our primary study aims, which will rule out any results driven by information bias, improving the quality of reported results. As these are the first results to come from this cohort study, we thought them important to share and they demonstrate that we have carefully thought through future analyses that may suffer from information bias.

VERSION 2 – REVIEW

REVIEWER	Wakako Tsuji Shiga General Hospital, Japan
REVIEW RETURNED	28-Mar-2018

GENERAL COMMENTS	Although the authors failed to show novel findings, it's relevant to state positive and negative predictive values based on this large cohort study. In the future, authors should report the long-term results from the pharmacogenomics study.

REVIEWER	Werner Schroth Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
REVIEW RETURNED	28-Mar-2018

GENERAL COMMENTS	In their revised version, the authors address a number of points including the control of drug adherence and by providing a clear-text table of planned SNP investigations. While this study presents as a solid pilot study to demonstrate the usefulness of the database/biobank to perform a future pharmacogenetics study, their objective to study potential classifiers of impaired tamoxifen metabolism in premenopausal women with breast cancer raises some concern regarding the choice of predictive genetic variants. According to their presented table, this refers to CYP2D6 and CYP3A. Because CYP2D6 is the key player to predict impaired endoxifen formation, the list of genetic variants associated with CYP2D6 function is incomplete or implausible: for example, rs28371706 is the major functional variant of the *17 "intermediate metabolizer" haplotype that is found mainly in Africans, so why including this in a study with Europeans? More importantly, their planned prediction of CYP2D6 activity is below standard, as a

	number of important alleles including *3, *5, *6, *7, *9 are not covered. In addition, a number of pharmacogenetic studies showed that prediction of CYP3A4 function is among the more important factors to explain impaired tamoxifen metabolism. To this end, the CYP3A4*22 variant was clearly demonstrated as one of the few known functional CYP3A4 predictors, and is a must in terms of pharmacogenetics. In summary, the goal to explain as much variability in pharmacokinetics of active tamoxifen metabolites as possible will not be achieved, based on the table 1 variant list; of course this criticism would be better placed in their actual pharmacogenetics study.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:
Reviewer: 1
Reviewer Name: Werner Schroth
Institution and Country: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Germany
Please state any competing interests or state 'None declared': None declared

*In their revised version, the authors address a number of points including the control of drug adherence and by providing a clear-text table of planned SNP investigations. While this study presents as a solid pilot study to demonstrate the usefulness of the database/biobank to perform a future pharmacogenetics study, their objective to study potential classifiers of impaired tamoxifen metabolism in premenopausal women with breast cancer raises some concern regarding the choice of predictive genetic variants. According to their presented table, this refers to CYP2D6 and CYP3A. Because CYP2D6 is the key player to predict impaired endoxifen formation, the list of genetic variants associated with CYP2D6 function is incomplete or implausible: for example, rs28371706 is the major functional variant of the *17 "intermediate metabolizer" haplotype that is found mainly in Africans, so why including this in a study with Europeans? More importantly, their planned prediction of CYP2D6 activity is below standard, as a number of important alleles including *3, *5, *6, *7, *9 are not covered. In addition, a number of pharmacogenetic studies showed that prediction of CYP3A4 function is among the more important factors to explain impaired tamoxifen metabolism. To this end, the CYP3A4*22 variant was clearly demonstrated as one of the few known functional CYP3A4 predictors, and is a must in terms of pharmacogenetics. In summary, the goal to explain as much variability in pharmacokinetics of active tamoxifen metabolites as possible will not be achieved, based on the table 1 variant list; of course this criticism would be better placed in their actual pharmacogenetics study.*

Response:

We thank the reviewer for the comments and appreciate the expertise in the pharmacogenetics of this project. We agree that under ideal circumstances we would have included a more comprehensive list of genetic variants for the pharmacogenetic aspects of this study. The process by which we selected the final list of variants in **Table 1**, was based on several important factors, some of which were practicalities. The original list included all reported functional genetic variants for enzymes in **Figure 1** (approximately 120 SNPs). We attempted to sequence all of these variants, but DNA quality extracted from FFPE led to too high a failure rate. This list was then shortened to include SNPs with a minor allele frequency in women of European descent >5%, a known functional polymorphism, and with an existing Taqman kit suitable for DNA extracted from FFPE. This list was further reduced based on the successful genotyping in our pilot studies and the success rate of genotyping using FFPE samples. We acknowledge that this may not be the optimal set of variants, however we have been able to include many important functional genetic variants involved in tamoxifen metabolism with high genotyping success rate and within the practical constraints.

Reviewer: 2
Reviewer Name: Wakako Tsuji
Institution and Country: Shiga General Hospital Please state any competing interests or state 'None declared': None declared

Summary:
Although the authors failed to show novel findings, it's relevant to state positive and negative predictive values based on this large cohort study. In the future, authors should report the long-term results from the pharmacogenomics study

Response:

We thank the reviewer for their comments and we plan to report the results from the pharmacogenomics study in the near future. Describing the cohort in detail will provide a valuable reference for the main aims and other papers, allowing us to refer readers to these methodologic details while allowing space in those papers to focus on the salient methodological details.