PREDICTION OF BREAST CANCER CASES AT GWERU PROVINCIAL HOSPITAL USING ARTIFICIAL NEURAL NETWORKS

Dr. Smartson. P. NYONI[1], Thabani NYONI[2]

[1] *Medical Doctor, ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe*
[2] *MSc Economics Scholar, Department of Economics, University of Zimbabwe, Harare, Zimbabwe*

ABSTRACT

In this endeavor, we apply the ANN (12, 12, 1) model to examine monthly breast cancer cases seen and managed at Gweru Provincial Hospital (GPH). The referral system of Zimbabwe is such that doctors at any health facilities in the country; refer cancer cases straight to central hospitals at the quaternary level. The cases analyzed here were initially managed at GPH. The data cover the period Janauary 2010 – December 2019. The out-of-sample period is January 2020 – December 2021. The applied model is stable as indicated by its residuals. The results of the study basically indicate that breast cancer cases will be ranging between 1 and 3 at GPH over the out-of-sample period. The study offers a 3-fold policy direction in order to prevent, control and effectively manage breast cancer cases not only GPH but also in the country at large.

Keyword: - *ANN, Breast Cancer, Forecasting*

INTRODUCTION

Breast cancer is the most common cancer in women worldwide (Sani *et al*., 2020). In Zimbabwe, breast cancer has become a significant burden to the healthcare system and financially toxic, especially to patients who will have to bear most of the costs of treatment (Kadzatsa & Ndarukwa-Jambwa, 2019). In fact, breast cancer is a major health concern that is plaguing the modern world (Borkar *et al*., 2015) and is also the principal cause of cancer deaths among women worldwide (Azubuike & Okwuokei, 2013). Breast cancer affects not only women but also men and animals (Beg & Jain, 2012). Only 1% of all the cases are found in men (Benson *et al*., 2009; Beg & Jain, 2012).

Breast cancer is a disease in which malignant cells come from breast tissue and proliferate increasingly while they pass immune system without causing any defending and aggressive reaction against it (Setayeshi *et al*., 2011; Rezvani, 2012). This disease usually starts as a hard mass in superior lethal region of breast and may expand gradually to the whole body (ACS, 2011). The main cause of breast cancer is not yet clear, however, a number of its risk factors are well known (Khoury-Collado & Bombard, 2004; NBCF, 2012), including genetic and racial factors, hormones, diet, obesity, menopausal after age of 50, long time use obstructive compulsive pills, hormone therapy, cancer family history and alcohol consumption (Claus *et al*., 1994; Khoury-Collado & Bombard, 2004).

Breast cancer is one of the most common female cancers in Zimbabwe (Muchuweti *et al*., 2017). In Zimbabwe, 11% of cancer deaths are due to breast cancer, with an incidence of 7% (Zimbabwe National Cancer Registry, 2012). Against this background, it becomes instruct to model and forecast breast cancer cases, especially in light of future planning and public health discourse. This study, which is the first of its kind in Zimbabwe; is mainly aimed at predicting the future trends of breast cancer cases at Gweru Provincial Hospital (GPH). The study is envisioned to go a long way in warning the policy makers on the need to "wake up" in the fight the fight against breast cancer in Zimbabwe.

LITERATURE REVIEW

In an USA study, Ayer *et al*. (2010) analyzed breast cancer using Artificial Neural Network (ANN) and their results shows that their ANN has superior discrimination and can effectively discriminate malignant abnormalities from benign ones and accurately predict the risk of breast cancer for individual abnormalities. In an Indian study, Borkar *et al*. (2015) predicted breast cancer using Artificial Neural Networks (ANNs). Actually, their study developed an ANN that determines if patients have breast cancer or not. The study used data from 699 patients to train the network. Their model was validated using the ROC analysis and their results show that ANN can be used to diagnose breast cancer in a rapid manner without causing a great deal of trouble to patients. In another USA study, Liming (2017) used time series analysis to predict cancer incidence rates. The author particularly applied the ARIMA and SARIMA models and found out that the SARIMA $(1, 1, 2)(1, 1, 1)_{12}$ was the best fit model and revealed that cancer incidence was declining in USA over the period 2015 – 2020. In a Nigeria study, Bawa (2017) compared various models on cancer rate modeling and forecasting and found out that the quadratic trend model was the best fit model and that cancer incidence was on the in the Niger State Hospital catchment area.

