Rwanda Medical Journal RMJ

ORIGINAL ARTICLE Open Access

Predictors of molecular subtypes in women with breast cancer in Rwanda

Authors: F. Ntirenganya[1,*]; J. D. Twagirumukiza[2]; G. Bucyibaruta[3]; F. Byiringiro[1]; B. Rugwizangoga[1]; S. Rulisa[1]

Affiliations: *[1]University of Rwanda, College of Medicine and Health Sciences, School of Medicine and Pharmacy; [2]University of Rwanda, College of Medicine and Health Sciences, School of Public Health; [3]Department of Statistics and Actuarial Science, University of Waterloo, Canada*

ABSTRACT

INTRODUCTION: Breast cancer (BC) constitutes a major public health problem worldwide. It remains a major scientific, clinical and societal challenge, generally in Africa and particularly in Rwanda. The purpose of this study was to determine clinical and histopathological predictors of BC molecular subtypes in Rwandan women.

METHODS: A retrospective cohort study including patients with histological confirmation of BC. Using R statistical software, a regression model for multinomial responses was developed. Univariate and multivariate logistic regression analyses were used to identify independent BC molecular subtypes predictors. A two-sided p<0.05 indicated a statistically significant difference.

RESULTS: Forty seven percent of cases presented with advanced stages (Stage III and IV). Postmenopausal BC (p=0.0142), absence of infertility (p=0.018) predicted Luminal A subtype with a predictive accuracy of 0.65. Age (p=0.003), postmenopausal BC (p=0.005), absence of axillar lymph nodes (p= 0.008) and poorly differentiated tumor (p=0.012) were predictors for Luminal B subtype with a predictive accuracy of 0.86. Age (p=0.045), BMI (p=0.005), rapid progression (p=0.032), tumor size T2-T3 (p<0.001) were predictors of HER2-Enriched subtype with a predictive accuracy of 0.70. Age below 40 (p=0.005), painless mass (p=0.030), nodal involvement (p=0.008), Nottingham grade 3 (p<0.001) predicted Triple Negative tumors with a predictive accuracy of 0.71.

CONCLUSION: Clinical and histopathological tumor characteristics can be used to predict BC molecular subtypes with acceptable accuracy. Further studies are needed to explore the possibility of developing a scoring system for clinical decision-making, especially in settings where immunohistochemistry testing is limited.

Keywords: Breast Neoplasm, Women, Immunohistochemistry, Pathology.

***Corresponding author:** Faustin Ntirenganya, Email: fostino21@yahoo.fr, University of Rwanda, College of Medicine and Health Sciences, School of Medicine and Pharmacy;
Potential Conflicts of Interest (CoI): All authors: no potential conflicts of interest disclosed; **Funding:** All authors: This study was sponsored by the Consortium for Advanced Research Training in Africa (CARTA). CARTA is jointly led by the African Population and Health Research Center and the University of the Witwatersrand and funded by the Carnegie Corporation of New York (Grant No--B 8606.R02), Sida (Grant No:54100113), the DELTAS Africa Initiative (Grant No: 107768/Z/15/Z) and Deutscher Akademischer Austauschdienst (DAAD). **Academic Integrity.** All authors confirm that they have made substantial academic contributions to this manuscript as defined by the ICMJE; **Ethics of human subject participation:** The study was approved by the local Institutional Review Board. Informed consent was sought and gained where applicable; **Originality:** All authors: this manuscript is original has not been published elsewhere; **Review:** This manuscript was peer-reviewed by three reviewers in a double-blind review process; **Type-editor:** Faloon (USA).

Received: 14th June 2022; **Initial decision given:** 29TH June 2022; **Revised manuscript received:** 14th July 2022; **Accepted:** 29th November 2022.
 Publisher: Rwanda Biomedical Centre (RBC)/Rwanda Health Communication Center, P. O. Box 4586, Kigali. ISSN: 2079-097X (print); 2410-8626 (online)

Citation for this article: F. Ntirenganya; J. D. Twagirumukiza; G. Bucyibaruta et al. Predictors of molecular subtypes in women with breast cancer in Rwanda. Rwanda Medical Journal, Vol. 79, no. 4, p. 65-77, 2022. https://dx.doi.org/10.4314/rmj.v79i4.7

INTRODUCTION

Breast cancer (BC) constitutes a major public health problem worldwide. It remains a major scientific, clinical and societal challenge generally in Africa and particularly in Rwanda.

BC molecular sub-typing aims to tailoring treatments to individual tumour characteristics in order to improve outcomes [1-4]. In the last decade, molecular sub-typing became the gold standard and cornerstone of modern breast cancer management. In fact, breast cancer being a heterogeneous disease with both intra-tumour and inter-tumour heterogeneity, it was not adequate to treat and/or follow all breast cancer patients the same way. Hence, molecular sub-typing permitted to overcome the breast cancer heterogeneity challenge and allowed individualized therapies based on each patient's tumour characteristics for better outcomes [1,2,5,6].

The advance in microarrays and immunohistochemistry (IHC) techniques has led to a new paradigm in breast cancer carcinogenesis and the understanding of breast cancer heterogeneity [7,8,9]. In that line, growing evidences are showing that breast cancer is composed of multiple subtypes that occur at different rates, in different groups, with different response to treatments and have a varied long-term survival rates [8,10-16]. For this reason, treatment planning and prognostication in breast cancer have become more demanding over the last decade. Indeed, different breast cancer subtypes may be associated with different risk factors and may have different preventive and early detection strategies. Consequently, tailoring screening, early detection and treatments to intrinsic molecular characteristics is crucial in order to improve breast cancer outcomes.

