To complete model calibration, please run the files in the order presented in the table below, and send corresponding output that is noted in “Files to send Kristen” to [Kristen\_brantley@dfci.harvard.edu](mailto:Kristen_brantley@dfci.harvard.edu). Please note: the first two files contain functions. These files will need to be read into your R environment, but no output is produced directly from them. Specific notes on how to run each model (Ext\_Model\_Calibration\_Gail.R, Ext\_Model\_Calibration\_iCARE.R, Ext\_Model\_Calibration\_BWHS.R, and Ext\_Model\_Calibration\_TyrerCuzick.R) are provided below the table.

The calibration codes are all written in .Rmarkdown format. They will be knitted at the end of the code using render (so don’t select “knit” in your Rstudio environment”- see instructions below). If you prefer to run these in batch mode in R that is fine, just send the .Rout files.

|  |  |  |
| --- | --- | --- |
| File | Intermediate datasets | Files to send to Kristen |
| Table1\_Fun.R | N/A | N/A |
| Functions\_ExistingModelValidation.R | N/A | N/A |
| BWHS\_incmortrates.R | N/A | N/A |
| Step1.DataClean\_CTS.R | **Clean dataset: FILEPATH/bcrpp\_valdata.csv** | N/A |
| Table1.R |  | Table1\_CTS.csv |
| Step2.Create2datasets.Rmd | 2 datasets will be created in your environment: **valdata and valdata2.** | Step2.Create2Datasets.html |
| Ext\_Calibration\_Gail.Rmd | N/A | Ext\_Calibration\_Gail.html |
| iCARE-Lit-Dependencies (see iCARE instructions below) | N/A | N/A |
| Ext\_Calibration\_iCARE.Rmd | N/A | Ext\_Calibration\_iCARE.html |
| Step2.Create2datasets\_BWHS.Rmd | 2 datasets will be created in your environment: **valdata and valdata2.** | Step2.Create2Datasets\_BWHS.html |
| Ext\_Calibration\_BWHS.Rmd | N/A | Ext\_Calibration\_BWHS.html |
| TC\_datasetup.R | This will create the datasets you can use with the IBIS risk evaluator.  **data\_1to5y\_assump1\_1.csv**  **data\_1to5y\_assump2\_1.csv**  **data\_1to5y\_assump3\_1.csv**  **data\_1to5y\_assump4\_1.csv** | N/A |
| Ext\_Model\_Calibration\_TyrerCuzick.Rmd | N/A | Ext\_ModelCalibration\_TyrerCuzick.html |

**Step 1: Clean data, Create Table 1**

1. Run **Step1.DataClean\_CTS.R –** this will save the dataset
2. In the same folder that you save the clean dataset, save:
   1. Functions\_ExistingModelValidation.R
   2. Table1\_Fun.R
3. **Run Table1.R**
4. Send resulting Table1.csv file

**Step 2: Create 2 datasets**

1. **Run Step2.Create2datasets.Rmd**
2. The final line of the code will create the .html file to share **“Create2datasets.html”**

**Step 3: Calibration**

**Details for Calibration models**: All models start by using data with follow-up 1-5 y (**valdata**) and 2-6 y (**valdata2**) created in Step2.Create2datasets.Rmd or Step2.Create2datasets\_BWHS.Rmd. The instructions within the .Rmd files are included so that you can chunk by chunk if needed (i.e., if an error arises). You can also simply save the .Rmd files to the folder that your dataset and all other R files are stored within and then run the command **rmarkdown::render(“filepath/FILENAME.Rmd”) in the R console**. See notes below on what is needed in your R environment before this will work. All the information that I need will be within the documents produced from the individual calibration codes. This includes the following information:

(1) distribution of absolute and relative risk in your dataset

(2) E/O ratio (overall and by absolute risk deciles)

(3) Absolute risk plots (observed v. predicted absolute risk by absolute risk deciles)

(4) AUC : based on the absolute risk, we will obtain an AUC and an age-adjusted AUC