In a Brazilian study, Rosales-Lopez *et al*. (2018), examined mortality rates from breast cancer using SARIMA models and found out that the SARIMA $(0, 1, 2)(1, 1, 2)_{12}$ was the best fit model and was effective in evaluating the Brazilian screening programme. In an Iranian study, Atashi *et al*. (2018) employed neural networks perceptrons and decision tree algorithm to analyze breast cancer and concluded that both models provided good results, neural network showed better diagnosis for positive cases. In another Iranian study, Khoshdel *et al*. (2019) predicted the incidence of the 3 most common cancers among Iranian military community during 2007 – 2019 using ARIMA models. Concerning breast cancer, their results show that the incidence will increase in the future. In another Nigerian study, Sani *et al*. (2020) used ARMA models to analyze breast cancer cases in Abuja and found out that the best fit model is the ARMA (23, 12) model. The present study will apply the ANN model to predict cancer cases at GPH in Gweru, Zimbabwe.

METHODOLOGY

ANN is a mathematical learning model deployed on a computer that is designed in the same way the biological neural network of the brain functions. The ANN designed, learns how to predict an outcome from a set of attributes. ANN has been successfully used in medical science, pharmacoepidemiology and medical data mining (Borkar *et al*., 2015). In this study, the ANN (12, 12, 1) is applied to predict breast cancer cases at GPH.

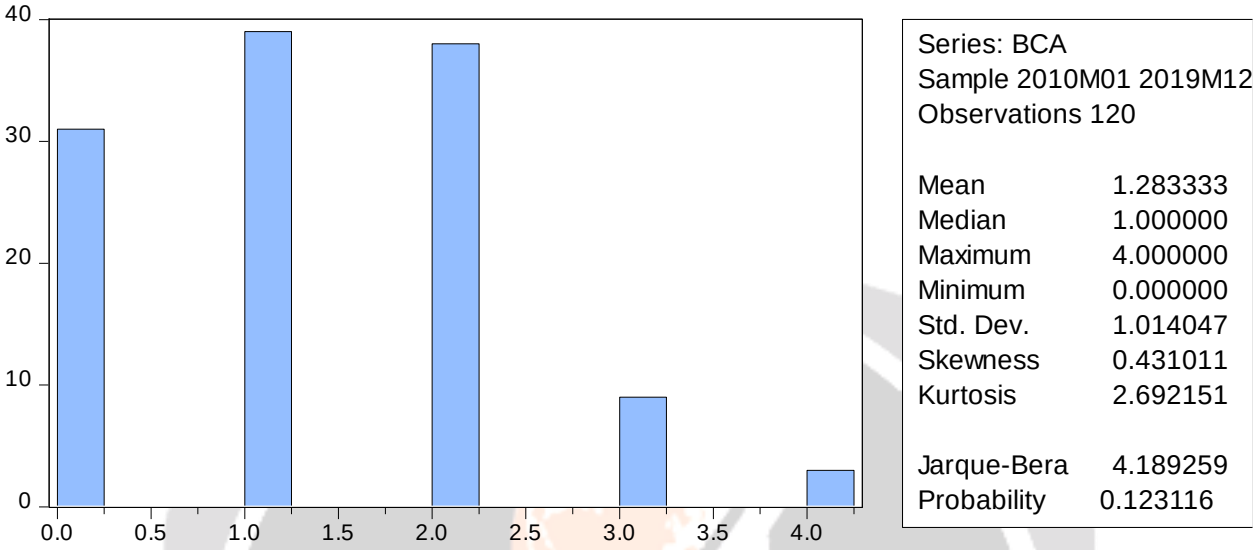
Data Issues

This paper is based on newly diagnosed monthly breast cancer cases (reffered to as BCA series in this study) in women aged 45 years and above at GPH. The data covers the period January 2010 to December 2019 and the out-of-sample forecast covers the period January 2020 to December 2021. All the data employed in this paper was gathered from GPH Health Information Department.

