BREAST CANCER RISK ASSESSMENT Survey, 2021

Study Documentation

Table of Contents

BREAST CANCER RISK ASSESSMENT Study Description

Citation	
Title Statement	
Title:	Breast Cancer Risk Assessment Survey Study
Responsibility Statement	
Authoring Entity:	Rensselaer Polytechnic Institute, Vasundhara Acharya
Production Statement	
Producer:	Vasundhara Acharya
Date of Production:	2021
Place of Production:	Rensselaer Polytechnic Institute, Troy, New York, United States of America
Distribution Statement	
Distributor:	Vasundhara Acharya, PhD student at Rensselaer Polytechnic Institute, Troy, New York, USA
Date of Distribution:	2021
Version Statement	
Version:	First Version
Bibliographic Citation	
	Breast Cancer survey, October 2021 [machine readable data file] / conducted by Vasundhara Acharya , Principal Investigator [producer and distributor] ,Rensselaer Polytechnic Institute. Troy, New York.

Study Scope	
Abstract	
	The survey is used as a tool to estimate risk for next 5 years and lifetime risk of developing breast cancer. The data of their first-degree relatives having breast cancer gives a better estimate of them developing the cancer in coming years. The breast cancer risk assessment tool can help in the clinical management of patient seeking advice concerning screening and prevention. It can be used to estimate a person's probability of developing breast cancer without undergoing an invasive technique.
Summary Data Description	
Time Period:	September 2021 - October 2021
Date(s) of Collection:	September 2021
Country:	United States of America
Geographic Coverage:	to women across United States of America and India.

Unit of Analysis:

The survey was sent

Individuals (Women)

Universe: The universe is the women living in different parts of United States of America and India

Kind of Data: Survey Data

Methodology and Processing

Data Collection Methodology

Data Collector: Vasundhara Acharya ,Principal Investigator, PhD Student at Rensselaer Polytechnic Institute, Troy, New York.

Mode of Data Collection: Google Form
.

Mode of Data Collection: Survey via Google Form

Type of Research Instrument: Structured

Weighting: For all attributes present in the data file except for the patient ID all have been given a single weight and used to compute the breast cancer risk estimate.

Data Labeling

Data labeling: The different cases collected via the survey form were labeled into three categories: Low Risk, Moderate Risk and High Risk.

Domain Expert :   Dr. Prasad J Kamath, M.D , D.G.O , Gynaecologist and Obstetrician

Hospital Name: Anupam Hospital

Location :   Near Shivaji Statue, Off Manpada Road, Dombivali (E), Thane-421203, India.

Contact :   Phone: +91-251-2862580

Data Appraisal Information

Response Rate: The survey was distributed to different parts of USA and India.   Most of the responses are recorded from India. The current version consists of 151 responses (records). The size of the file is approximately 35KB.

Data Access

Dataset Availability

Location: GitHub Repository. https://github.com/Vasundhara1993-acharya/Breast-Cancer-Survey.git

Dataset Availability

Location: Online

Extent of Collection: 1 data file + 1 machine- readable documentation (PDF) containing the metadata (DDI format)

Number of Files: 2

Other Study Description Materials

Related Tools

"Weka 3: Machine Learning Software in Java," Weka 3 - Data Mining with Open Source Machine Learning Software in Java. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/. [Accessed: 01-Oct-2021].

Related Material

"The Breast Cancer Risk Assessment Tool," Breast Cancer Risk Assessment Tool. [Online]. Available: https://bcrisktool.cancer.gov/. [Accessed: 01-Oct-2021].

Related Publication(s)

- Erbil N, Nursel D, Çiğdem İ, Bölükbaş N. Breast cancer risk assessment using the Gail model: a Turkish study
- Keane H, Huilgol YS, Shieh Y, Tice JA, Belkora J, Sepucha K, Shibley WP, Wang T, Che M, Goodman D, Ozanne E. Development and pilot of an online, personalized risk assessment tool for a breast cancer precision medicine trial. npj Breast Cancer. 2021 Jun 17;7(1):1-9.
- Pillai D, Hossain SS, Chattu VK. Developing a community-based breast cancer risk prediction tool for resource-poor settings. Journal of education and health promotion. 2019;8.
- Samimi G, Heckman-Stoddard BM, Holmberg C, Tennant B, Sheppard BB, Coa KI, Kay SS, Ford LG, Szabo E, Minasian LM. Assessment of and interventions for women at high risk for breast or ovarian cancer: A survey of primary care physicians. Cancer Prevention Research. 2021 Feb 1;14(2):205-14.

