Original Article

Access this article online

Quick Response Code:

Website: www.jehp.net

DOI: 10.4103/jehp.jehp_384_18

# Developing a community-based breast cancer risk prediction tool for resource-poor settings

**Divya Pillai, Shaikh Shah Hossain[1], Vijay Kumar Chattu[2,3]**

*Department of Clinical Portfolio Management, Clinical Development Services Agency, Translational Health Science and Technology Institute, Faridabad, Haryana, [1]Department of Public Health, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka, [3]Department of Public Health Research, Global Institute of Public Halth, Thiruvananthapuram, Kerala, India, [2]Department of Paraclinical Sciences, Faculty of Medical Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago*

**Address for correspondence:**
Dr. Vijay Kumar Chattu, Faculty of Medical Sciences, University of the West Indies, St. Augustine, Trinidad and Tobago.
E-mail: drvkumar.ch@gmail.com

Received: 13-11-2018
Accepted: 31-12-2018

**Abstract:**

**BACKGROUND:** With an estimation of every two women newly diagnosed with breast cancer, one dies. It is accounted that 1 in 28 women is likely to develop breast cancer during her lifetime. Developing a risk prediction tool by assessing the prevalence of known risk factors in the community will help public health intervention.

**METHODOLOGY:** A cross-sectional study was conducted among 18–64-year-old women to gather the prevalence of known breast cancer risk factors, through a community survey (sample survey). In this multistage random number-based cluster sampling study, the results were compiled, collated, and analyzed in rates and proportions. Statistical conclusions were made using spreadsheets (Microsoft) and the values were converted into ordinal values using modified Likert scale and median was used to estimate central values. The estimated prevalence of these known risk factors was re-assorted for analysis and these re-assorted data were categorized into range of values across the communities. The internal validity of the survey questionnaire was measured using Cronbach's alpha ($\alpha$).

**RESULTS:** The analysis of 558 participants was performed for the known risk factors for breast cancer including participant's age, age at menarche, marriage, first childbirth, menopause, family history of breast cancer and benign breast disease, history of abortion, and body mass index. Based on the estimated prevalence of these risk factors, a community-based risk prediction tool was developed with Cronbach's $\alpha$ score of medium internal validity.

**CONCLUSIONS:** The risk assessment tool has collated most of the risk factors of breast cancer that are capable of being measured at community level. The survey findings concluded that the community under survey was bearing moderate risk for breast cancer for women.

**Keywords:**

BRCA1, breast cancer, cluster sampling, Gail Model, risk assessment tool, risk factors, screening

## Introduction

Breast cancer is ranked as number one cancer among women in India, and the age-standardized incidence rate was found to be 25.8/100,000 females.[1-4] However, the role of various risk factors leading to breast cancer is not adequately understood in our country setting. Based on the breast cancer risk factor profiling, countries have developed their own risk prediction tools or models and these risk assessment models are established and validated based on the known risk factors for breast cancer in white women. The Breast Cancer Risk Assessment Tool developed by the United States National Institute of Health for Cancer is based on inputs over medical history of breast cancer or precancerous conditions including age of individual, age at menarche, age at first childbirth, family history of breast cancer, biopsy findings, and ethnicity.[5] The Gail Model has been modified to account for the history of atypia and race or ethnicity, but until now, it has included only nonmodifiable risk factors (age, reproductive history, biopsy



**How to cite this article:** Pillai D, Hossain SS, Chattu VK. Developing a community-based breast cancer risk prediction tool for resource-poor settings. J Edu Health Promot 2019;8:106.

history, etc.,) among white American women. Barlow in 2006 constructed separate risk models which are slightly more complex than the Gail Model and are based on breast density as an indicator or breast cancer risk.[6-9]

Some of these models and similar instruments have been tested on Asian population; however, ethnic and geographic differentials make them less predictive. The physical examination of the breast by paramedical workers has also been evaluated as an alternative screening tool. A regular self-breast examination is also helping the woman in detecting the onset of breast cancer. Detecting cancer, when it is in the very early stage, can improve the cure rate from breast cancer. For effective breast cancer prevention, the assessment of community risk is of critical importance so that a targeted intervention is possible, optimizing resource utilization and yielding best results. Risk communication combined with suitable and customized intervention is the key to a successful anticancer campaign. The current study aims at developing a questionnaire-based risk assessment score for breast cancer in the community.