So far, IHC allowed identifying at least 5 main breast cancer molecular subtypes differing completely in progression and outcomes. These are Luminal A, Luminal B, HER2-Enriched, Triple Negative and Unclassified [1,6,17-20]. Different authors identified luminal-A sub-group of ER positive, PR positive but negative HER2/Neu tumors as being associated with the best outcomes, while Triple Negative and HER2/Neu enriched tumors are associated with the worst outcomes [5,21,22]. Unfortunately, Immunohistochemistry (IHC) technology is still expensive and not widely accessible for the majority of patients from countries with limited resources like Rwanda. Hence, patients may receive quite often blind or empirical medical treatments, which may be one of the main contributors to the poor outcomes currently seen in LMICs.

Consequently, it may be important to think about alternatives tools which may help to sub-classify BC patients into different subgroups to bridge the gap created by the non-availability of IHC technology in the majority of LMICs. Indeed, the majority of centres in Africa can conduct a proper clinical exam and have access to standard breast cancer histopathology analysis. For that reason, a predictive model, taking into consideration breast cancer clinical and histopathology characteristics may be useful to guide clinicians' decision making, helping them to offer individualized breast cancer treatments for improved outcomes. We conducted this study to determine clinical and histopathological predictors of BC molecular subtypes in Rwandan women.

METHODS

This was a retrospective cohort study including patients with histological confirmation of breast cancer. A pre-established questionnaire was administered for socio-demographic, clinical and histopathological characteristics.

Histopathology analysis was done using usual hematoxylin and eosin (H&E) stains, highlighting the unique micro-architectural and morphological aspect of the tumour. Breast cancer histology type, differentiation, grade, vascular invasion, lymphatic invasion and lymph nodes involvement were reported by a consultant pathologist and validated by a second pathologist in the same laboratory.

Immunohistochemistry (IHC) used specific antigens identified in Formalin-Fixed, Paraffin-Embedded (FFPE) tissues. The method was based on antigen-antibody interaction (Taylor and Burns, 1974). In IHC, antigen-antibody reactions were visualized through light microscopy by means of colour signal which is produced by labelling or tagging the antibody. The morphology of the tissue around the specific antigen was clearly visualized by counter staining with hematoxylin. Results of stained IHC markers were reported semi quantitatively by pathologists.

IHC staining was conducted in 3 steps:

Step 1: Fixation

For this study, core needle biopsies or mastectomy

specimens were fixed immediately in 10% neutral buffered formalin to avoid problems of interpretation which may arise due to under fixation due to elution of the stain or over fixation which causes masking of antigen sites, and hence false negative results.

Step 2: Antigen retrieval

Antigen retrieval heating technique was used to restore the tertiary structure.

For that purpose, the FFPE tissues were cut into 3- to 4-micron thin sections, on glass slide coated with Poly L Lysine (PLL).

Step 3: Antigen-antibody interaction and labeling/ Detection: For this step, the direct methods using labeled monospecific antibodies were used.

Estrogen receptors (ER), progesterone receptors (PR) and Human epidermal growth factor receptor 2 (HER2/Neu) have been analysed and reported as positive or negative. However, for HER2/Neu, if the IHC result is 3+, the cancer is HER2/Neu positive. If the IHC result is 1+, the cancer is HER2/Neu negative. However, if the result is 2+ the HER2/Neu status is not clear (Equivocal) and needs further testing by FISH to clarify the result. Unfortunately, FISH technology is not available in the country. For the purpose of this study, tumors with equivocal HER2/Neu status were considered "unclassified".

Using different combinations of ER, PR and HER 2/Neu results, breast cancers were classified into luminal A, luminal B, HER2-type, basal-like (triple negative) and unclassified molecular subtypes.

Statistical analysis

Statistical software R was utilized. Continuous values were compared with the Student's t-test. All continuous variables were verified for normality by the Shapiro–Wilk test. Categorical variables were compared with Pearson's Chi-square test or Fisher's exact test. Univariate and multivariate logistic regression analyses were considered to identify independent predictors of the molecular subtypes of breast cancer. A two-sided $p<0.05$ was considered to indicate a statistically significant difference. Classification performances were evaluated based on the receiver operating characteristic (ROC) curve and area under the curve (AUC) in the validation cohort.

Data were cleaned by removing missing values, adjusting levels of some categorical variables and selecting key variables. At the end, the data set contained 1 response variable and 23 predictors.

The entire data set was split into two sets: a training set and a testing set. Once we have two data sets, we used the training set to build and train the model. Once the model is ready, we tested it on the testing set for accuracy and how well it performs. The objective being to have the model performing on any data with the highest accuracy.

There are various methods that can be used to split the data into training and testing sets. Generally, the observations will be assigned to training and testing sets randomly so that both sets resemble in their properties as much as possible. Split sizes can also differ based on scenario: it could be 50:50, 60:40, or 2/3rd and 1/3rd. While there are many empirical studies and papers on the best way to split data, 80/20 or 70/30 split are widely used. In our case we consider the scenario of 80/20 which allows us to have more data in training set.

A combined model was built by performing multivariate logistic regression that included all predictor variables and molecular subtypes.

We evaluated association between predictors and each of the molecular subtypes and assessed the predictive accuracy for the prediction of different molecular subtypes using Stepwise logistic regression analysis.

Stepwise logistic regression analysis

Stepwise variable selection is a practical alternative to examining all possible models that should be built from the predictor variables. Beginning with the 23 variables, we used the backwards selection procedure available in "MASS" package (R software) by using the function stepAIC. The variables that do not meet the significance level $p<0.05$ were omitted. The final model includes all the variables that are statistically significant.

As an alternative model, we fit a logistic model with all 23 candidate variables under a manually procedure. Those that resulted to be significant at this first step were considered in the second step. We repeated the process until getting the last model with significant variables. In addition, one of the criteria of keeping or dropping out the variable in the model was its contribution in terms of prediction. If removing variable reduces the prediction power of the model, in that case we kept it otherwise it was removed.