Data Files Description

File Structure (hierarchical) and File by File Description

File Name:	Breast Cancer Risk Assessment Survey Study
File Contents:	The survey to estimate the breast cancer risk estimates the absolute 5 year risk and lifetime risk of developing breast cancer. The data of their first-degree relatives having breast cancer gives a better estimate of them developing the cancer in coming years. This file contains responses to the various quantitative and qualitative questions asked. The responses to questions relating to their personal medical history , their first degree relatives having cancer and so on are recorded in this file.

Label:	Individual Women
Dimensions:	• No. of Variables:18
	• No. of Cases: 151
	• Record Length:151

Breast Cancer Risk Assessment Survey File Variable Description

Time Stamp

Length: width:17; decimal: 0

Range of Valid Data Values: Valid date and time

Universe: All participants (Women)

Text:　　　　Time at which the survey was answered. The time is stored following ISO standard

Value	Label
Valid date and Time	Time

Consent

Length: width:1; decimal: 0

Range of Valid Data Values: 0-1

Universe: All participants (Women)

Text:　　　Please Indicate your Consent Before Proceeding

Value	Label
0	No (Consent is not given for the study)
1	Yes (Consent is given for the study)

Description*: The patients are asked to give their consent regarding their participation in the study. They can withdraw their consent at any point during the study by emailing the Principal Investigator

Patient ID

Length: width:1-3; decimal: 0

Range of Valid Data Values: 1-151

Universe: All participants (Women)

Text:　　　　Patient are identified with id instead of names to maintain confidentiality.

Value	Label
1-151	Patient_ID

Age

Length: width:2-3; decimal: 0

Range of Valid Data Values: 24-100

Universe: All participants (Women)

Text:　　　Please Select Your Age

Value	Label
24-100	Age

Description*: The patients are asked to enter their age.

Race/ Ethnicity

Length: width:1; decimal: 0

Range of Valid Data Values: 0-4

Universe: All participants (Women)

Text: Select the Race/Ethnicity

Description*: The patients are asked to select their race/ethnicity. Race refers to a person's physical characteristics, such as bone structure and skin, hair, or eye color

Value	Label
0	White
1	Asian
2	Caucasian
3	Hispanic
4	Other

Q1: Age at first menstrual period

Length: width:1; decimal: 0

Range of Valid Data Values: 1-3

Universe: All participants (Women)

Text: What was your age at the time of your first menstrual period?

Description*: Menstrual period refers to a woman's monthly bleeding. It is often termed as period. The patients are asked to enter their age during which they first got their menstrual period.

Value	Label
1	7-11
2	12-14
3	Above 14

Q2: Age during the first birth of the child

Length: width:1; decimal: 0, negative:1

Range of Valid Data Values: -1-2

Universe: All participants (Women)

Text: What was your age when you gave birth to your first child?

Value	Label
0	Under 20
1	20-30
2	Above 30
-1	Do not have a child

Description*: Women who have their first child after 35 years are more prone to risk of breast cancer when compared to women who give birth to child when they are under 20.

Q3: Relatives in the family having Breast Cancer

Length: width:1; decimal: 0

Range of Valid Data Values: 0-5

Universe: All participants (Women)

Text: If any women in your family have had breast cancer, check their relationship to you (check all that apply)

Value	Label
0	None (No one had cases of breast cancer in the family)
1	Mother
2	Sister
3	Daughter
4	Grandmother
5	Aunt (mother's or father's sister)

Description*: If first degree relatives have been diagnosed with breast cancer then it increases the risk in Women.

Q4:Number of first-degree relatives having breast cancer

Length: width:1; decimal: 0,negative number:1

Range of Valid Data Values: -1-2

Universe: All participants (Women)

Text: How many of your first-degree relatives (mother, sisters, daughters) have had breast cancer?

Value	Label
-1	None (No one had cases of breast cancer in the family)
0	One
1	More than 1
2	Unknown

Description*: The risk is doubled even if the woman has any one first degree relatives with breast cancer

Q5: If their first-degree relatives were diagnosed with breast cancer before the age of 50?

