Breast Cancer Decision Support System for Rural People

Thoranin Intarajak and Seung Hwan Kang
Computer Information Systems, Payap University, Chiang Mai, 50000 Thailand
Thoranin_i@payap.ac.th, Seung_h@payap.ac.th

***Abstract*- In Thailand, breast cancer is one of the top five cancers and the percentage of breast cancer patients is surprisingly increasing. Siriraj Hospital has collected data of breast cancer patients for more than ten years. Unfortunately, those data were not arranged in a useable form. Therefore, a system to record breast cancer information and treatment methods has been created. It enables medical users to rapidly diagnose and treat breast cancer patients. The scope of this system and the database is that it can record breast cancer information and recommend the treatment method for breast cancer patients. The system uses the data of the Siriraj Hospital algorithm for classifying the cancer stage, which is a modification of the National Comprehensive Cancer Network technique in order to find the treatment method and collect survival rates after treatment.**

***Keywords*- Decision support system, Rural decision support system, Breast cancer**

I. INTRODUCTION

Cancers develop as a result of cells dividing uncontrollably which is caused by a complex mix of factors related to environment, lifestyle, and genetics. Eventually, altered genes and uncontrolled growth may produce a tumor that can be benign or malignant. Malignant tumors can invade, damage, and destroy nearby tissues and spread to other parts of the body. A benign tumor won't spread to other parts of the body, but local tissue may be damaged and the growth may need to be removed.

Breast cells contain a variety of genes that normally work cooperatively with a woman's natural hormones, diet, family history, age, gender, menopause, first birth, depression, and environment to keep her breasts healthy [1]. Certain genes routinely maintain breast cells from dividing and growing out of control and forming tumors. Breast cancer usually occurs at single cells that change from normal cells to malignant over a period of time. Presently, no one can predict exactly when cancer will occur or how it will progress when breast cancer is diagnosed, even if detected at the earliest stage [1]. In Thailand, breast cancer is one of the top five cancers and is increasing very fast. Siriraj Hospital is one hospital in Thailand that has gathered data from patients for more than 10 years. Unfortunately breast cancer databases have redundancies, complexes, and constrains. Siriraj Hospital wanted to analyse risk factor, and investigate treatment methods. The National Comprehensive Cancer Network (NCCN)[2] in the USA is dedicated to improving the quality and usefulness of care provided to patients with cancer. NCCN develops resources that present precious information to the many stakeholders in the health care delivery system. NCCN promotes the important of continuous quality improvement and recognizes the significance of creating clinical practice guidelines appropriate for use by patients, clinicians, and other health care decision-makers. The NCCN provides an efficient method for deciding on the treatment of breast cancer. This method is a very effective way to classify the stages and uses neoadjuvant to recommend treatment of breast cancer patients. However Siriraj Hospital has its own system, which differs in detail from the NCCN algorithm. Therefore this work developed, a web application using the Siriraj algorithm to classify breast cancer stage and find a suitable method to treat breast cancer patients. The database has been created to eliminate complex and redundant patient information and make it easy to add and update data. This system helps physicians to classify the state of breast cancer which allows a doctor to treat patients in the right way. An algorithm for doing Cox regression, to forecast probability of survival has also been included so that when historical data has been entered into the database, analysis of risk factors for the Thai context can be investigated.

II. RELATED WORKS

Nowadays breast cancer is one of the top five cancers in the world. There are many risk factors that corresponding with this disease such as age of menarche, family history, genes, hormone, previous biopsy, age of menopause, etc but in Thailand there is a very little of research about how to find risk factors for Thai people, Anne Stotter, et al [3], found the way to improve data collection on breast cancer incidence and survival by using Cox model. They suggest that age was important to predict survival. Vincent Vinh-Hung, et al [4] studied the number of lymph nodes as a risk factor and using cox model and they found that the number of lymph nods is a risk factor for breast cancer. In the decision part, this system uses decision support system concept to make a recommendation to the physician. Curtis H.K. Tsang and Chris Bloor [5] studied a genetic frame work for complex medical expert system interface by use PC-Kappa as a system development environment. Bridgett, N. A., Brandt, J. and Harris studied [6] in Associative Memory Neural Network to increase accuracy of diagnosis and treatment of breast cancer but it is only a concept not implemented. Genetic algorithm Elisabet Golobardes, et.al [7] used case based reasoning and genetic algorithms to diagnose breast cancer by using mammography image but human expert is better than this system. Phayung Meesad and Garry G. Yen [8] used incremental leaning fuzzy neural network algorithm to classify breast cancer in mammography images. This system can classify breast cancer better than a human expert. Paul A.Fearn, et al [9] used web application and database to build the reusable clinical database for predictive modeling and for coming wave of translational laboratory research and now this system cover prostate cancer.