FINDINGS OF THE STUDY

DESCRIPTIVE STATISTICS

Figure 1: Descriptive statistics



The average number of breast cancer cases over the study period is 1 case per month while the maximum is 4 cases per month. For a provincial hospital like GPH, these can be considered as large numbers of breast cancer cases, especially if the statistics are aggregated over time. Therefore, there is need for policy makers in the country to take action in order to reduce breast cancer cases where possible. Striking to note also is that the series under consideration is normally distributed as shown by the kurtosis statistic which approximately equal to the rule of thumb of 3 and the probability value which is statistically insignificant. The validity of this assumption is desirable even if it does not matter in the application of the ANN models. The point is, normally distributed data sets are often easier to predict as compared to those data sets which are not normally distributed.

ANN MODEL SUMMARY FOR DYSENTRY CASES IN CHITUNGWIZA URBAN DISTRICT

Table 1: ANN model summary

Variable	BCA
Observations	108 (After Adjusting Endpoints)
Neural Network Architecture:	
Input Layer Neurons	12
Hidden Layer Neurons	12
Output Layer Neurons	1
Activation Function	Hyperbolic Tangent Function
Back Propagation Learning:	
Learning Rate	0.005
Momentum	0.05
Criteria:	
Error	0.141724
MSE	0.099189
MAE	0.256140

Table 1 shows the applied ANN (12, 12, 1) model.