**Gail**

1. If valdata and valdata2 are in your R environment jump to Step3.
2. Run **Step1.DataClean\_CTS.R** & **Step2.Create2datasets.Rmd.** 
   1. No need to knit this again, we just need the datasets in your R environment.Do not clear the environment after running this step.
3. Save **Functions\_ExistingModelValidation.R** into your folder so that you can source this file within the next part.
4. Open the: **Ext\_Model\_Calibration\_Gail.Rmd** code & follow instructions. Chunks titled {r instructions} contain text only and are there to help guide you. Replace all “filepath” text with the appropriate filepaths for your datasets. Steps are commented on in the code. Briefly, the code will reformat your dataset to fit within the context of the Gail model, will read in files required for the BCRAT model to work. You will predict absolute and relative 5-year risks for participants in CTS using this code. Then, iCARE-Calibrate will be used with accompanying functions (provided: Functions\_ExistingModelValidation.R) to produce the results.
5. In the console type: **rmarkdown::render("FILEPATH/Ext\_Model\_Calibration\_Gail.Rmd").** This will produce thehtml file, saved within the same folder.

**iCARE**

1. Run **Step1.DataClean\_CTS.R** & **Step2.Create2datasets.Rmd.** 
   1. No need to knit this again, we just need the datasets valdata and valdata2 in your current R environment.Do not clear the environment after running this step.
2. **Load iCARE model parameters (iCARE-Lit-dependencies**). Simply open the folder containing the dependencies when R is open. R will ask if you want to load these into your environment. You say yes to the files you want to load. These will include the model parameter files. You have created your own reference datasets and have incidence and mortality rates from the UK to use, so make sure to use those and not the ones stored in the zip file (they are US based).
3. Open the: **Ext\_Model\_Calibration\_iCARELit.R** code & follow instructions. Replace all “filepath” with the correct filepaths.
4. In the console type: **rmarkdown::render("FILEPATH/Ext\_Model\_Calibration\_iCARELit.Rmd").** This will produce the html file to send, saved within the same folder.

**BWHS**

**NOTE: You likely don’t have enough black women to run this model. If you have >=100 cases of black women with BC diagnosed in years 1-5y then you can run this. Otherwise, do not bother.**

1. Best to clear your R environment before running this, as new datasets are created
2. Using the clean dataset created in Step1.DataClean\_CTS.R : Run **Step2.Create2datasets\_BWHS.Rmd**. In the R console type: rmarkdown::render("FILEPATH/Step2.Create2datasets\_BWHS.Rmd") and send html output.
   1. Please note this is slightly different than that used for the other models, as we are restricting to Black women and changing the age range. **Do not clear the environment after running this step.** Now valdata and valdata2 are specific for Black women 30-70y
3. You will need to make sure **BWHS\_incmortrates.R** is saved in your working folder, as it is sourced within the code
4. Open the **Ext\_Model\_Calibration\_BWHS.Rmd** code.
5. In the console, type: rmarkdown::render("FILEPATH/Step2.Create2datasets\_BWHS.Rmd")