III. RISK FACTORS

Breast cancer is the most common malignancy affecting women in North America and Europe. Every woman is at risk of breast cancer. Close to 200,000 cases of breast cancer were diagnosed in the United States in 2001. Breast cancer is the second leading cause of cancer death in American women behind lung cancer. The lifetime risk of any particular woman getting breast cancer is about 1 in 8 although the lifetime risk of dying from breast cancer is much lower at 1 in 28. Risk factors for breast cancer can be divided into those that you cannot change and those that can be changed. Some factors that increase the risk of breast cancer that cannot be altered include being a woman, getting older, having a previous history of breast cancer, having had radiation therapy to the chest region, education, age at menarche, parity nulliparae, menopausal status, age at menopause, body mass index, total calorie intake, [10,11,12] not only being Caucasian, starting periods young, having menopause late, never having children or having them when over 30, reproductive life, hormonal factors [2], and having a genetic mutation that increases your risk, also people from metropolis comparison with developing country. The researcher found that people from metropolis have more risk than developing country because of not only family history of breast cancer, nulliparity, first pregnancy at late age, early age at menarche and menopause, obesity in post-menopausal women, previous breast disease and a genetic disposition. Also depression is another risk factor of breast cancer [13]. Genetic mutations for breast cancer have become a hot topic of research lately. Between 3% to 10% of breast cancers may be related to changes in either the gene BRCA1 or the gene BRCA2. Women can inherit these mutations from their parents and it may be worth testing for either mutation if a woman has a particularly strong family history of breast cancer. If a woman is found to carry either mutation, she has a 50% chance of getting breast cancer before she is 70. Family members may elect to get tested to see if they carry the mutation as well. If a woman does have the mutation, she can get more rigorous screening or even undergo preventive mastectomies to decrease her chances of contracting cancer.

IV. SCREENING TEST FOR BREAST CANCER

Screening for breast cancer uses mammograms, clinical breast exams, and breast self-exams. Woman between the ages of 20 and 39 should have a clinical breast exam every 3 years; and after the age of 40 every woman should have a clinical breast exam done each year. Specific patterns of breast cancer consist of 4 stages and 2 specific patterns:

Stage 0 : Very early breast cancer. Cancer has not spread within or outside the breast so called DSIS, LCIS, or breast cancer in situ or noninvasive cancer.

Stage I :Cancer is no larger than about 1 inch in size and has not spread outside the breast. (Also described as early breast cancer.)

Stage II :

- The cancer is no larger than 1 inch, but has spread to the lymph nodes under the arm.
- The cancer is between 1 and 2 inches. It may or may not have spread to the lymph nodes under the arm.
- The cancer is larger than 2 inches, but has not spread to the lymph nodes under the arm.

Stage III Stage III is divided into stages IIIA and stages IIIB:

Stage IIIA doctor may find either of the following:

- The cancer is smaller than 2 inches and [14]has spread to the lymph nodes under the arm. The cancer also is spreading further to other lymph nodes.
- The cancer is larger than 2 inches and has spread to the lymph nodes under the arm.

Stage IIIB doctor may find either of the following:

- The cancer has spread to tissues near the breast (skin, chest wall, including the ribs and the muscles in the chest).
- The cancer has spread to lymph nodes inside the chest wall along the breast bone.

Stage IV The cancer has spread to other parts of the body, most often the bones, lungs, liver, or brain. Or, the tumor has spread locally to the skin and lymph nodes inside the neck, near the collarbone.

V. TREATMENT FOR BREAST CANCER

Surgery

The purpose of this method is to eliminate as much of the cancer as possible. Some women will be candidates for surgery that called breast conservation therapy (BCT). BCT constantly needs to be combined with radiation therapy to make it an option for treating breast cancer. In early stage cancers (stage I and stage II), BCT is as effective as removal of the entire breast via mastectomy. Most patients with DCIS that have a lumpectomy are treated with radiation therapy to prevent the local recurrence of DCIS [10].