Residual Analysis for the ANN model

Figure 2: Residual analysis

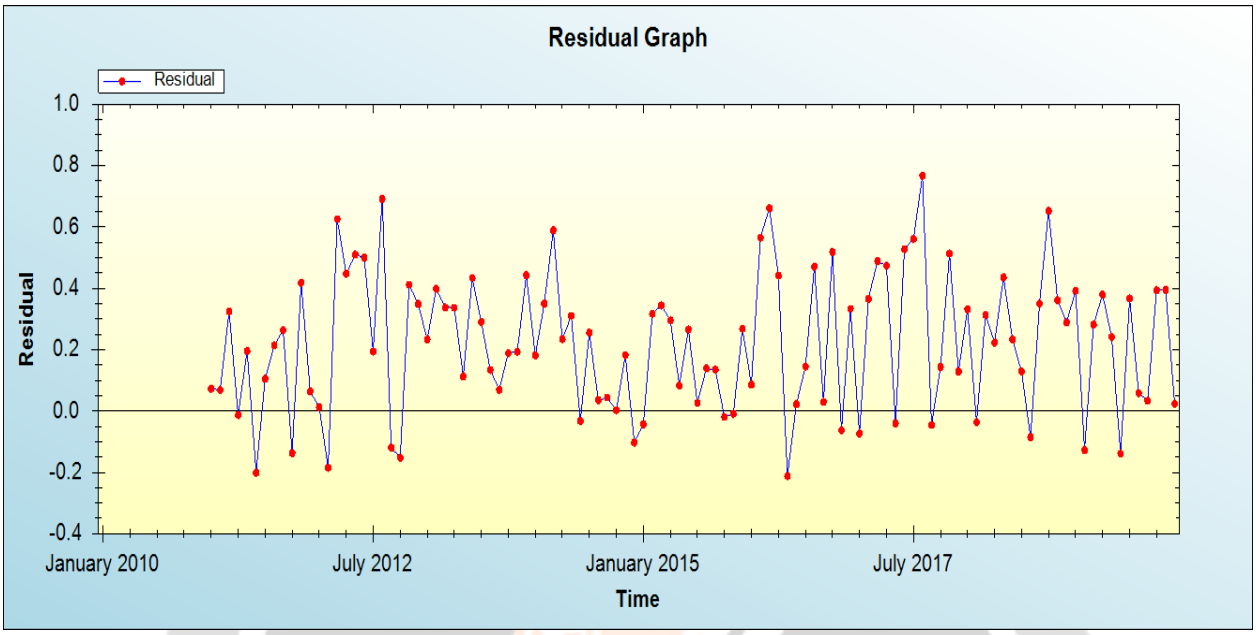


Figure 2 shows the residual plot. The residual have been extracted from the applied model. It is clear that the model