Author’s response to reviews

Title: Coronary Heart Disease and Mortality Following a Breast Cancer Diagnosis

Authors:
Aixia Guo (gax8118@gmail.com)
Kathleen Zhang (kwzhang@wustl.edu)
Kristi Reynolds (risti.reynolds@kp.org)
Randi Foraker (randi.foraker@wustl.edu)

Version: 1 Date: 13 Mar 2020

Author’s response to reviews:

Response to reviewer comments

We would like to thank reviewers for their helpful comments. Please find our responses to the reviews below. We hope you find the revised manuscript suitable for publication.

Ruoqi Liu (Reviewer 1): The study examines the independent and joint effects of CVH and cancer treatments on CHD and mortality via statistical analysis and machine learning techniques. While the topic is significant and the experiments are comprehensive, the paper needs some improvements.

1. The methodology part:

   a. How to represent the features for each patient in the machine learning prediction model? What's the dimension of the feature vectors in this framework?

   b. How to represent the interaction of CVH and treatment as the features in the model (i.e., directly concatenate or take average)? Please illustrate how do you preprocess and construct these features.

Thank you for the great questions. In this study, we utilized 5 CVH submetrics and 8 treatment categories for each patient. We then combined the 5 CVH submetrics into an overall CVH score/feature following published criteria (i.e., the American Heart Association CVH metric), and also combined the 8 categories of treatment as one score/feature, (i.e., receipt of treatment or not).

We also tested the 5 CVH submetrics and all 8 treatment categories as input features for all classification models. We obtained similar results as using the CVH score and the treatment...
score. In order to make the model easier to interpret, we used the latter construction of features for the manuscript. (We have added this language on page 4-5.)

2. The experiment part:
   a. The explanation and analysis of results in Table 3 and Figure 3 are not sufficient. How do these results contribute to your study purpose (the examination of independent & joint effects of CVH and treatments)? The higher accuracy of CVH+treatment than only using one single feature is highly likely because more information is used to train the model. Please highlight the indication lies behind the results.

   We agree with the reviewer, and have added the following to the results section of the manuscript as further explanation of Table 3 and Figure 3: “The results from Table 3 and Figure 3 indicate that all three models achieved higher accuracy with the inclusion of joint effects as compared to only individual effects. Specifically, models which include both CVH and receipt of treatment data provide additional information and improve the prediction of CHD and death. Patients with poor overall CVH who received cancer treatments had the highest risk of CHD and death.”

Na-Jin Park (Reviewer 2):

3. The study addresses the current gap of literature on cardiovascular health in breast cancer survivors: a joint effect (interaction) of pre-treatment cardiovascular health prior to breast cancer (BC) and cardiotoxic BC treatment on post-treatment coronary heart disease (CHD) and all-cause mortality following 10 years of BC diagnosis. That gap is due to, mainly, the lack of available datasets including both pre-treatment/diagnosis cardiovascular risk factors and details of post-BC treatment. The authors used EHR data from a medical center. Limitations of EHR data were briefly addressed but should include deficit of important information, such as physical activity and diet for the complete list of cardiovascular health (CVH) measures. I also wonder why details of BC data on diagnosis and treatment are missing from the EHR, which would be helpful to understand the study sample of BC survivors.

   Thank you for pointing out these omissions. In the limitations section of the revised manuscript, we have expanded upon our previous statements as follows: “Physical activity and diet data are not commonly recorded in the EHR as structured, actionable data elements. If physical activity and diet data do exist in the EHR, they are usually recorded as clinical notes using free text. Importantly, these data are not easily translated into the American Heart Association’s metric definitions and thus are not actionable at the point-of-care or easily incorporated into risk scoring algorithms. Similarly, data on diagnosis and treatment, have not always been stored as structured data elements in the EHR. Conducting this analysis required mining data from legacy EHR systems and for pragmatic reasons we accessed only structured data elements for this analysis, which resulted in incomplete data ascertainment.”
4. While the topic itself is innovative, there are similar publications of cohort observational studies addressing independent effects of existing (post-BC) cardiovascular risk factors and BC treatments (in a separate or combined regimen) on cardiovascular disease (CVD) with mixed findings in variously different characteristics of BC survivors. Unlike other previous studies, this study particularly used machine learning algorithms labeled as "novel" statistical techniques without appropriate justifications and implications relevant to clinical decision-making, which is the main focus of the journal. Therefore, the goal statement here is incomplete without that.