## Methodology

A cross-sectional survey with multistage random number-based cluster sampling technique was used for validating the usefulness of the instrument. A study population considered for the interview was women of 18–64 years. The sample size was calculated as per the World Health Organization STEPS Manual for Noncommunicable Disease Risk Factors Survey. A target sample size was taken at 500. With a 10% nonresponse, the actual target was fixed at 550.[10] The eligible participants were identified from each randomly selected panchayats and municipalities by the study investigator with the help of junior public health nurses at subcenter or primary health-care centers. The study was conducted with the objective of developing a risk-based assessment score on the grounds of the prevalence of known risk factors of breast cancer in the community.

The risk factors identified for the assessment of breast cancer risk factors in the community were as follows: (1) prevalence of ≥40 years, (2) percentage of participants married at 30 years and above, (3) percentage of participants with age at menarche below 13 years, (4) percentage of participants by age at first childbirth, (5) percentage of participants with age at menopause of 49 or above among all menopausal participants, (6) percentage of abortions among ever-married women, (7) average fat consumption/week in KCal, (8) percentage of body mass index 25 and above, (9) percentage of breast diseases (benign and malignant) cases, and (10) percentage of family history of breast cancer.

Each participant was explained the purpose of the study. Written consent was obtained from the study participants. The presented study proposal was approved by the Institutional Ethics Committee of the National Centre for Disease Control, New Delhi.

### Data analysis

The values were obtained in different scales/units for each risk factor considered for the study. Age was measured in completed years, fat consumption in grams per week, etc., These quantitative values were converted into ordinal values for comparison. To facilitate a composite estimation of risk, these values were classified in 5-point modified Likert scale from minimum to maximum value.[10]

Results were compiled, collated, and analyzed in rates and proportions. For comparing the proportions obtained in the current study with other studies StataV9, statistical software was used. Test of significance to measure internal validity of scoring by modified Likert scale was Cronbach's alpha (α). The equation used was as follows: α Cronbach's = $(k/k\text{-}1) \times \{1-\Sigma S_i^2/S_{Total}^2\}$[11,12] where,

- K = number of items in the scored questionnaire
- $S_i$ = standard deviation of scores for each item
- $S_{Total}$ = standard deviation of total score of all the items.

## Results

### The risk assessment score

A total of 564 participants in the age group of 18–64 years from 345 households were approached for the survey. The analysis of 558 participants was performed excluding six incomplete forms. The known risk factors for breast cancer included history of reproductive factors, nutritional intake, and anthropometry.

The prevalence of these known risk factors of breast cancer estimated was re-assorted for analysis of in the study community. These re-assorted data are categorized into range of values across the communities [Table 1].

On the basis of this range of values, a 5-point modified Likert scale has been prepared to capture lowest to highest value. The scale is given with a scoring key for each risk factor. With the Likert scale score for each item reflected, statistical test was applied to check validity and robustness of the scale. The composite scores obtained have been reproduced in Table 2 which summarizes the scoring scale for assessment of breast cancer risk factors in the community.

The test of significance used in this study is Cronbach's α or α Cronbach's. Table 3 gives Cronbach's α value for breast cancer risk factor survey and its basis.

**Table 1: Re-assorted prevalence of risk factors of breast cancer in the communities**

| Range | Age | Age at | | | | Abortions | Average fat intake/week in KCal | BMI | Breast disease | Family history of Breast Cancer |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Marriage (≥30 years) | Menarche (<13 years) | First child ≥30 years | Menopause (≥49 years) | | | | | |
| Highest | 62 | 11.1 | 83.3 | 11.8 | 87.5 | 26.7 | 2988 | 40.4 | 10 | 8.8 |
| Median | 40.5 | 2.65 | 46.6 | 4.6 | 66.7 | 18.6 | 2320 | 14.3 | 3.35 | 2 |
| Lowest | 28.1 | 0 | 0 | 0 | 37.5 | 7.7 | 1840 | 7.3 | 0 | 0 |
| Score key for the risk factors | | | | | | | | | | |
| 1 | 20-29 | <2 | <30 | <2 | <40 | <10 | <2000 | <15 | <2 | <2 |
| 2 | 30-39 | 2,3 | 30-39 | 2,3 | 40-54 | 10-14 | 2000-2199 | 15-19 | 2 | 2-3 |
| 3 | 40-49 | 4,5 | 40-49 | 4,5 | 55-69 | 15-19 | 2200-2399 | 20-24 | 3 | 4-5 |
| 4 | 50-59 | 6,7 | 50-59 | 6,7 | 70-84 | 20-24 | 2400-2699 | 25-29 | 4 | 6-7 |
| 5 | 60-69 | ≥8 | ≥60 | ≥8 | ≥85 | ≥25 | ≥2700 | ≥30 | ≥5 | ≥8 |