is stable as shown by the residuals which revolve closer to zero.

In-sample Forecast for BCA

Figure 3: In-sample forecast for the BCA series

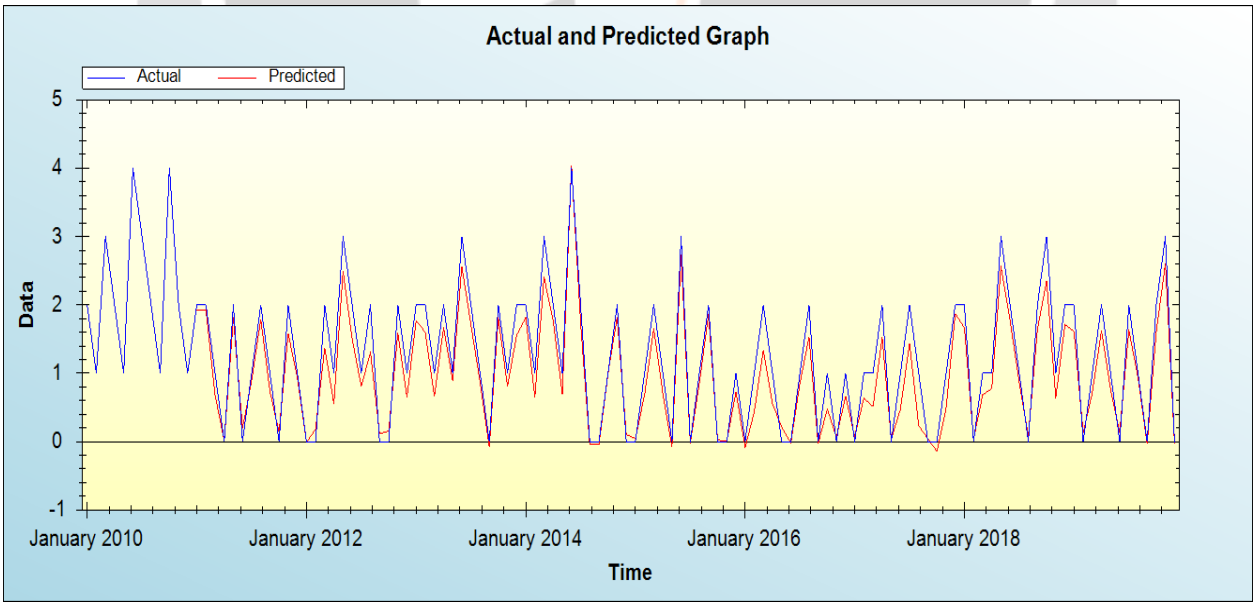
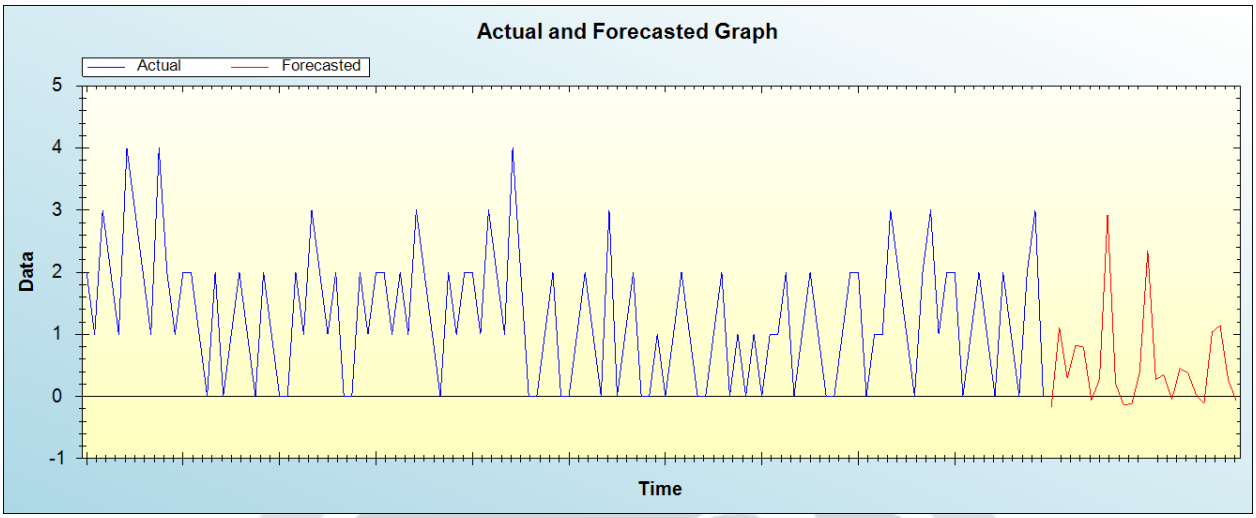