Chemotherapy

Many breast cancer patients are offered chemotherapy, in order to decrease a patient's risk of recurrence after surgery. Chemotherapy using anti-cancer drugs which treat the entire body. Sometimes patients have a recurrence of cancer, or present in stage IV with disease outside the breast. These patients need chemotherapy, and a multiplicity of different agents may be tried until a response is obtained. Sometimes doctor give chemotherapy before surgery. This is called neoadjuvant chemotherapy. This is usually reserved for very advanced cancers that need to be shrunk before the doctor can be operated on it [10].

Radiotherapy

Radiation therapy is commonly used which uses high energy rays (similar to x-rays) to destroy cancer cells. The treatment takes just a few minutes, and it is painless. Radiation therapy is used in patients who receive breast conservation therapy (BCT). It is also recommended for patients after a mastectomy on a patient who had a large tumor, lymph node involvement, or close/positive margins [10].

Hormonal Therapy

Patients who display tumors expressed estrogen receptors are recommended for therapy with a estrogen blocking drug called Tamoxifen. This drug has been shown to significantly reduce risk of recurrence [10].

Nottingham Prognostic Index(NPI)

The NPI status was 1st published about 1992 as an attempt at using some fairly objective parameters to determine the odds that a newly diagnosed case of invasive ductal adenocarcinoma would benefit from adjuvant chemotherapy. By means of, the status of the axillary lymph nodes must be determined at least as to whether any nodes are positive for metastatic cancer (ALN+) or negative (ALN-). If there are any positive nodes, more details are needed. Decisions may have to be made prior to having the whole tumor out and size-measurable. By whatever measure, the following are the criteria for assignment of size points [15].

NPI = (0.2×size of tumor(cm)) + grade point + lymph node point

Grade point score

Negative node	= 1 point
Positive nodes, low axillary only	= 2 point
Positive nodes, internal mammary only	= 2 point
Positive nodes, axillary plus internal mammary	= 3 point
Positive nodes, apical axillary node positivity	= 3 point

Lymph node score

Stage A denoted no involvement of regional node	= 1 point
Stage B denoted involvement of ≤ 3 axillary node or involvement of internal mammary node	= 2 point
Stage C denote > 3 axillary node or both internal mammary and axillary node involvement	= 3 point
NPI < 3.4	= doubtful
$3.4 \leq$ NPI >5.4	= may benefit with chemotherapy
NPI $\geq$ 5.4	= Must have Chemotherapy

Van Nuys Prognostic Index (VNPI)

The original Van Nuys prognostic index (VNPI) was introduced in 1996 as an aid to the complex treatment decision-making process for patients with ductal carcinoma in situ (DCIS) of the breast. This update adds patient age to the previous predictors of local recurrence in breast preservation patients [16, 17].

VNPI Score = tumor size point + margin point + pathology point

Tumor

Tumor size 15	parameter = 1
Tumor size > 15 – 40	parameter = 2
Tumor size > 40	parameter = 3

Margin

Margin size 10 mm	parameter = 1
Margin size 2 – 9 mm	parameter = 2
Margin size < 1 mm	parameter = 3

Pathology

Non high grade no necrosis	parameter = 1

Non high grade with necrosis	parameter = 2
High grade with necrosis	parameter = 3
VNPI score 3 – 4	excision only
VNPI score 5,7, or 7	radiation therapy
VNPI score 8-9	mastectomy

VI. METHODOLOGY

As shown in Figure 1 , the Siriraj hospital decision algorithm system checks the values of certain decision variables in order to recommend treatment. In the Siriraj case these are size of tumor, size of breast, number of masses, radiation to chest wall, collagen vascular disease, and hormone receptor then recommend the most suitable treatment method. This decision algorithm has been written in a way to make it easy to modify to accommodate variations in the decision algorithms used as there could be variations between hospitals in their decision processes. It should be emphasised that the criteria used in this decision making are based on the information provided by the doctors who will be using the system. The IT expert is converting the doctors' expertise into a logical, comprehensive decision trees. It is not the IT expert trying to tell the doctor how to do their job. The risk factor analysis application uses Cox regression analysis to analyze the risk factors of breast cancer and the chances of survival[18]. The Cox model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. Survival analysis typically examines the relationship of the survival distribution to covariates. This examination entails the specification of a linear-like model for the log hazard

