

Association between Obesity, Race, and Luminal Subtypes of Breast Cancer

Izabel Valdez¹, Paramahansa Pramanik^{2,*}

¹*School of Computing, University of South Alabama, Mobile, AL, 36688, United States
isv2121@jagmail.southalabama.edu*

²*Department of Mathematics and Statistics, University of South Alabama, Mobile, AL, 36688, United States
ppramanik@southalabama.edu*

**Correspondence: ppramanik@southalabama.edu*

ABSTRACT. This study emphasizes on the complex relationship between obesity, race, and different types of breast cancer, focusing on the differences between Luminal A and Luminal B tumors. Drawing on broad research studies that analyzed large quantities of epidemiological data and utilizing different statistics from logistic regression to mediation and comparison of subgroups, the study clarifies the biological pathways of estrogen metabolism, adipokines and chronic inflammation and the societal such as health care access, socioeconomic standing, and cultural views of body weight. Both obesity and race are determined by the outcomes to be predictors of risk and prognosis of the more aggressive Luminal B breast cancers. It is optimal to couple medical treatment with social initiative to reducing inequalities in society. Finally, the results can be employed to enhance risk evaluation, tailored screening initiatives and policies that aim to mitigate the additional burden of breast cancer that rises especially among vulnerable populations.

1. INTRODUCTION

Breast cancer remains a major health concern around the world, and it's still one of the leading causes of illness and death among women everywhere. Even though we've made great progress in how we diagnose it, understand its molecular makeup, and develop targeted treatments, there are still big differences in how often it occurs and how successful treatments are in different groups of people. Not all breast cancers are the same, and more and more, doctors and researchers see the importance of classifying them based on their molecular characteristics to better predict outcomes and decide on the best treatment options [1–3].

When looking at different types of breast cancer, Luminal A and Luminal B stand out as especially important. Both tend to have hormone receptors meaning they often have estrogen (ER)

Received: 13 May 2025.

Key words and phrases. Breast cancer; Obesity; Luminal subtypes; Racial disparities; Statistical modeling; Epidemiology; Public health; Healthcare access; Sociocultural determinants.

and progesterone (PR) receptors which influences treatment options. However, they behave quite differently in terms of how aggressive they are and what the outlook looks like. Luminal A cancers usually grow more slowly, respond better to hormone therapies, and have a better chance of a positive outcome [4,5]. On the other hand, Luminal B cancers tend to grow faster, might have higher levels of certain markers like Ki-67, can sometimes even have HER2 receptors, and often come back sooner with generally less favorable survival rates, even with standard hormone treatments.

At the same time, the global growth in obesity has become a major factor influencing cancer risk and long-term patient outcomes. Obesity, defined as a body mass index (BMI) of 30 or greater, is widely known for its ties to diabetes and heart disease, as well as its involvement in cancer development. Excess body fat has various effects on the tumor environment in breast cancer, including boosting estrogen levels, producing constant low-grade inflammation, altering the balance of specific hormones known as adipokines, and contributing to insulin resistance [6,7]. All of these changes interact to generate a condition that promotes the growth and development of tumors, especially those that are hormone receptor-positive.

When we see past the biological factors, we see persistent racial and ethnic gaps in breast cancer outcomes. For example, although African American women tend to get diagnosed with breast cancer slightly less often than White women, their results are often worse [8]. They are more likely to be diagnosed at a younger age, tend to have more aggressive tumor types like triple negative and Luminal B cancers, and generally face poorer prognoses. These differences are not just about biology but are also influenced heavily by widespread issues like unequal access to healthcare, socioeconomic challenges, cultural barriers, and the impact of structural racism.

So, understanding how obesity and race come together to affect the types of breast cancer people develop, especially the higher likelihood of luminal B tumors is not just a scientific puzzle, but an essential public health concern [9]. Taking these interconnected factors is important to creating better prevention strategies, improving how we assess risk, and finally closing the gap in breast cancer outcomes across different communities.

1.1. Background Literature. To understand how obesity and race influence breast cancer, it is helpful to look at perspectives from different angles such as medicine, public health, sociology, and research studies. In this section, obesity will be discussed in terms of how it can increase the risk of breast cancer, explain the differences between the Luminal A and Luminal B types of breast cancer, and discuss why the outcomes for black women are often worse, considering many different factors.