Ethical approval and consent to participate: The study was approved by IRB of the College of Medicine and Health Sciences, University of Rwanda (Approval notice: No110/CMHS/IRB/2019) and CHUK and BCCOE review committees. Written informed consent was obtained prior to prospective

Table 1: Clinical characteristics

Clinical characteristics	Category	Menopausal status		P value
		Premenopausal N=170	Postmenopausal N= 170	
Self-reported progression pattern	Slow	13 (7.6%)	32 (18.8%)	<.0001
	Intermediate	93 (54.7%)	108 (63.5%)	
	Rapid	64 (37.6%)	30 (17.6%)	
Axillary lymph nodes	Yes	124 (72.9%)	118 (69.4%)	0.473
	No	46 (27.1%)	52 (30.6%)	
Node status	1 (0 node)	42 (25.3%)	49 (29.5%)	0.073
	2 (1-3 nodes)	102 (61.4%)	107 (64.5%)	
	3 (>3 nodes)	22 (13.3%)	10 (6.0%)	
Breast site	Right	86 (50.6%)	82 (48.5%)	0.703
	Left	84 (49.4%)	87 (51.5%)	
Tumor size according to TNM	T1	8 (4.7%)	15 (8.8%)	0.306
	T2	70 (41.2%)	77 (45.3%)	
	T3	74 (43.5%)	63 (37.1%)	
	T4	18 (10.6%)	15 (8.8%)	
Distant metastasis	Yes	23 (13.5%)	15 (8.8%)	0.169
	No	147 (86.5%)	155 (91.2%)	
Clinical stage TNM	Stage 1	5 (2.9%)	16 (9.4%)	0.024
	Stage 2	78 (45.9%)	78 (45.9%)	
	Stage 3	63 (37.1%)	62 (36.5%)	
	Stage 4	24 (14.1%)	14 (8.2%)	
Presence of chronic disease	Yes	17 (10.0%)	38 (22.4%)	0.002
	No	153 (90.0%)	132 (77.6%)	

data collection

RESULTS

Three hundred and forty participants were recruited into the study. The median age was 49 years (Range 28-89 years). Forty eight percent of cases presented advanced stages of the disease (stage III and IV) (Table 1). The majority of patients had invasive ductal carcinoma (95.8%). Subtypes of poor prognosis (HER2 enriched 14.7%, triple negative 12.9%, unclassified 32.9%) represented 60.6% (Table 2).

Interpreting the above results, two things have to be considered: the sign and size of the effect of the predictor over the response. To do so, the table

Table 2: Histological characteristics

Histological characteristics	Category	Menopausal status		p
		Premenopausal	Postmenopausal	
Histology type	Invasive ductal carcinoma	167 (98.2%)	159 (93.5%)	0.053
	Others*	3 (1.8%)	11 (6.5%)	
Nottingham histologic grade	Grade 1	11 (6.5%)	28 (16.5%)	0.036
	Grade 2	76 (44.7%)	68 (40.0%)	
	Grade 3	83 (48.8%)	74 (43.5%)	
Differentiation	Well differentiated	23 (13.5%)	105 (61.8%)	<.0001
	Poorly differentiated	124 (72.9%)	33 (19.4%)	
	Undifferentiated	23 (13.5%)	32 (18.8%)	
Presence of poor prognostic factors	Comedo necrosis	10 (5.9%)	7 (4.1%)	0.142
	Lymphatic invasion	9 (5.3%)	3 (1.8%)	
	None/unknown	151 (88.8%)	160 (94.1%)	
*Molecular subtypes	Luminal A	45 (26.5%)	30 (17.6%)	0.062
	Lumina B	22 (12.9%)	13 (7.6%)	
	Her2 Enriched	25 (14.7%)	24 (14.1%)	
	Triple negative	22 (12.9%)	32 (18.8%)	
	Unclassified	56 (32.9%)	71 (41.8%)	

** Molecular subtypes classification: Luminal A: ER+/ PR+ ; HER2/Neu Negative; Luminal B: ER+/PR+; HER/Neu Positive; HER2 Enriched: ER-/PR - ; HER2/Neu Positive; Triple Negative: ER -/ PR-; HER2/Neu Negative; Unclassified: Any other combination.*

of results contained a column called "z values": the effect size. When negative, it meant that the referenced level of predictor favors the outcome of interest. In other words, the probability in favor of the outcome of interest decreases with respect to the corresponding level of that predictor. When positive, meant that the probability in favor of the outcome of interest increases with respect to that level.

Overall, age, fertility, menopausal status, tumor differentiation, lymph nodes involvement and Nottingham grade have been retained in the final model as predictors of breast cancer molecular subtypes in general (Table 3). However, for specific molecular subtypes, only postmenopausal breast cancer (p=0.0142), and no history of infertility (p=0.018) have been retained as Luminal A subtype predictors (Table 4) with model accuracy of 0.65 (Figure 1).

Predictive model for Luminal B retained age (p=0.003), postmenopausal cancer (p=0.005), absence of axillar lymph nodes (p= 0.008) and poorly differentiated tumor (p=0.012) (Table 5) with a predictive accuracy of 0.86 (Figure 2).

For HER2-Enriched subtype, age (p=0.045), BMI (p=0.005), rapid progression (p=0.032), T2 tumor size (p<0.001), T3 tumor size (p=0.008), histology types other than invasive ductal carcinoma

Table 3: Predictive model for molecular subtypes in general

Predictors	Likelihood ratio	Df	P value
Age	17.594	3	<0.001
Menopausal status	6.212	3	0.1017
Tumor differentiation	10.998	6	0.0884
Infertility	8.685	3	0.0337
Lymph nodes	8.144	3	0.0431
Nottingham grade	34.458	6	<0.001
Nodal status	14.459	9	0.1068

(p=0.049) have been retained in the final fitting model (Table 6) with a predictive accuracy of 0.70 (Figure 3).

For triple negative tumors, age (p=0.005), painless mass (p=0.030), no family history of breast cancer (p=0.046), nodal involvement (p=0.008), Nottingham grade 3 (p<0.001) have been retained in the final model (Table 7) with 0.71 as prediction accuracy (Figure 4).

DISCUSSION

We studied clinical and histopathological predictors of breast cancer molecular subtypes in Rwandan women.