Figure 3 is the in-sample forecast of the applied model.

Out-of-Sample Forecast for BCA: Actual and Forecasted Graph

Figure 4: Out-of-sample forecast for BCA: actual and forecasted graph



Out-of-Sample Forecast for BCA: Forecasts only

Table 2: Tabulated out-of-sample forecasts

Month/Year	Predicted BCA
January 2020	-0.1758
February 2020	1.1110
March 2020	0.2953
April 2020	0.8201
May 2020	0.7962
June 2020	-0.0593
July 2020	0.2803
August 2020	2.9251
September 2020	0.2044
October 2020	-0.1359
November 2020	-0.1181
December 2020	0.3907
January 2021	2.3429
February 2021	0.2733
March 2021	0.3498
April 2021	-0.0511

May 2021	0.4503
June 2021	0.3811
July 2021	0.0349
August 2021	-0.1151
September 2021	1.0356
October 2021	1.1445
November 2021	0.2673
December 2021	-0.0637

Figure 5: Graphical presentation of out-of-sample forecasts

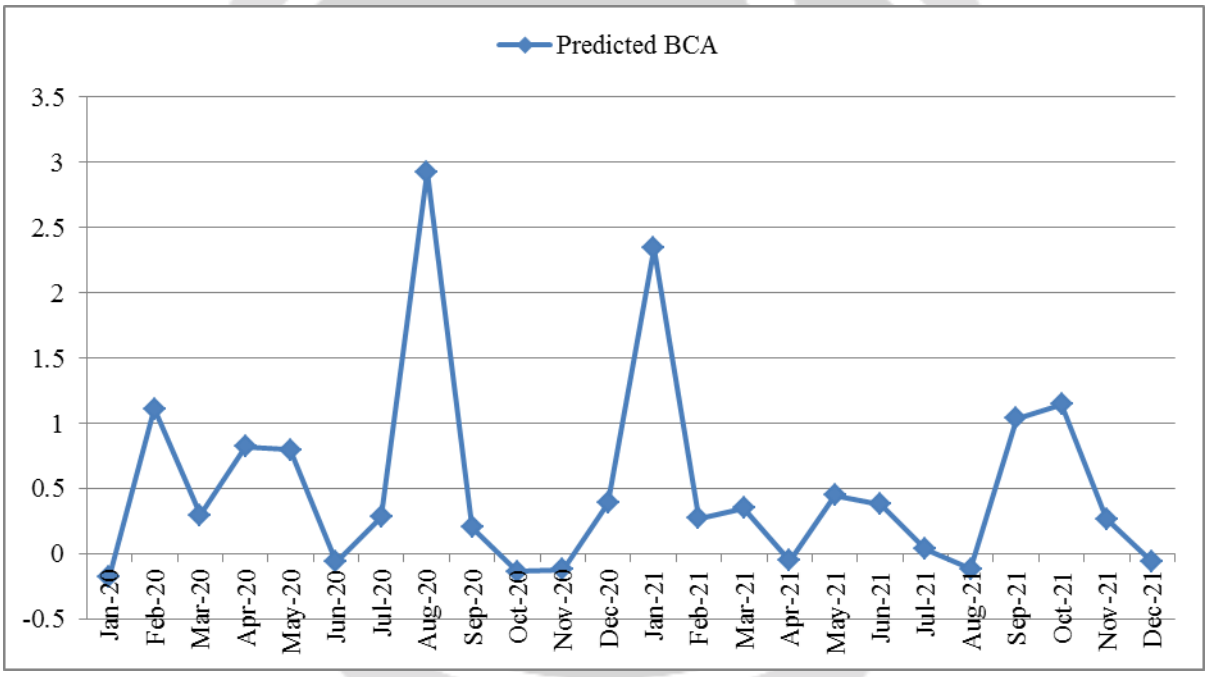


Figure 4 and 5 as well as table 2 show the out-of-sample forecasts of the applied ANN (12, 12, 1) model. Striking to note is the fact that the predicted breast cancer cases at GPH are generally declining over the out-of-sample period. A closer look at figure 5 points the fact that the predicted breast cancer cases range between 1 and 3 cases per month over the out-of-sample period. GPH and the relevant government authorities have a role to play in controlling, preventing and managing breast cancer cases at GPH just like in any other health facility in the country.

RECOMMENDATIONS

i. GPH should engage in educational campaigns that must include sensitization on healthy foods that help prevent breast cancer, for example, citrus fruits (especially peaches and berries), beans, eggs, fish, and vegetables (especially kale and broccoli). People should also be discouraged from smoking as it increases the risk of breast cancer. People should also be taught on how to carry out self-examination and that they

should immediately seek medical attention once a breast lump or something suspicious has been observed. Furthermore, people ought to exercise in order to ensure a healthy and physically active body.

ii. The government of Zimbabwe should avail resources for treatment of breast cancer at GPH. Early detection and treatment increase patient survival. Advanced cases of breast cancer should be reffered urgently to central hospitals for urgent attention. This can potentially save many lives.

iii. There is also need to limit dose and duration of hormone therapy as this can increase the risk of breast cancer.

CONCLUSION

Breast cancer is a real public health threat in Zimbabwe. It is no longer a mystery. The current study predicted the number of breast cancers at GPH over the period January 2020 to December 2021. Even though, the predictions show a general decline in breast cancer cases for GPH and its catchment area, it is clear that the disease will always remain with us until drastic action taken to prevent, control and treat it. The study applied to ANN models which are currently trending in terms of analyzing epidemiological data sets these days. Further studies are encouraged to apply ANN models to examine breast cancer risk factors: this can potentially bring forth interesting results which can complement existing research output.

REFERENCES

[1] American Cancer Society (ACS) (2011). Breast Cancer Facts and Figures, *ACS*, New York.

[2] Atashi, A., Sohrabi, S., & Dadashi, A. (2018). Applying Two Computational Classification Methods tot Predict the Risk of Breast Cancer: A Comparative Study, *Multidisciplinary Cancer Investigation*, 2 (2): 1 – 6.