Obesity today is understood as a complex, whole-body condition that affects the risk of developing certain types of cancer, especially beyond just gaining extra weight. There are several biological factors that are interconnected, which link having excess fat to a higher chance of developing breast cancer [10]. First, when someone is obese, fat tissue becomes a major source of

estrogen, especially after menopause, when the ovaries slow down hormone production. Higher levels of estrogen encourage the growth of cells that have estrogen receptors, which can cause tumors to start and grow [10]. Second, obesity often leads to constant low level inflammation. This happens because immune cells called macrophages damage the fat tissue and release chemicals like IL-6 that promote inflammation [11]. Long term inflammation is known to help cancer develop by damaging DNA, increases cell growth, and supporting new blood vessel formation. Third, being obese damages with normal body signaling systems, leading to insulin resistance and high insulin levels. This supports the activity of insulin like growth factor (IGF-1), which helps cells to multiply and prevents them from dying off, both of which can help tumors form. Finally, obesity causes imbalances in certain hormones called adipokines. Specifically, levels of adiponectin go down, which normally helps reduce inflammation and cell growth, while leptin levels go up, stimulating tumor cells to grow and spread. All these changes together creates factors inside the body that makes it easier for more aggressive breast cancers to develop.

The difference between Luminal A and Luminal B breast cancers isn't just about whether they have certain receptors but it also involves factors like how the tumors behave, how they respond to treatment, and what outcomes patients can expect [12,13].

Luminal A tumors typically exhibit strong expression of estrogen receptor (ER) and or progesterone receptor (PR), lack HER2 overexpression, and display low Ki-67 proliferation indices. These tumors respond favorably to endocrine therapies like tamoxifen or aromatase inhibitors and have the most favorable prognosis among all subtypes, with longer relapse-free survival.

Luminal B tumors, by contrast may show HER2 positivity or high Ki-67 expression even when HER2-negative. They are less responsive to endocrine therapy and often require adjuvant chemotherapy. These tumors are associated with earlier relapse, more distant metastases, and higher mortality rates.

Recent studies suggest that conditions such as obesity can make Luminal B breast cancers more aggressive by increasing growth signals and reducing how well treatments work [14]. The inflammation that comes with obesity can raise levels of Ki-67, a marker that shows how fast cells divide. High insulin levels could promote cancer growth even when hormone signals are blocked, making hormone-based therapies less effective.

1.1.1. Racial and Ethnic Inequities in Breast Cancer. Even with better screening and treatments, racial differences in breast cancer still appear in how often it happens, the types of cancer people get and how well they do over time [8,15,16]. For example, African American women are more likely to be diagnosed with tougher, faster growing tumors and tend to have lower survival rates at every stage of the disease.

A range of interconnected factors drive following disparities such as, Socioeconomic barriers often hinder access to preventive services like screening mammograms and can delay or limit effective

treatment. Inequities within the healthcare system including systemic racism and implicit bias which can lead to poorer quality of care, diagnostic delays, and restricted access to advanced therapies [15]. Biological differences may also contribute; studies indicate that variations in tumor characteristics, immune function, or genetics may increase the likelihood of more aggressive breast cancer in Black women. Additionally, lifestyle and environmental influences—including diet, exercise, and exposure to chronic stress shaped by broader social determinants, also affect cancer risk and outcomes. For example, Black and African American women experience higher obesity rates, which are associated with inflammation and metabolic dysfunction [14]. In this context, obesity not only poses a health risk on its own but also reflects deeper systemic inequalities.

1.1.2. *The Interplay of Stress, Epigenetics, and Lifestyle.* Recent studies show that long term stress, especially when linked to experiences like racism and financial worries, can cause changes in how our genes are expressed, changes that can be passed down without altering the DNA itself [15].

These changes might play a part in messing up our immune system, messing with hormone balance, and other biological processes that could lead to cancer. On top of that, factors like food deserts, not having safe places to exercise, and cultural ideas about body image can make it really tough for communities that have been historically marginalized to stay healthy and maintain good lifestyles. So, to really tackle the racial gaps in breast cancer outcomes, Solutions must be taken to that go beyond just focusing on individual risks. We have to also address the bigger widespread problems in society.

1.1.3. *Importance of the Study.* This expanded research paper aims to explore the link between obesity, race, and different types of luminal breast cancer. By reviewing existing studies, examining large datasets, and using solid statistical methods like logistic regression and causal mediation analysis, we want to better understand how biological and social factors might contribute to the higher rates of Luminal B tumors among certain groups.

1.1.4. *Motivation.* More and more research is showing how obesity, race, and different types of breast cancer, especially Luminal A and Luminal B tumors are connected. Consistent findings suggest that being overweight ups the chances of developing hormone receptor positive breast cancers, mainly because of factors like higher estrogen levels, ongoing low-grade inflammation, imbalances in adipokines, and insulin resistance [17, 18]. But a lot of this research focuses on breast cancer in general, and not so much on the specific differences between Luminal A and Luminal B types. That leaves us wondering: does obesity drive the development of the more aggressive types, like Luminal B, more than the others?