340 patients with histologically confirmed breast cancer were recruited into the study. The median age was 49 years (Range 28-89 years). Of all cases, 47.9% presented advanced stages of the disease (stage III and IV) and had invasive ductal carcinoma (95.8%). Subtypes of poor prognosis (HER2 enriched 14.7%, triple negative 12.9%, unclassified 32.9%) represented 60.6%, reflecting the general picture of breast cancer as seen in Africa [23,24].

Breast cancer molecular sub-typing aims to tailoring treatments to individual tumor characteristics in order to improve outcomes [1-4]. In the last decade, molecular sub-typing became the gold standard and corner stone of modern breast cancer management. In fact, breast cancer being a heterogeneous disease with both intra-tumor and inter-tumor heterogeneity as previously seen; it was not adequate to treat and/or follow all breast cancer patients the same way. Hence, molecular sub-typing permits to overcome the breast cancer heterogeneity challenge and allows individualizing therapies on each patient's tumor characteristics for better outcomes [1,2,5,6].

In the current literature, there is a clear consensus that Luminal A tumors represent the vast majority of breast cancers. Furthermore, it seems that most established breast cancer risk factors reflect those of luminal-A subtype. Hence, other molecular sub-types may be not yet totally understood with the hypothesis that breast cancer risk factors may be differently associated with other intrinsic tumor subtypes[1,6,18,19,20]. Indeed, few studies investigated predictors of breast cancer molecular subtypes. In our study, it was found that having breast cancer in postmenopausal period (p=0.0142), in otherwise fertile women (p=0.018) predicted Luminal A subtype with accuracy of 0.65. The Luminal A tumors frequently have low histological grade, low degree of nuclear pleomorphism, low mitotic activity and include special histological types (i.e., tubular, invasive cribriform, mucinous and lobular) with good prognosis [25]. However, there is no study yet which calculated the predictive value of the above mentioned characteristics.

Luminal-B tumors comprise 15%-20% of breast cancers and have a more aggressive phenotype, higher histological grade, proliferative index and a worse prognosis. This subtype has a higher recurrence rate and lower survival rates after relapse compared to luminal-A subtype [26,27]. Regarding the histological grade, a significant difference was

Table 4: Predictive model for Luminal A subtype

Predictors	OR	95% CI	SE	Z	P value
Intercept	5.244	0.648-46.535	1.072	1.545	0.031
Menopausal status					
Premenopausal	1				
Postmenopausal	0.402	0.191-0.823	0.371	-2.453	0.0142
Family history of breast cancer					
Yes	1				
No	2.204	0.764-7.261	0.567	1.394	0.163
Infertility					
Yes	1				
No	0.162	0.031-0.689	0.769	-2.366	0.018
Disease progression					
Slow	1				
Intermediate	0.646	0.189-2.143	0.611	-0.714	0.475
Rapid	0.442	0.105-1.779	0.715	-1.139	0.254
Nottingham grade					
Grade I	1				
Grade II	1.178	0.289-5.092	0.718	0.228	0.819
Grade III	0.355	0.079-1.651	0.762	-1.358	0.174

noted among the molecular subtypes by Manal et al. [27]. In fact, Luminal A was associated with the lowest proportion of histological grade III. Luminal A had also the lowest proportion of tumors with diameter larger than 5 cm. These findings are consistent with previous studies that have shown that Luminal A tumors tend to be slow-growing and are associated with a good prognosis [22,28,29]. In these studies, Luminal A subtype had a better prognosis than Luminal B subtype. Indeed, in comparison with Luminal A, Luminal B had a higher percentage of tumors with a large diameter (>5.0 cm), a larger proportion of histological grade III, and a higher percentage of vascular emboli and lymph node involvement. The above mentioned data are consistent with our findings where luminal B breast cancer was predicted by age (p=0.003), postmenopausal cancer (p=0.005), absence of axillar lymph nodes (p= 0.008) and poorly differentiated tumor (p=0.012) with a predictive accuracy of 0.86.

It has been recorded in the literature that between 15-25% of breast cancers possess overexpression of HER2 and yield unfavorable clinical outcome [30,31]. These tumors display the highest frequency of poorly differentiated cancers and metastatic lymph nodes. The registered rates of HER2 + subtypes are quite higher in Asian and African populations [21,27,32]. Such regional and ethnic differences in the grades of the tumor are most probably related to genetic, biological and environmental factors. Studies so far published correlating the stage of breast cancer at the time of diagnosis with the clinicopathological characteristics of the affected patients demonstrated that 64.4% and 67.2% exhibiting Luminal A and Luminal B subtypes respectively were diagnosed at Stages I and II whereas 68% and 62% of those harbouring the TN and HER2+ respectively presented at advanced stages (III and IV), aligning with our findings which shows that age (p=0.045), BMI(p=0.005), rapid progression (p=0.032), T2 tumor size (p<0.001), T3 tumor size (p=0.008), histology types other than invasive ductal carcinoma (p=0.049) are predictors of HER2-Enriched breast cancer with a predictive accuracy of 0.70.

In our study, triple negative tumors represented 12.9%, a figure below the high prevalence of triple negative tumors usually reported in African literature [33,34,35]. It is important to note that