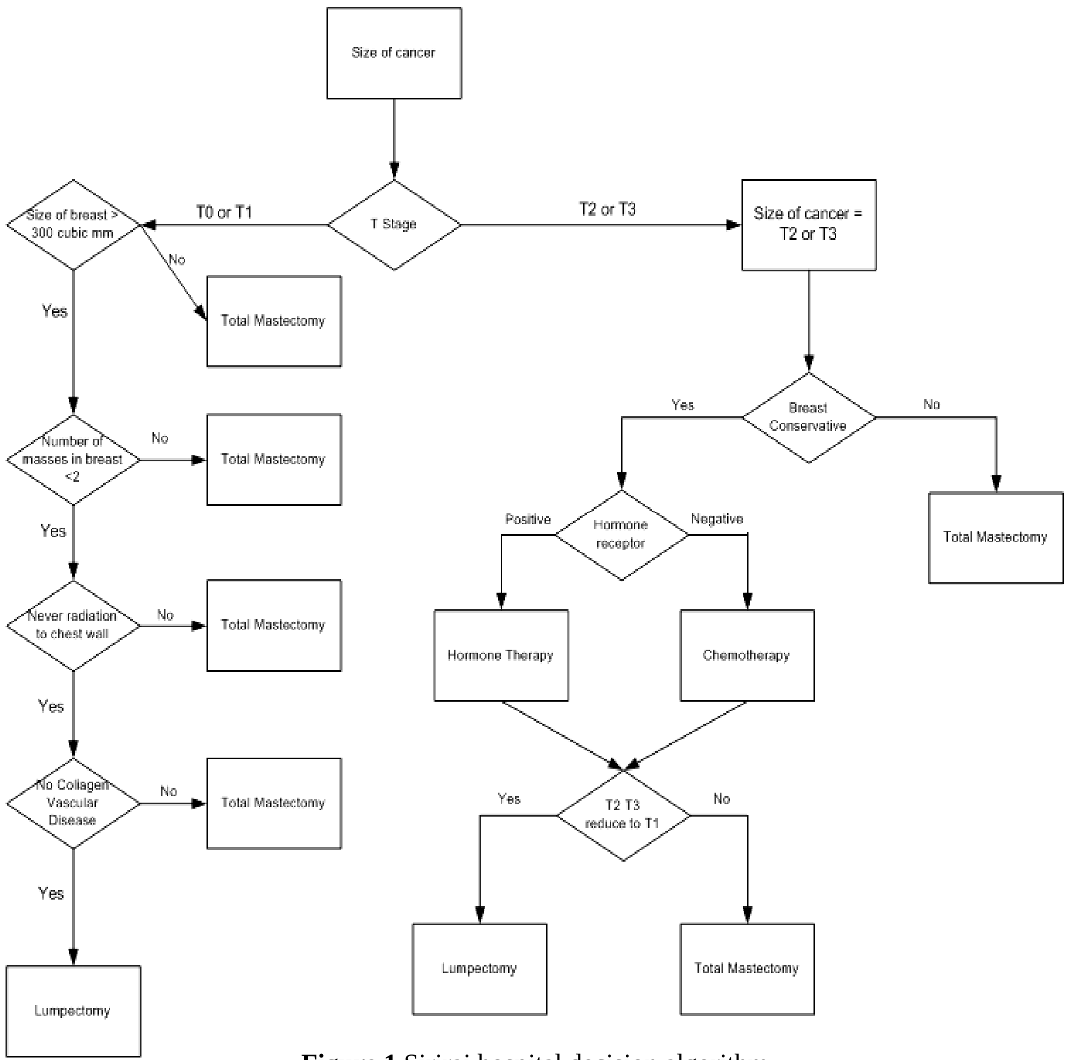

Figure 1 Siriraj hospital decision algorithm

In the risk factor analysis phase, this web application uses Cox regression analysis model to analyze the risk factors of breast cancer. The Cox model [19] is a statistical method for exploring the relationship between the survival of a patient and the variables. The cox model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. Survival analysis typically examines the relationship of the survival distribution to covariates. This examination entails the specification of a linear-like model for the log hazard. A parametric model based on the exponential distribution may be written as

or, equivalently,

$$h_i(t) = \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{ik} + \ldots. + \beta_k x_{ik})$$

The Cox model, in contrast, leaves the baseline hazard function

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{ik} + \ldots. + \beta_k x_{ik}$$

or, again equivalently,

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{ik} + \ldots. + \beta_k x_{ik})$$

This model is semi-parametric because while the baseline hazard can take any form, the covariates enter the model linear. Consider, now, two observations i and i' that differ in their x-values, with the corresponding linear predictors

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{ik} + \ldots. + \beta_k x_{ik}$$

and

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'k} + \ldots. + \beta_k x_{i'k}$$

The hazard ratio for these two observations, is given as

$$\frac{h_i(t)}{h_{i'}(t)} = \frac{h_0(t)e^{\eta_i}}{h_0(t)e^{\eta_{i'}}} = \frac{e^{\eta_i}}{e^{\eta_{i'}}}$$

h0(t) is the baseline hazard at time t, representing the hazard for a person with the value 0 for all the predictor variables.