BMI=Body mass index

**Table 2: Composite scores for the 12 study communities for calculating Cronbach's Alpha**

| Risk factors | Community (C) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 |
| Percentage age ≥40 | 2 | 3 | 2 | 3 | 5 | 3 | 2 | 3 | 2 | 3 | 2 | 1 |
| Percentage age at marriage ≥30 | 3 | 2 | 2 | 2 | 2 | 5 | 2 | 5 | 1 | 1 | 1 | 1 |
| Percentage age at menarche <13 | 4 | 4 | 1 | 3 | 2 | 3 | 3 | 2 | 5 | 3 | 4 | 3 |
| Percentage age at first childbirth ≥30 | 3 | 4 | 1 | 3 | 5 | 4 | 2 | 5 | 1 | 2 | 1 | 4 |
| Percentage age at menopause ≥49 | 4 | 5 | 3 | 4 | 3 | 5 | 3 | 3 | 3 | 3 | 4 | 1 |
| Percentage abortions | 3 | 2 | 4 | 1 | 5 | 3 | 4 | 2 | 3 | 4 | 5 | 3 |
| Average fat intake in KCal | 3 | 5 | 2 | 5 | 3 | 2 | 3 | 3 | 1 | 4 | 3 | 4 |
| Percentage BMI ≥25 | 5 | 4 | 1 | 1 | 5 | 1 | 1 | 1 | 1 | 2 | 1 | 5 |
| Percentage breast diseases | 4 | 3 | 1 | 1 | 5 | 1 | 4 | 2 | 2 | 5 | 5 | 3 |
| Percentage family history of BrCa | 3 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 4 | 1 | 5 |

**Table 3: Cronbach's alpha for breast cancer risk factor survey**

| Component | Value | Computation |
|---|---|---|
| K | 10 | |
| K-1 | | 9 |
| (K/K-1) | | 1.111111 |
| $\sum Si^2$ | 18.01515 | |
| $\sum S_{total}^2$ | 30.51515 | |
| $\sum Si^2/\sum S_{total}^2$ | | 0.590367 |
| $1 - (\sum Si^2/\sum Stotal^2)$ | | 0.409633 |
| $\alpha_{Cronbach's}$ | 0.455147 or 46% | |

## Discussion

According to the recent 2017 estimates of the Global Burden of Disease study, the prevalence of breast cancer globally is around 16,697,300 globally [Table 4].[13] There are various behavioral, environmental, metabolic, and occupational risk factors that contribute to the growing burden (disability-adjusted life years) of breast cancer[14,15] as shown in Table 5. The prevalence of risk factors of breast cancer in the current study gives a cross-sectional picture of the community. The number of first-degree relatives diagnosed with breast cancer is one of the major risk factors which is also included in the Gail Model Breast Cancer Risk Assessment Tool,[7,16] and these findings were also validated from other studies.[17-19] The scoring system used in Likert scale is based on the actual range of values on the ground and validates the relevance of known risk factors of breast cancer as observed in studies conducted abroad in Indian conditions. Internal validity of this tool is reflected in the Cronbach's α value of 45.5, which is moderately good.

Inadequate knowledge of breast cancer risk factors and poor cancer screening practices for breast cancer among female patients makes it difficult to identify high-risk groups. These high-risk groups should be educated and requires careful follow-ups with good screening modalities. Furthermore, improved ways of follow-up using identification of various putative genes such as BRCA-1 and BRCA-2 can also be employed. Hence, the study of family history in a population helps in screening of breast cancer and predisposition to breast cancer.[20] The Epidemiology and Genomics Research Program under the Division of Cancer Control and Population Sciences in National Cancer Institute discussed the breast cancer risk prediction models in detail. The available models are categorized to the Absolute Risk Prediction Models, Gene Carrier Status Risk Prediction Models, Risk Prediction Models of Women at High Risk, and Online Risk Assessment Tools and Calculators.