[3] Ayer, T., *et al*. (2010). Breast Cancer Risk Estimation With Artificial Neural Networks Revisited: Discrimination and Calibration, *Cancer*, 116 (4): 3310 – 3321.

[4] Azubuike, S. O., & Okwuokei, S. O. (2013). Knowledge, Attitude and Practices of Women Towards Breast Cancer in Benin City, Nigeria, *Annals of Medical and Health Sciences Research*, 3 (2): 155 – 160.

[5] Bawa, M. (2017). Comparison of Various Models on Cancer Rate and Forecasting, *Journal of Applied Science and Environmental Management*, 21 (5): 957 – 959.

[6] Beg, M. M., & Jain, M. (2012). An Analysis of the Methods Employed for Breast Cancer Diagnosis, *Galgotias College of Engineering & Technology*, Noida.

[7] Benson, J. R., Jatoi, I., & Keish, M. (2009). Early Breast Cancer, *Lancet*, 373: 1642 – 1643.

[8] Borkar, M., Deulkar, K., & Garg, A. (2015). Prediction of Breast Cancer Using Artificial Neural Networks, *International Journal of Engineering Research and Technology*, 4 (9): 475 – 477.

[9] Claus, E. B., Risch, N., & Thompson, W. D. (1994). Autosomal Dominant Inheritance of Early-onset Breast Cancer Implications for Risk Prediction, *Cancer*, 73: 643 – 651.

[10] Kadzatsa, W., & Ndarukwa-Jambwa, S. (2019). Breast Cancer Treatment in Resource Constrained Countries: A Zimbabwe Perspective, *Current Breast Cancer Reports*, 11: 170 – 174.

[11] Khoshdel, A. R., *et al*. (2019). The Prediction Incidence of the Three Most Common Cancers Among Iranian Military Community During 2007 – 2019: A Time Series Analysis, *Journal of Preventive Medicine and Hygiene*, 60: 256 – 261.

[12] Khoury-Collado, F., & Bombard, A. T. (2004). Hereditary Breast and Ovarian Cancer: What the Primary Care Physician Should Know, *Obstetric Gynecology*, 59 (7): 537 – 542.

[13] Liming, X. (2017). Time Series Analysis and Prediction on Cancer Incidence Rates, *Journal of Medical Discovery*, 2 (3): 1 – 8.

[14] Muchuweti, D., Nyandoro, G., Muguti, E. G., & Muchaziwepi, T. (2017). Factors Contributing to Delayed Breast Cancer Presentation: A Prospective Study at Parirenyatwa Group of Hospitals, Harare, Zimbabwe, 2010 – 2013, *Journal of Cancer and Tumor International*, 5 (1): 1 – 10.

[15] National Breast Cancer Foundation (NBCF) (2012). Early Detection, *NBCF*, New York.

[16] Rezvani, B. (2012). The Relationship Between TRU91 Polymorphism in Vitamin D Receptor (VDR) Gene and Breast Cancer, *Maters Thesis*, Islamic Azad University, Damghan.

[17] Rosales-Lopez, A., et al. (2018). The Use of Intervention Analysis of the Mortality Rates From Breast Cancer in Assessing the Brazilian Screening Programme, *Research on Biomedical Engineering*, 38 (4): 285 – 290.

[18] Sani, M., Olanrewaju, S. O., & Oguntade, E. S. (2020). Analysis of Reported Cases of Breast Cancer in Abuja, Nigeria: A Time Series Approach, *Iconic Research and Engineering Journals*, 3 (9): 59 – 69.

[19] Setayeshi, S., *et al*. (2011). Breast Cancer and Technical Analysis of it Diagnostics, *Bitarafan*, Tehran.

[20] Zimbabwe National Cancer Registry (2012). Breast Cancer, *ZNCR*, Harare.