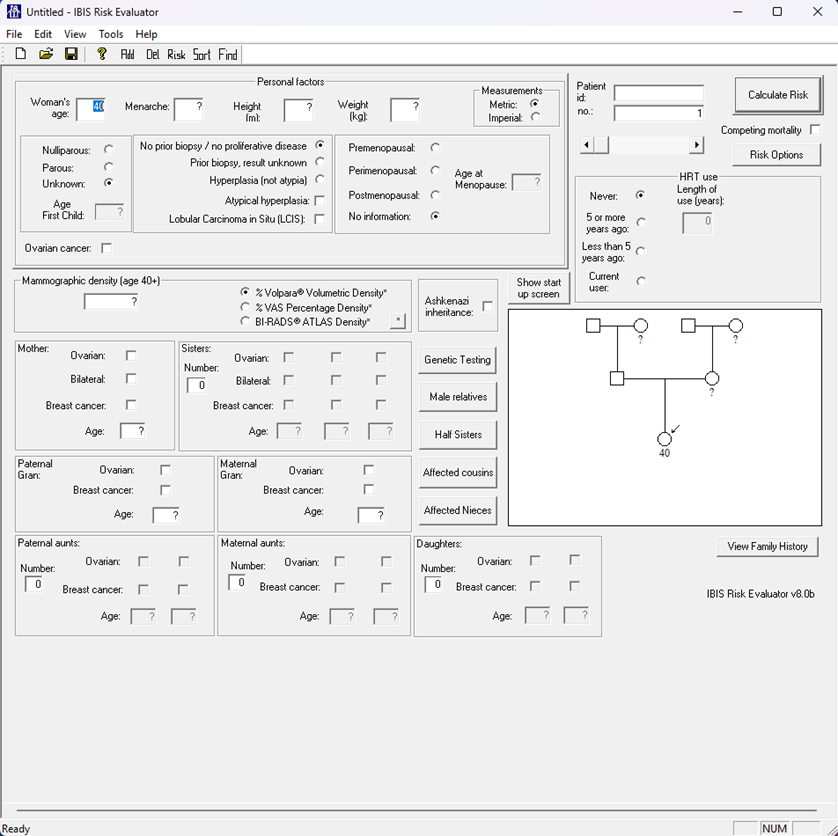
**TC**

1. Best to clear your environment before this
2. Run **Step1.DataClean\_CTS.R** & **Step2.Create2datasets.Rmd.** The datasets valdata and valdata2 will be used first.
   1. No need to knit this again, we just need the datasets valdata and valdata2 in your current R environment.Do not clear the environment after running this step
3. Run **Ext\_TC\_datasetup.R.** This creates the code in the TC format needed. As it is written this code assumes your cohort has NO information on whether sisters or daughters have BC. IF YOUR COHORT DOES HAVE THIS INFORMATION please stop here and connect with Kristen Brantley.
   1. Please pay attention to instructions at the end of the R file. If you have more than 25000 individuals in your cohort, you will need to create multiple files, otherwise the calculator tends to fail. You will also need to make sure your files replace “N” with the number of individuals in each file (see R code instructions). **There will be 4 datasets created, please name so that it includes the period covered \_1to5y, the assumption covered (e.g., \_assump1) and the file number (\_1). The file number will be 1 for the first 25000 participants. If you have more individuals you will need to create \_2 starting at 25001:50000, etc. The last file needs to have the exact number of individuals specified.**
4. **Files produced [number of datasets will vary based on dataset size]:** 
   1. data\_1to5y\_assump1\_1.csv
   2. data\_1to5y\_assump2\_1.csv
   3. data\_2to6y\_assump1\_1.csv
   4. data\_2to6y\_assump2\_1.csv
5. Download the IBIS Risk Evaluator v8: <https://ems-trials.org/riskevaluator/>

A screenshot of a software

AI-generated content may be incorrect.

1. Open the IBIS Risk Evaluator – it will look like this:



1. Click the “open folder” button on the top left. Select your file of interest. Click “open” . This will load your data into the risk calculator.

A screenshot of a computer

AI-generated content may be incorrect.

1. Select “Risk” at the top menu. It will prompt you to save the risk estimates. Name your files aligning with earlier names. The “\_1” at the end references the number of your dataset.
   1. **data\_1to5y\_assump1\_risks\_1**
   2. **data\_1to5y\_assump2\_risks\_1**
   3. **data\_2to6y\_assump1\_risks\_1**
   4. **data\_2to6y\_assump2\_risks\_1**

A screenshot of a computer

AI-generated content may be incorrect.

1. Now your risk files are saved, and it is time to implement iCARE Calibrate
2. Open R code: **Ext\_Model\_Calibration\_TyrerCuzick.Rmd.** Follow instructions within R code to complete model calibration.
   1. Note: the section “Combine Files” you will need to make sure the valdata and valdata2 datasets are loaded in your R environment, from Create2datasets.Rmd
3. In the console type: rmarkdown::render("FILEPATH/Ext\_Model\_Calibration\_TyrerCuzick.Rmd") and send the html file