VII. IMPLEMENTATION

The system architectures are show in Figure 2. The system starts when a user wishes to insert, or delete, or update data into database. Data are stored in a database and the user can print a report or send a signal to the analysis of risk factors routine. The output from the risk factor analysis is sent to another part of the decision support system to train the intelligent part. The final output will be the treatment recommendation or report. Figure 2 shows the different use cases. The main users are the doctors and the hospital administrative or nursing staff. Each user is uniquely identified. From Figure 2, the general data flow of this system is when users connect to the breast cancer system they must log in and get their level of information access. After that the system separates users in to 2 groups; the first one is doctor and administrator the other is other hospital staff. Doctors and system administrators have authority to insert, update, search, and analyze patient data. On the other hand other hospital staff can only insert breast cancer patient information and print reports.

The design of the database for this system took some effort due to the large number of fields to be stored for each patient visit as well as the variable number of visits and the fact that the record length for a visit depends on the type of treatment applied on that occasion. There is also a requirement that identifying personal information for patients should be only available to their doctor and that it be possible to display the system and print out patient record information with that identifying information not displayed. The confidentiality of patient information has to be adhered to in any electronic medical systems.

VIII. UTILIZATION

This system can eliminate time to diagnosis and treatment breast cancer patients in rural areas, categorize information for medical users to use in future, reduce time and cost in part of investigation for help people in rural country.

IX. SUMMARY

In Thailand Breast Cancer is one of top five cancers and its occurrence is increasing. Siriraj hospital, one of biggest hospitals in Thailand, has collected data from breast cancer patients. Unluckily the data of breast cancer are not arranged and analyzed. This work uses information technology to solve the problem to arrange data of breast cancer patients in rural country by use database technology, and in the analysis part it uses the Cox model to analyze the risk factors of breast cancer, and in the decision part it uses a decision support system to recommend suitable methods for treating breast cancer patients.