These are a great points. In accordance with your suggestion to update references (listed in this document as #8), we have included many of these publications. To enhance the implications of this work and to make it more relevant to clinical decision-making, we have elaborated on our ongoing efforts to develop and validate clinical decision support tools which incorporate risk prediction algorithms for cancer survivors. As indicated in our response to #7, below, we include this in the conclusions section. We have also incorporated this into the goal statement.

5. From the background, (left-sided) radiation therapy seems the focal exposure among many types of cardiotoxic therapies, which suddenly disappeared in reporting actual data. Total patients who received BC treatments were reported as 17.6% in the study in Table 2. Isn't that odd that in 2006-2007, women with BC diagnosis had treatment so low like that? Only 2% had anthracyclines. There is no radiation therapy reported. Aromatase inhibitors (Ais) are the major hormonal or endocrine therapy, but hormone therapy here appears not to include AIs. Then what include hormone therapy here? I guess tamoxifen? How about Herceptin for HER+ BC? 72% had hormone therapy or aromatase inhibitors, which makes sense in that over 70% of BC cases are hormone receptor-positive in national statistics. It is plausible that those who received any therapies had more advanced BC and poorer prognosis/mortality outcomes given the very low treatment record. My experience additionally tells me that those with advanced BC have poorer CVH compared to early-stage disease. BC-related death statistics (along with CHD/CVD mortality) is missing, which is also important to better understand all-cause mortality in this study sample of BC survivors.

The reviewer identifies a very important future direction for this study, which is to characterize all-cause mortality among the sample of breast cancer survivors. For this work, we focused on incident cardiovascular disease. Please see the table below for the types of therapy that were queried with respect to drug class.

Table. Example names of each class of drug

Anthracyclines

- Doxorubicin
- Epirubicin
- Donarubicin
• Daunorubicin
• Idarubicin
• Valrubicin

Hormone therapy
• Tamoxifen
• Evista / Raloxifene
• Fareston / Toremifene
• Faslodex / Fulvestrant
• Flutamide
• Zoladex
• Casodex
• Lupron

Aromatase inhibitors
• Arimidex / Anastrozole
• Femara / Letrozole
• Aromasin / Exemestane

Monoclonal antibody
• Herceptin / Trastuzumab
• Zevalin / Ibritubomab
• Tiuxetan
• Rituximab
• Campath / Alemtuzumab
• Adcetris / Brentuximab vedotin
• Kadcyla / Ado-trastuzumab emtansine
• Ontak / Denileukin diftitox
• Blincyto / Blinatumomab

Antimicrotubule agents

• Taxanes
• Paclitaxel
• Nab-paclitaxel
• Docetaxel
• Eribulin
• Cabaxitaxel
• Ixempra / Ixabepilone

Alkylating agents

• Platinol / Platinol-AQ / Cisplatin
• Paraplatin / Carboplatin
• Cytoxan / Neosar / Cytoxan Lyophilized / Cyclophosphamide
• Ifex / Ifosfamide
• Myleran / Busulfan
• Treanda / Bendamustine
• Leukeran / Chlorambucil
• Temodar / Temozolomide
• Alkeran / Evomela / Melphalan
• Busulfex / Bendeka / Busulfan / Bendamustine
• BiCNU / Carmustine
• CCNSB Capsules / CeeNU / Lomustine
• Eloxatin / Oxaliplatin
• Mustargen / Meclorethamine
• Tepadina / Thioplex / Thiotepa
• Yondelis / Trabectedin
• Zanosar / Streptozocin

Antimetabolites
• Adrucil / 5-fluorouracil
• Arabinosylcytosine
• Leustatin / Novaplus / Cladribine
• Otrexup / Trexall / Rheumatrex / Methotrexate
• Purinethol / Mercaptopurine
• Alimta / Pemetrexed
• Xeloda / Capecitabine
• Gemzar / Gemcitabine
• Hydrea / Droxia / Hydroxyurea
• Fludara / Oforta / Fludarabine
• Folotyn / Pralatrexate
• Arranon / Nelarabine
• Clofarabine
• Dacogen / Decitabine
• DepoCyt / Cytarabine liposomal
• FUDR / Floxuridine
6. CVH metric by the AHA include 7 health factors, including 5 used in the study plus physical activity and diet. However, Table 1 with 5 measures may mislead readers because the clear description of justification on that decision is missing (I guess the EHR had no reliable data about them). I also want to see specific numbers and percentages for each measure and final categories of poor, intermediate, and ideal health as described. I think that basic information is particularly important with arbitral use of CVH components in this study. I don't understand why CVH scores in Figure 1 (intermediate) had 1 or 2 (ideal) without 0 (poor). I also want the authors to be clear about these categories that were determined by CVH 5 factors. Given that, it may be interesting and better make sense to look at interactions of individual factors and BC treatment on outcomes, which is another major gap in the current literature.