**Table 4: Global burden, years lived with disability, and disability-adjusted life years of breast cancer**

| Prevalence (thousands) 2017 counts | Incidence (thousands) 2017 counts | YLDs (thousands) 2017 counts | Percentage change in YLDs counts, 1990-2007 | Percentage change in YLDs counts, 2007-2017 | All age- DALYs (thousands) 2017 | Percentage change in DALYs 2007-2017 |
|---|---|---|---|---|---|---|
| 16,697.3 (16178.9-17171·7) | 1960.7 (1891.4-2023.2) | 1307.9 (932.0-1769.3) | 67.7% (60.6-72.7)* | 33.8% (29.0-37.7)* | 17400 (16900-18700) | 24.6% (18.1-29.3)* |

*Statistically significant increase. YLDs=Years lived with disability, DALYs=Disability-adjusted life years

**Table 5: Global estimation of various risk factors contributing to breast cancer**

| Behavioral/environmental/ occupational risk factor | 2007 deaths (thousands) | 2017 deaths (thousands) | Percentage change in deaths, 2007-20117 | 2007 DALYs | 2017 DALYs | Percentage change in DALYs, 2007-2017 |
|---|---|---|---|---|---|---|
| Tobacco smoking | 16 (12-21) | 17 (12-22) | 5.2% (1.9-8.4)* | 447 (317-582) | 451 (320-593) | 0.8% (–2.5-3.8)* |
| Second hand smoke | 12 (3-21) | 15 (4-26) | 21.2% (14.3-26.9)* | 404 (98-688) | 478 (118-829) | 18.5% (11.6-24.6)* |
| Alcohol use | 54 (46-62) | 59 (49-69) | 9.7% (4.3-14.1)* | 1560 (1320-1790) | 1670 (1390-1940) | 7.0% (2.0-11.3)* |
| Low physical activity | 7 (0-17) | 9 (0-21) | 25.9% (20.8-30.5)* | 205 (4-470) | 253 (5-575) | 23.4% (17.7-28.5)* |
| High fasting plasma glucose | 33 (6-73) | 44 (9-99) | 35.2% (27.9-43.9)* | 805 (153-1800) | 1070 (205-2430) | 33.5% (25.4-42.4)* |
| High body mass index | 26 (10-48) | 40 (16-71) | 54.8% (39.2-88.4)* | 487 (125-983) | 817 (267-1530) | 67.7% (45.5-153.1)* |

*Statistically significant increase. DALYs=Disability-adjusted life years

These statistical models estimate the probability of developing cancer over a defined period of time. This will help clinicians identify individuals at higher risk of specific cancers. This will further, encourage them in imparting routine screening and genetic counseling as best health-seeking behaviors among the potential patients leading to the reduction in the burden of the disease. These types of models also will be useful for designing future chemoprevention and screening intervention trials in individuals at high risk of specific cancers in the general population.[21]

The most common model used to predict woman's risk of developing breast cancer is Gail Model and it is useful in targeting screening and prevention efforts. The multivariate risk model developed by Gail *et al*. is widely used for quantifying the risk of breast cancer within a specified time in women at various ages and with certain risk factors. A case–control study was conducted at Regional Cancer Center, Thiruvananthapuram, to determine the usefulness of Gail Model Breast Cancer Risk Assessment Tool in identifying women at high risk for breast cancer proved that Gail Model cannot be used to predict high-risk women in Kerala.[22] Numerous mathematical models were also made using these risk factors to predict individual women's risk of developing breast cancer. There is a paucity of data regarding the risk factors in our society.

## Conclusions

The Cronbach's α value with the survey conducted with these items was 45.5 indicating good internal validity. The risk assessment tool was found suitable for use in the current study field situation. The current study design has the scope of being up scaled for use in a larger study that may lead to a more internally valid reflection of risk in the population. The risk assessment tool has collated most of the risk factors of breast cancer that are capable of being measured at community level. Based on the experience of this study, a risk assessment tool with scoring system has also been evolved. This may be used for community surveys for risk assessment, and based on the Cronbach's α value of <40, 40–69, and ≥70, the validity of such assessment results may be taken as low, moderate, or high.