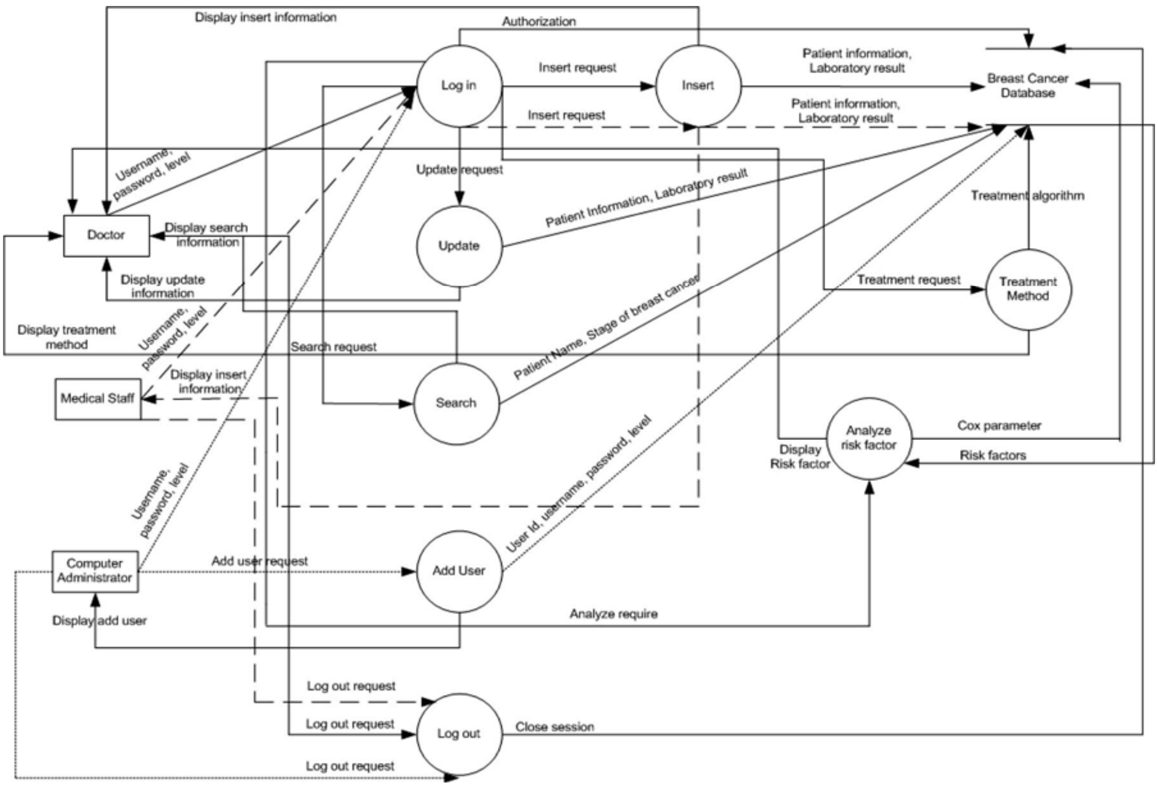

Figure 2 System Architecture design

REFERENCES

[1] Armstrong K, Eisen A, Weber B, "Assessing the risk of breast cancer", N Engl JMed 2000;342(8):564-71.

[2] Sasco AJ, "Breast cancer and the environment", Horm Res. 2003;60 Suppl 3:50 2003;60(Suppl 3:50).

[3] Anne Stotter, Nicola Bright, Paul B Silcocks, Johannes L Botha, "Effect of improved data collection on breast cancer incidence and survival:reconciliation of a registry with a clinical database", BMJ volume 321 22july2000

[4] Resource [Online] Available from : "http://breast-cancer-research.com"

[5] Curtis H.K. Tsang, Chris Bloor. A Medical Expert System Using Object-Oriented Framework . IEEE. (1994) . 176-181

[6] Bridgett, N. A., Brandt, J. and Harris, C. J., "Artificial Neural Networks for use in the Diagnosis and Treatment of Breast Cancer", In Proceedings of 4th Int.Conf. on Artificial Neural Networks, 448--453 04 may 1999

[7] Elizabeth B. Claus, Meredith stowe, and Darryl carter., "Breast Carcinoma In Situ: Risk Factors and Screening Patterns", Journal of the National Cancer Institute, Vol. 93, No. 23, December 5, 2001.

[8] Phayung Meesad, Gary G. Yen, "Combined numerical and linguistic knowledge representation and its application to medical diagnosis", IEEE

Transactions on Systems, Man, and Cybernetics, Part A 33(2): 206-222 (2003).

[9] Paul A. Fearn, Kevin Regan, Frank Sculli, Jared Katz, and Michael W. Kattan, "A Chronological Database as Backbone for Clinical Practice and Research Data Management". IEEE. (2003).

[10] Wenten M, Gilliland FD, Baumgartner K, Samet JM, "Associations of weight, weight change, and body mass with breast cancer risk in Hispanic and non-Hispanic white women" Ann Epidemiol 2002;12(6):435-4.

[11] Claus EB, Stowe M, Carter D, "Breast carcinoma in situ: risk factors and screening patterns", J Natl Cancer Inst 2001;93(23):1811-7.

[12] Tavani A, Gallus S, La Vecchia C, Negri E, Montella M, Dal Maso L, et al, "Risk factors for breast cancer in women under 40 years", Eur J Cancer 1999;35(9):1361-7.

[13] Nyklicek I, Louwman WJ, Van Nierop PW, Wijnands CJ, Coebergh JW, Pop VJ, "Depression and the lower risk for breast cancer development in middle-aged women: a prospective study", Psychol Med 2003;33(6):1111-7.

[14] Morrow M, Strom EA, Bassett LW, Dershaw DD, Fowble B, Giuliano A, et al, "Standard for breast conservation therapy in the management of invasive breast carcinoma", CA Cancer J Clin 2002;52(5):277-300.

[15]Resource[Online]Availablefrom:http://poptop.hypermart.net/brcanpi.html

[16] Resource [Online] Available from : http://newton.nap.edu/books

/0309092132/html/53.html.

[17] Silverstein MJ, Lagios MD, Craig PH, Waisman JR, Lewinsky BS, Colburn WJ, et al, "A prognostic index for ductal carcinoma in situ of the breast. Cancer " 1996;77(11):2267-74.

[18]Resource [Online] Available from :
http://thailand.prd.go.th/the_news_update_view.php?id=3543

[19] John Fox, "Cox Proportional-Hazards Regression for Survival Data", 2002.