Table 5: Predictive model for Luminal B subtype

Predictors	OR	95% CI	SE	Z	P value
Intercept	7.959	0.049-1006.41	2.467	0.841	0.400
Age	0.887	0.815-0.956	0.0404	-2.953	0.003
Menopausal status					
Premenopausal	1				
Postmenopausal	12.629	2.282-81.73	0.904	2.805	0.005
Breast swelling					
Yes	1				
No	6.656	1.020-139.267	1.155	1.641	0.101
Lymph nodes					
Yes	1				
No	0.159	0.036-0.573	0.691	-2.657	0.008
Clinical stage					
Stage 1	1				
Stage 2	1.616	0.186-35.611	1.225	0.392	0.694
Stage 3	0.242	0.022-5.868	1.312	-1.800	0.280
Presence of NCD					
Yes	1				
No	0.283	0.075-1.105	0.672	-1.875	0.061
Differentiation					
Well differentiated	1				
Poorly differentiated	5.912	1.581-26.447	0.710	2.501	0.012
Undifferentiated	0.838	0.099-5.064	0.961	-0.184	0.584

the numbers of triple negative tumors reported in African literature vary considerably to make them questionable. In fact, breast cancers reported to be "Triple negative" in Africa range from 20 to 90%. This variability makes many authors doubting the quality of Immunohistochemistry (IHC) done in Africa [27,36-40]. Whether the various numbers of triple negative tumors seen in Africa represent geo-ethnic factors or simply technical and procedural Immunohistochemistry (IHC) errors has yet to be determined. Certainly, when IHC is not done in optimal conditions, it may be source of many false negative which may increase the number of triple negative tumors [27,36-40]. ASor triple negative tumors, age (p=0.005), painless mass (p=0.030), no family history of breast cancer (p=0.046), nodal involvement (p=0.008), Nottingham grade 3 (p<0.001) have been retained in the final model with 0.71 as prediction accuracy.

The majority of authors agree that breast cancer is a heterogeneous disease with five main intrinsic tumor subtypes so far identified: Luminal A (ER+/PR-, ER-/PR +, HER2 -), Luminal B (ER+/PR-, Er-/PR+ and HER2+), HER2 Enriched (ER-,PR-, HER2+), Triple Negative (ER-,PR-, HER-) and Unclassified [17,41,42]. These subtypes are different in tumor expression, phenotypes and outcomes [1,2,3,4] and have revolutionized breast cancer management. However, Immunohistochemistry (IHC) technology is still expensive and not yet neither readily available nor accessible for the majority of patients in countries with variable resources like Rwanda. Hence, patients may receive generally blind or empirical medical treatments, which may be one of the main contributors to current poor outcomes seen in LMICs.

Multinomial regression uses a maximum

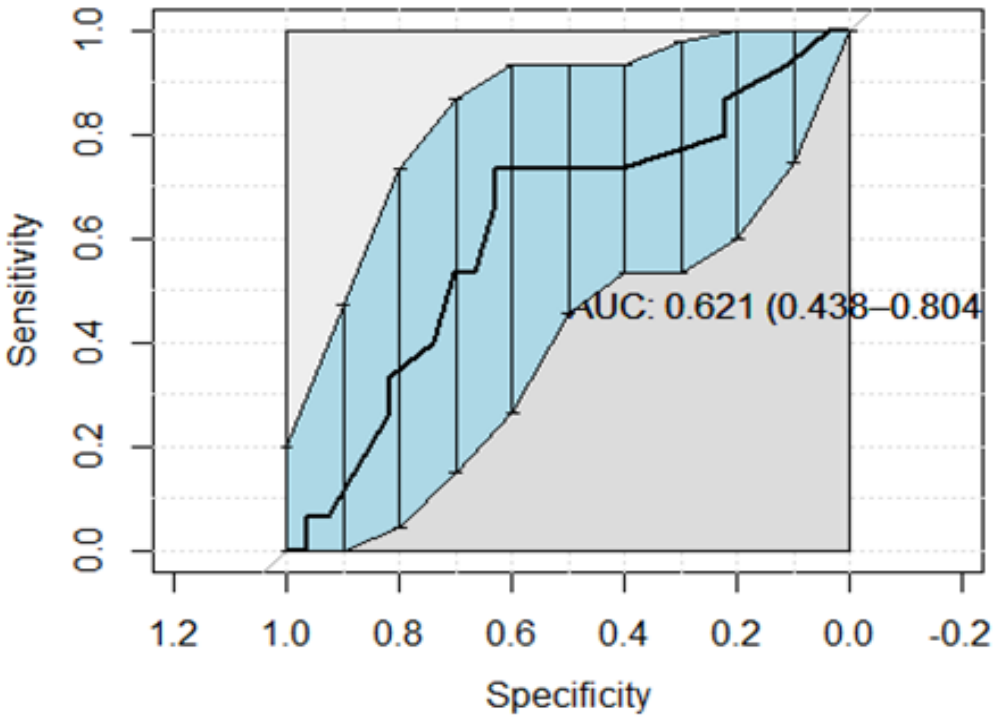

Figure 1: ROC Curve for Luminal A molecular subtype

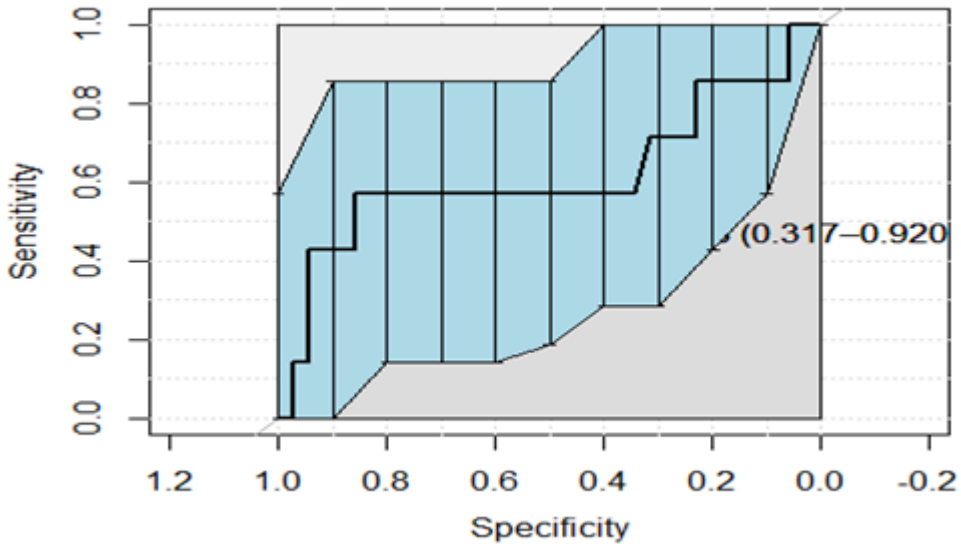

Figure 2: ROC curve for luminal B

likelihood estimation method and multiple equations. This implies that it may require a larger sample size compared to ordinal or binary logistic regression. Furthermore, if a cell has very few cases (a small cell), the model may become unstable or it might not even run at all. For that reason, we had to remove empty or small cells by doing a cross-tabulation between categorical predictors and the outcome variable.