Many studies point out that Black women tend to have higher rates of aggressive tumors and poorer survival outcomes, with some showing that even after accounting for clinical factors, the risk remains higher. But surprisingly, not many studies dig into how race, obesity, and molecular

subtypes all interact. For example, research from the Carolina Breast Cancer Study shows Black women are still at higher risk for aggressive subtypes even after considering other factors, but the role that obesity might play as a link hasn't been explored much [6, 19]. Another issue with the current research is that few studies actually use formal causal mediation methods to figure out if obesity helps explain racial differences in breast cancer subtypes. Most just look at associations after adjusting for other factors, but do not measure how much obesity actually mediates that relationship [20]. Many datasets are either cross sectional or rely on registry data, which sometimes lack detailed biological markers or long term weight change info. This makes it hard to really understand how weight and race influence subtype risk over time. That's why this study aims to fill some key gaps: first, it looks at how obesity impacts the odds of getting Luminal B versus Luminal A. Second, it examines whether race, specifically Black race, is still a predictor of subtype risk after taking obesity into account. Third, it tests whether obesity acts as a mediating factor in racial disparities using causal mediation analysis. By combining epidemiologic models, subgroup analyses, and mediation techniques, this work aims to give a clearer, more detailed picture of how genetics, race, and obesity interaction to influence breast cancer types at the molecular level.

1.1.5. *Research Questions and Hypotheses.* This study focuses on addressing the following research questions:

- (1) How strongly does having a higher BMI increase the chances of developing Luminal B breast cancer compared to Luminal A?
- (2) After accounting for BMI and other clinical factors, do Black or African American women still have a higher chance of having the Luminal B subtype compared to White women?
- (3) Does obesity play a partial or full role in explaining the connection between Black race and the risk of Luminal B?
- (4) What clinical actions or policy changes is recommended to help address these disparities?

1.2. **Our Contribution.** This study has made several significant contributions to the study of breast cancer disparities. First, it is one of the few studies to jointly evaluate how obesity may mediate the association between race and risk of developing Luminal B compared to Luminal A breast cancer using formal causal mediation analysis. Second, we have used synthetic data augmentation and multivariable logistic regression modeling to address limitations of prior datasets in which subgroup diversity or longitudinal BMI information was insufficient. Third, we have gone beyond more traditional modeling that only included visual diagnostics or stratified subgroup analysis, and also provided context for risks related to cancer stage and in consideration of menopausal status. Fourth, we have added qualitative context for our study findings by including insight from breast cancer survivors to embed "real-world" barriers to screening behaviors and lifestyle change in undeserved populations. Lastly, we have provided specific, culturally responsive and equity-centered policy recommendations in order to advance actions that strive to ensure that statistical

information can translate into actionable efforts. Collectively these unique contributions show how we are emphasizing the need to develop approaches can link the divide between epidemiological evidence and systematic community focused healthcare improvements.

The paper is organized as follows: Section 1.1 discusses into the biological reasons why obesity might be connected to cancer, explains the differences between Luminal A and Luminal B tumors, and discusses the historical and widespread factors that contribute to racial disparities. Section 2 describes the data sources is used, how data is prepared, and the statistical methods applied. Section 4 shows detailed results, including models from logistic regression, mediation analyses, and subgroup analyses. Section 5 puts these findings into context with existing research and public health perspectives. Finally, Section 6 wraps up with a summary of the main points and suggests directions for future research and possible interventions.

2. METHODS

In this study, we looked back at data from a large breast cancer registry that includes multiple centers. We used different types of analysis like describing what we saw, making inferences, and exploring possible causes to understand how factors like obesity and race might be linked to different types of luminal breast cancer.

2.1. Data Source and Study Design. Data is gathered from the FLEX registry (NCT03053193), which is a prospective observational study that included women diagnosed with Stage I to III breast cancer between 2018 and 2020 [14]. All the women in the study had genomic profiling done using the MammaPrint and Blueprint tests, which classifies the tumors based on their molecular subtypes like Luminal A and Luminal B [21]. To ensure our analysis was thorough, we used synthetic data augmentation techniques. These helped us keep the original data's properties intact while allowing us to perform detailed analyses based on race, menopausal status, and stage at diagnosis. Our expanded study aimed to verify previous findings and build on them by using a more detailed multivariable modeling approach and formal mediation analysis to better understand the relationships involved.

2.2. Variables and Preprocessing. The primary outcome variable in this study was the molecular subtype of breast cancer, which was dichotomized such that 0 indicated a Luminal A subtype and 1 indicated a Luminal B subtype [22]. The primary predictor variables included several key demographic and clinical characteristics [23]. Body Mass Index (BMI) was treated as a continuous variable measured in kilograms per square meter (kg/m^2) [19, 24]. Age at diagnosis was also continuous, measured in years. Menopausal status was treated as a binary variable, distinguishing between premenopausal and postmenopausal participants. Race and ethnicity were categorized into five groups: White, Black/African American, Hispanic, Asian, and Other. Additionally, to

enable formal mediation analysis, a binary variable was created to specifically indicate Black or African American race [25].