Length: width:1; decimal: 0

Range of Valid Data Values: 1-2

Universe: All participants (Women)

Text: Was your mother, sibling, or child diagnosed with breast cancer before the age of 50?

Value	Label
1	Yes
2	No

Description*: The risk remains the same as others. This does not add much to the risk of the participant.

Q6: Uterine Cancer

Length: width:1; decimal: 0

Range of Valid Data Values: 1-2

Universe: All participants (Women)

Text: Have you had or do you have uterine cancer?

Value	Label
1	Yes
2	No

Description*: Women will have increased risk of uterine cancer if they had breast cancer in the past.

Q7: Ovarian Cancer

Length: width:1; decimal: 0

Range of Valid Data Values: 1-2

Universe: All participants (Women)

Text: Was your mother, sibling, or child diagnosed with breast cancer before the age of 50?

Value	Label
1	Yes
2	No

Description*: Women will have increased risk of ovarine cancer if they had breast cancer in the past.

Q8:Dense Breasts

Length: width:1; decimal: 0

Range of Valid Data Values: 1-2

Universe: All participants (Women)

Text: Has a mammogram shown that you have dense breasts?

Value	Label
1	Yes
2	No

Description*: X-ray picture of breast is represented using the Mammogram. Women with dense breasts with no other risk factors for breast cancer, are considered to have a higher risk of breast cancer than average.

Q9: Intake of diethylstilbestrol tablet

Length: width:1; decimal: 0

Range of Valid Data Values: 1-2

Universe: All participants (Women)

Text: Have you taken the medicine diethylstilbestrol (DES), or did your mother take it while pregnant with you?

Value	Label
1	Yes
2	No

Description*: Diethylstilbestrol (DES) represents female hormone estrogen. It was prescribed to pregnant women to prevent miscarriage, premature labor, and related complications of pregnancy. Daughters of mothers who took this tablet are prone to twice the risk than average. This usually happens when the daughters reach the age of 40.

Q10: Intake of hormone-based birth control pills

Length: width:1; decimal: 0

Range of Valid Data Values: 1-2

Universe: All participants (Women)

Text: Are you currently taking hormone-based birth control, such as birth control pills?

Value	Label
1	Yes
2	No

Description*: Usage of the birth control pills is linked to a slight increase in the risk of breast cancer.

Q11: Breast Biopsy (Benign one)

Length: width:1; decimal: 0,negative_number=1

Range of Valid Data Values: -1-1

Universe: All participants (Women)

Text: Have you ever had a breast biopsy with a benign (not cancer) diagnosis?

Value	Label
-1	Unknown
0	No
1	Yes

Description*: Benign growths are unusual growths or other changes in the breast tissue that are not cancer. Based on the histological abnormality in the benign biopsy, women can be checked to see if they have higher risk of breast cancer.

Q12: Number of Breast Biopsy

Length: width:1; decimal: 0,negative_number=1

Range of Valid Data Values: -2-2

Universe: Only Women who have undergone benign biopsies (does not include women who have answered "No" to the question

Of benign breast biopsies)

Text: How many breast biopsies with a benign diagnosis have you had?

Value	Label
0	None
1	1
2	More than one
-1	Unknown

Description*: Number of benign breast biopsies undergone is counted here.

Q13: Breast Biopsy with Atypical Hyperplasia

Length: width:1; decimal: 0,negative_number=1

Range of Valid Data Values: -2-2

Universe: : Only Women who have undergone benign biopsies (does not include women who have answered "No" to the question of benign breast biopsies)

Text: How many breast biopsies with a benign diagnosis have you had?

Value	Label
-1	Unknown
0	No
1	Yes

Description*: Atypical Hyperplasia is a precancerous condition that affects the cells in the breast. Women diagnosed with Atypical hyperplasia have been found to have higher risk of breast cancer.

Label: Final Label

Length: width:1; decimal: 0

Range of Valid Data Values: 0-2

Universe: All participants (Women)

Text: Final Categorization of Instances

Value	Label
0	Low

1	Moderate
2	High

Description*: The cases obtained are categorized into Low, Moderate and High with assistance of the domain expert.