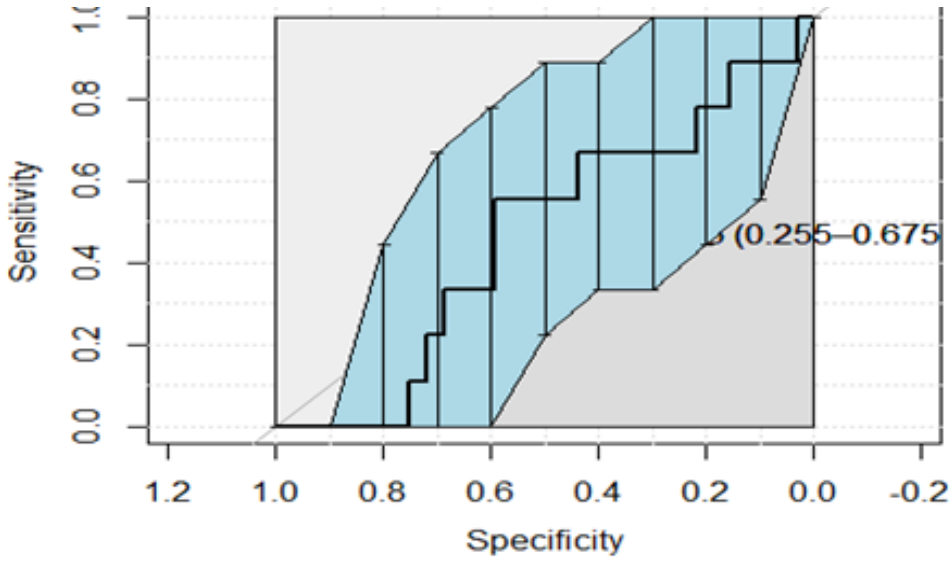

Figure 3: ROC curve for Her2 enriched subtype

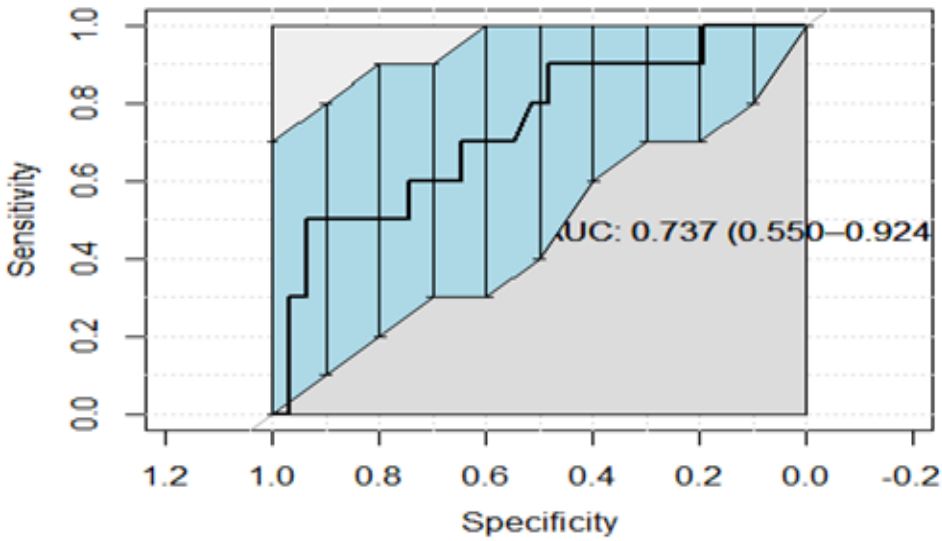

Figure 4: ROC curve for triple negative subtype

CONCLUSION

Breast cancer exhibited different clinical and histopathological predictors per molecular subtypes. Further studies are needed to explore the possibility of developing a clinical and histopathology-based scoring system which may help in clinical decision-making, especially in settings with scarce resources where access to immunohistochemistry testing is limited.

Availability of data and materials: The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

REFERENCES

1. Yang, X.R.; Pfeiffer, R.M.; Garcia-Closas, M.; Rimm, D.L.; Lissowska, J.; Brinton, L.A.; Peplonska, B.; Hewitt, S.M.; Cartun, R.W.; Mandich, D.; et al. Hormonal Markers in Breast Cancer: Coexpression, Relationship with Pathologic Characteristics, and Risk Factor Associations in a Population-Based