The known risk factors of breast cancer validated by international studies were used for assessment of risk factor prevalence of breast cancer in this study. This has given a fair idea of prevalence of these risk factors in India. This list of 10 risk factors for the assessment of community risk for breast cancer can be used in other surveys of similar kind. It is recommended that this composite comprehensive risk assessment tool for risk assessment of breast cancer in the community as a composite instrument with the 10 items with given scores. A larger study using the instrument shall lead to refinement and piloting of an important community-based risk assessment tool.

## Acknowledgments

The study was fully supported by the National Centre for Disease Control and Directorate of Health Services, Government of Kerala, India. The authors would like to thank Dr. K. Suresh Kishanrao, Dr. Manish Kakkar, and Dr. Dhananjay Srivastava for their valuable guidance to the study. We thank all the study participants who made this study successful through their active participation and contribution.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## References


1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, *et al*. GLOBOCAN 2012 v10. Cancer incidence and mortality worldwide: IARC CancerBase No. 11. 2012.
2. Breast Cancer Estimated Incidence, Mortality and Prevalence Worldwide; 2012. Available from: http://www.globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. [Last accessed on 2016 Jun 07].
3. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer 2013;132:1133-45.
4. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev 2010;19:1893-907.
5. Breast Cancer Risk Assessment Tool. National Cancer Institute, USA. Available from: http://www.cancer.gov/bcrisktool/. [Last accessed on 2018 Oct 07].
6. Breast Cancer Risk Calculator. National Cancer Institute, USA. Available from: https://www.bcrisktool.cancer.gov/about.html#OtherRiskAssessmentTools. [Last accessed on 2018 Oct 07].
7. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, *et al*. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879-86.
8. Bondy ML, Newman LA. Assessing Breast Cancer Risk: Evolution of the Gail Model. Available from: http://www.jnci.oxfordjournals.org/cgi/content/full/98/17/1172. [Last accessed on 2018 Sep 24].
9. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, *et al*. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006;98:1204-14.
10. World Health Organization. WHO STEPS Surveillance Manual: The WHO STEPwise Approach to Chronic Disease Risk Factor Surveillance. Geneva: World Health Organization; 2005.
11. Daly LE, Bourke GJ, Mcgilvray J. Interpretation and Uses of Medical Statistics. Oxford: Blackwell Scientific Publications; 1991. p. 104-6.
12. Campbell MJ, Machin D. Medical Statistics – A Commonsense Approach. 3rd ed. England: John Wiley & Sons Ltd.; 1999.
13. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789-858.
14. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1923-94.
15. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1859-922.
16. Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, *et al.* Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst 2011;103:951-61.
17. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358:1389-99.
18. Hirose K, Tajima K, Hamajima N, Inoue M, Takezaki T, Kuroishi T, *et al.* A large-scale, hospital-based case-control study of risk factors of breast cancer according to menopausal status. Jpn J Cancer Res 1995;86:146-54.
19. Lech R, Przemysław O. Epidemiological models for breast cancer risk estimation. Ginekol Pol 2011;82:451-4.
20. Varughese AA, Poothiode U, Manjula VD. Descriptive study on selected risk factors and histopathology of breast carcinoma in a tertiary care centre in Kerala, India with special reference to women under 40 years old. Asian Pac J Cancer Prev 2015;16:181-4.
21. Breast Cancer Risk Prediction Models, Division of Cancer Control and Population Sciences, National Cancer Institute, USA. Available from: https://www.epi.grants.cancer.gov/cancer_risk_prediction/breast.html. [Last accessed on 2018 Oct 07].
22. Augustine P, Jose R, Peter A, Lal AA, Prabhakar J, Sreedharan J, *et al.* Risk factors of breast cancer in Kerala, India – A case control study. Acad Med J India 2014;2:7-13. Available from: http://www.medicaljournal.in/volume2-issue1-feb-2014-59-original-research-risk-factors-breast-cancer-kerala-india-case-control-study/. [Last accessed on 2018 Sep 12].