We appreciate this comment, and have included information about the missing two metrics (physical activity and diet) in the methods section of the manuscript. According to the reviewer’s prior suggestion, we have elaborated on this gap in the limitation section (please see our response to #3, above). Please see below for a table which shows the distribution of categories of poor, intermediate, and ideal CVH.

We agree that the reviewer’s suggestion to evaluate interactions of individual CVH metrics with breast cancer treatments is a logical next step for these analyses. We have suggested this as a next step in the discussions section of the manuscript.

Table. Counts and percentages of ideal, intermediate, poor category of the 5 CVH metrics

<table>
<thead>
<tr>
<th>CVH/Category</th>
<th>n (%)</th>
<th>Ideal</th>
<th>Intermediate</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body mass index (BMI)</td>
<td></td>
<td>826 (99.3)</td>
<td>3 (0.4)</td>
<td>3 (0.4)</td>
</tr>
<tr>
<td>Cholesterol</td>
<td></td>
<td>205 (29.2)</td>
<td>437 (62.2)</td>
<td>61 (8.7)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td></td>
<td>282 (33.3)</td>
<td>432 (51.1)</td>
<td>132 (15.6)</td>
</tr>
<tr>
<td>Glucose/hemoglobin A1c</td>
<td></td>
<td>40 (8.8)</td>
<td>329 (72.3)</td>
<td>86 (18.9)</td>
</tr>
<tr>
<td>Smoking status*</td>
<td></td>
<td>0 (0)</td>
<td>1868 (96.6)</td>
<td>66 (3.4)</td>
</tr>
</tbody>
</table>

* We classified current smoking as 0 points, and not smoking as 1 point, as there were no data available to indicate if they had never smoked or had quit for more than one year (representing the “ideal” category).

In the methods section of the manuscript, we have stated that “not all CVH data were complete for all women; therefore, we imputed a value of 2 points for all missing submetric values. We
tested the robustness of this strategy by imputing a value of 1 or 0, respectively, for missing values.” We imputed a value of 2 points for the missing submetric values due to 2 reasons: By imputing values of (1) and (0), respectively, we observed similar results, trends, and patterns to those shown in the Figure 2 and 3; and we assumed from a clinical perspective that a patient had a missing value because the providers assumed that it was not necessary to do the test as it was ideal. Given the above two reasons, we show the results from imputing missing values as 2 points.

7. Discussion needs more specifics and elaborations relevant to clinical decision-making process in oncology survivorship practice.

Thank you for this suggestion. We have elaborated in the discussion section to include information on our ongoing work, which involves deploying and evaluating clinical decision support in the cancer survivorship setting for managing cardiovascular late effects among cancer survivors. Our clinical decision support system (CDSS) presents CVH and cancer treatment data separately in the EHR-embedded data visualization. Our goal is to one day integrate a validated cardiovascular risk algorithm into our existing CDSS to better target cardiovascular disease prevention and management efforts in cancer survivorship.

8. The whole list of references should be updated to 2020. In last 5 years, there has been an increasing data on BC and cardiovascular disease (CVD) outcomes, including the 2018 AHA publication on Circulation. However, authors provided many 2011 publications despite newly available references. Also, statistics of BC survivors (references 4 & 5) is old.

We agree, and have added more recent and relevant references, as shown below.


• Han X, Zhou Y, Liu W. Precision cardio-oncology: understanding the cardiotoxicity of cancer therapy. npj Precis Oncol. 2017. doi:10.1038/s41698-017-0034-x


We additionally replaced the old references 4 & 5 with the following two:


7. Please consider to revise and reorganize the tables and figures. I consider some figures or information could be better in supplementary materials. Overall, it would be much improved with focused presentation of data and providing specific interpretation for BC survivors and decision-making implications for improved BC care. The revision needs to be more reader-friendly.

Thank you for the great suggestion. We have moved Table 3 to supplementary materials and more clearly described our results per recommendations from Reviewer 1.