Study. Cancer Research 2007, 67, 10608–10617, doi:10.1158/0008-5472.CAN-07-2142.
2. Morrison, D.H.; Rahardja, D.; King, E.; Peng, Y.; Sarode, V.R. Tumour Biomarker Expression Relative to Age and Molecular Subtypes of Invasive Breast Cancer. British Journal of Cancer 2012, 107, 382–387, doi:10.1038/bjc.2012.219.
3. Nourmoussavi, M.; Pansegrau, G.; Popesku, J.; Hammond, G.L.; Kwon, J.S.; Carey, M.S. Ovarian Ablation for Premenopausal Breast Cancer: A Review of Treatment Considerations and the Impact of Premature Menopause. Cancer Treatment Reviews 2017, 55, 26–35, doi:10.1016/j.ctrv.2017.02.005.
4. Park, W.-C. Role of Ovarian Function Suppression in Premenopausal Women with Early-Stage Breast Cancer. Journal of Breast Cancer 2016, 19, 341–348, doi:10.1007/s11864-017-0442-8.
5. Masood, S. Breast Cancer Subtypes: Morphologic and Biologic Characterization. Women's Health 2016, 12, 103–119, doi:10.2217/whe.15.99.
6. Harvey, J.M.; Clark, G.M.; Osborne, C.K.; Allred, D.C. Estrogen Receptor Status. 1999, 17, 2.
7. Li, H.; Sun, X.; Miller, E.; Wang, Q.; Tao, P.; Liu, L.; Zhao, Y.; Wang, M.; Qi, Y.; Li, J. BMI, Reproductive Factors, and Breast Cancer Molecular Subtypes: A Case-Control Study and Meta-Analysis. Journal of Epidemiology 2017, 27, 143–151, doi:10.1016/j.je.2016.05.002.
8. Dai, X.; Li, T.; Bai, Z.; Yang, Y.; Liu, X.; Zhan, J.; Shi, B. Breast Cancer Intrinsic Subtype Classification , Clinical Use and Future Trends. 2015, 5, 2929–2943.
9. Polyak, K. Science in Medicine Breast Cancer : Origins and Evolution. Cell 2007, 117, 3155–3163, doi:10.1172/JCI33295.group.
10. Iqbal, J.; Ferdousy, T.; Dipi, R.; Salim, R.; Wu, W.; Narod, S.A.; Kotsopoulos, J.; Mostafa, M.G.; Ginsburg, O. Risk Factors for Premenopausal Breast Cancer in Bangladesh. 2015, 2015, doi:10.1155/2015/612042.
11. Weir, R.; Ali, W. Risk Factors for Breast Cancer in Women A Systematic Review of the Literature; 2007; Vol. 10; ISBN 9781877455117.
12. Dartois, L.; Fagherazzi, G.; Baglietto, L.; Delaloge, S. Proportion of Premenopausal and Postmenopausal Breast Cancers Attributable to Known Risk Factors : Estimates from the E3N-EPIC Cohort., doi:10.1002/ijc.29987.
13. Nelson, H.D.; Zakher, B.; Cantor, A.; Kerlikowske, K.; Ravesteyn, N.T. Van; Trentham-, A. NIH Public Access. 2013, 156, 635–648, doi:10.1059/0003-4819-156-9-201205010-00006.Risk.
14. Saadatmand, S.; Bretveld, R.; Siesling, S.; Tilanus-Linthorst, M.M.A. Influence of Tumour Stage at Breast Cancer Detection on Survival in Modern Times: Population Based Study in 173 797 Patients. BMJ (Online) 2015, 351, doi:10.1136/bmj.h4901.
15. Mwakigonja, A.R.; Lushina, N.E.; Mwanga, A. Characterization of Hormonal Receptors and Human Epidermal Growth Factor Receptor-2 in Tissues of Women with Breast Cancer at Muhimbili National Hospital, Dar Es Salaam, Tanzania. Infectious Agents and Cancer 2017, 12, 1–12, doi:10.1186/s13027-017-0170-5.
16. Sawe, R.T.; Kerper, M.; Badve, S.; Li, J.; Sandoval-Cooper, M.; Xie, J.; Shi, Z.; Patel, K.; Chumba, D.; Ofulla, A.; et al. Aggressive Breast Cancer in Western Kenya Has Early Onset, High Proliferation, and Immune Cell Infiltration. BMC Cancer 2016, 16, 1–15, doi:10.1186/s12885-016-2204-6.
17. Esteva, F.J.; Hortobagyi, G.N. Prognostic Molecular Markers in Early Breast Cancer. Breast Cancer Research 2004, 6, 109–118, doi:10.1186/bcr777.
18. Natrajan, R.; Weigelt, B. Risk Stratification and Intrinsic Subtype Classification of Breast Cancer: A Multiparameter Test to Rule Them All? Journal of the National Cancer Institute 2016, 108, 2016–2018, doi:10.1093/jnci/djw118.
19. Vanderpuye, V.; Grover, S.; Hammad, N.; Prabhakar, P.; Simonds, H.; Olopade, F.; Stefan, D.C. An Update on the Management of Breast Cancer in Africa. Infectious Agents and Cancer 2017, 12, 1–12, doi:10.1186/s13027-017-0124-y.
20. Sutter, S.A.; Slinker, A.; Balumuka, D.D.; Mitchell, K.B. Surgical Management of Breast Cancer in Africa: A Continent-Wide Review of Intervention Practices, Barriers to Care, and Adjuvant Therapy. Journal of Global Oncology 2017, 3, 162–168, doi:10.1200/JGO.2016.003095.
21. Awadelkarim, K.D.; Arizzi, C.; Elamin, E.O.M.; Hamad, H.M.A.; De Blasio, P.; Mekki, S.O.; Osman, I.; Biunno, I.; Elwali, N.E.; Barberis, M.C.; et al. Basal-Like Phenotype in a Breast Carcinoma Case Series from Sudan: Prevalence and Clinical/Pathological Correlations, Basal-Like Phenotype in a Breast Carcinoma Case Series from Sudan: Prevalence and Clinical/Pathological Correlations. Pathology Research International, Pathology Research International 2011, 2011, 2011, e806831,

doi:10.4061/2011/806831, 10.4061/2011/806831.
22. Salhia, B.; Tapia, C.; Ishak, E.A.; Gaber, S.; Berghuis, B.; Hussain, K.H.; DuQuette, R.A.; Resau, J.; Carpten, J. Molecular Subtype Analysis Determines the Association of Advanced Breast Cancer in Egypt with Favorable Biology. BMC Women's Health 2011, 11, doi:10.1186/1472-6874-11-44.
23. Galukande, M, :Wabinga, H; Asea, A. Molecular breast cancer subtypes prevalence in an indigenous sub-saharan african population.The panafrican medical journal.2014;17:249 doi:10.11604/pamj.2014.17.249.330
24. Odongo, J.; Makumbi, T.; Kalungi, S.; Galukande, M. Patient Delay Factors in Women Presenting with Breast Cancer in a Low Income Country. BMC Research Notes 2015, 4–9, doi:10.1186/s13104-015-1438-8.
25. Yersal, O.; Barutca, S. Ozlem Yersal, Sabri Barutca. World J Clin Oncol 2014, 5, doi:10.5306/wjco.v5.i3.412.
26. Boisserie-Lacroix, M.; Hurtevent-Labrot, G.; Ferron, S.; Lippa, N.; Bonnefoi, H.; Mac Grogan, G. Correlation between Imaging and Molecular Classification of Breast Cancers. Diagnostic and Interventional Imaging 2013, 94, 1069–1080.
27. Elidrissi Errahhali, M.; Elidrissi Errahhali, M.; Ouarzane, M.; El Harroudi, T.; Afqir, S.; Bellaoui, M. First Report on Molecular Breast Cancer Subtypes and Their Clinico-Pathological Characteristics in Eastern Morocco: Series of 2260 Cases. BMC Women's Health 2017, 17, 1–11, doi:10.1186/s12905-016-0361-z.
28. Turkoz, F.P.; Solak, M.; Petekkaya, I.; Keskin, O.; Kertmen, N.; Sarici, F.; Arik, Z.; Babacan, T.; Ozisik, Y.; Altundag, K. Association between Common Risk Factors and Molecular Subtypes in Breast Cancer Patients. The Breast 2013, 22, 344–350, doi:10.1016/J.BREAST.2012.08.005.
29. Huber, K.E.; Carey, L.A.; Wazer, D.E. Breast Cancer Molecular Subtypes in Patients With Locally Advanced Disease: Impact on Prognosis, Patterns of Recurrence, and Response to Therapy. Seminars in Radiation Oncology 2009, 19, 204–210.
30. Hemminki, K.; Mousavi, S.M.; Sundquist, J.; Brandt, A. Does the Breast Cancer Age at Diagnosis Differ by Ethnicity? A Study on Immigrants to Sweden. The Oncologist 2011, 16, 146–154, doi:10.1634/theoncologist.2010-0104.
31. Maru, D.; Middleton, L.; Wang, S.; Valero, V.; Sahin, A. HER-2/Neu and P53 Overexpression as Biomarkers of Breast Carcinoma in Women Age 30 Years and Younger. Cancer 2005, 103.
32. Heer, E.; Harper, A.; Escandor, N.; Sung, H.; McCormack, V.; Fidler-Benaoudia, M.M. Global Burden and Trends in Premenopausal and Postmenopausal Breast Cancer: A Population-Based Study. The Lancet Global Health 2020, 8, e1027–e1037, doi:10.1016/S2214-109X(20)30215-1.
33. Dietze, E.C.; Sistrunk, C.; Miranda-Carboni, G.; O'Regan, R.; Seewaldt, V.L. Triple-Negative Breast Cancer in African-American Women: Disparities versus Biology. Nature Reviews Cancer 2015, 15.
34. Stark, A.; Kleer, C.G.; Martin, I.; Awuah, B.; Nsiah-Asare, A.; Takyi, V.; Braman, M.; E. Quayson, S.; Zarbo, R.; Wicha, M.; et al. African Ancestry and Higher Prevalence of Triple-Negative Breast Cancer. Cancer 2010, 116, 4926–4932, doi:10.1002/cncr.25276.
35. Brinton, L.A.; Figueroa, J.D.; Awuah, B.; Yarney, J.; Wiafe, S.; Wood, S.N.; Ansong, D.; Nyarko, K.; Wiafe-Addai, B.; Clegg-Lamptey, J.N. Breast Cancer in Sub-Saharan Africa: Opportunities for Prevention. Breast Cancer Research and Treatment 2014, 144.
36. Awadelkarim, K.D.; Arizzi, C.; Omer, E.; Elamin, M.; Hamad, H.M.A.; Blasio, P. De; Mekki, S.O.; Osman, I.; Biunno, I.; Elwali, N.E.; et al. Basal-Like Phenotype in a Breast Carcinoma Case Series from Sudan : Prevalence and Clinical / Pathological Correlations. 2011, 2011, doi:10.4061/2011/806831.
37. Sengal, A.T.; Haj-mukhtar, N.S.; Elhaj, A.M.; Bedri, S.; Kantelhardt, E.J.; Mohamedani, A.A. Immunohistochemistry Defined Subtypes of Breast Cancer in 678 Sudanese and Eritrean Women ; Hospitals Based Case Series. 2017, 1–9, doi:10.1186/s12885-017-3805-4.
38. Eng, A.; McCormack, V.; dos-Santos-Silva, I. Receptor-Defined Subtypes of Breast Cancer in Indigenous Populations in Africa: A Systematic Review and Meta-Analysis. PLoS Medicine 2014, 11, doi:10.1371/journal.pmed.1001720.
39. Anderson, a S.; Armes, J.; Berditchevski, F.; Blaydes, J.P.; Brennan, K.; Brown, N.J.; Bryant, H.E.; Bundred, N.J.; Burchell, J.M.; Campbell, a M.; et al. Critical Research Gaps and Translational Priorities for the Successful Prevention and Treatment of Breast Cancer. Breast cancer research : BCR 2013, 15, R92, doi:10.1186/bcr3493 [doi].
40. Rugwizangoga, B.; Ndayisaba, M.; Izimukwiye, A.; Baryabagaya, J.D.D.; Bosco, J.; Nsenguwera, A.; Niyikora, N. Accuracy of the Duration of Tissue Fixation and the Receptor Status Profile of Primary

Breast Cancers in a Tertiary Hospital of Rwanda. 2016, 756–762.
41. Gray, J.M.; Rasanayagam, S.; Engel, C.; Rizzo, J. State of the Evidence 2017: An Update on the Connection between Breast Cancer and the Environment. Environmental Health: A Global Access Science Source 2017, 16, 1–61, doi:10.1186/s12940-017-0287-4.
42. Rosner, B.; Eliassen, A.H.; Michels, K.B.; Willett, W.C. Subtype in the Nurses ' Health Studies. 2017, 9, 732–738, doi:10.1158/1940-6207.CAPR-16-0085.Height.