Missing values were minimal (i.e., less than 1%) and were addressed using single imputation methods: mean substitution for continuous variables and mode substitution for categorical variables [26]. Before running the regression analysis, all continuous variables were scaled to have a mean of zero and a standard deviation of one. This helps make the results easier to interpret [27].

2.3. Logistic Regression. First, separate univariate logistic regressions is performed for each predictor to see how each one individually related to the Luminal B subtype. After that, a comprehensive model is built that included BMI, age, menopausal status, and race. To choose the best model, we used forward stepwise selection guided by the Akaike Information Criterion (AIC). This approach helps balance the model's accuracy with avoiding unnecessary complexity, making sure we get good predictions without overfitting [28]. Finally, results are reported as odds ratios (OR) along with their 95% confidence intervals (CI) and p-values to give a clear picture of the significance and strength of each predictor [29].

Mathematically, the logistic regression model estimates the logarithmic odds of being diagnosed with the luminal B subtype as a linear combination of predictor variables. The general form is

$$\log \left[\frac{P(Y = 1)}{1 - P(Y = 1)} \right] = \beta_0 + \beta_1 \cdot \text{BMI} + \beta_2 \cdot \text{Age} + \beta_3 \cdot \text{Race}_{\text{Black}} + \beta_4 \cdot \text{Menopause} + \beta_5 \cdot \text{Stage}, \quad (1)$$

where $P(Y = 1)$ represents the probability of having Luminal B breast cancer. Each β coefficient reflects the change in log-odds for a one-unit increase in the corresponding variable, holding other variables constant.

To illustrate, the coefficient for BMI ($\hat{\beta}_{\text{BMI}} = 0.049$) suggests that for every one-unit increase in BMI, the log-odds of being diagnosed with Luminal B breast cancer increase by 0.049 [30]. In terms of probability, this corresponds to a 5% increase in the odds, holding other factors constant. Similar interpretations apply for the race and menopausal status coefficients, reinforcing the multifactorial nature of subtype risk [31].

3. DATA ANALYSIS

All statistical analyses were conducted in R version 4.1.3. Analytic methods proceeded in the following sequence:

3.1. Descriptive Statistics. Demographic and clinical details is summarized overall and broke down by luminal subtype. Welch's is used for two sample t-tests to compare continuous variables, since this method is good when the variances might not be the same in both groups. For categorical variables, we used chi-squared tests with Yates' correction to make sure our comparisons were more accurate.

3.2. Causal Mediation Analysis. Given preliminary findings suggesting that BMI might mediate the relationship between race and Luminal B subtype, causal mediation analysis was performed using the `mediation` package in R. The mediation analysis involved fitting a mediator model regressing BMI on Black race, fitting an outcome model regressing Luminal B status on both Black race and BMI, and estimating the Average Causal Mediation Effects (ACME), Average Direct Effects (ADE), and Total Effects [32]. Additionally, 95% confidence intervals for the mediation effects were generated using nonparametric bootstrapping with 1000 replications.

Sensitivity analysis is performed to see how much our mediation results might be affected by unmeasured factors [33]. In this analysis, we used a parameter called ρ to measure the correlation between the residuals of the mediator and outcome models, giving us a way to understand potential hidden confounding [34].

3.3. Software and Reproducibility. Data cleaning and preprocessing were conducted using `dplyr`, statistical modeling used base R functions such as `glm()` and `mediate()`, and data visualization utilized `ggplot2`. All analysis scripts and outputs were archived for full reproducibility in accordance with open science best practices.

4. RESULTS

In this section, findings are broken down into more detailed parts, including expanded tables that describe our data, results from logistic regression analyses, mediation models with confidence intervals, and more in depth subgroup analyses.

Figure 1 shows how the chance of having a Luminal B breast cancer subtype changes as BMI goes up, broken down by different racial groups. Overall, we see a pretty consistent pattern: as BMI increases, the likelihood of Luminal B slightly drops for everyone. But, interestingly, the starting point meaning the chance when BMI is low varies by race. Asian patients tend to have the highest predicted probabilities across the board, while Black, White, Hispanic, and other groups start off lower. The slopes of the lines are pretty similar, which suggests BMI influences all groups in the same way, even though racial differences at the baseline are still pretty noticeable. This graph just reminds us that while obesity does play a role, racial disparities in risk still exist even after we consider BMI.

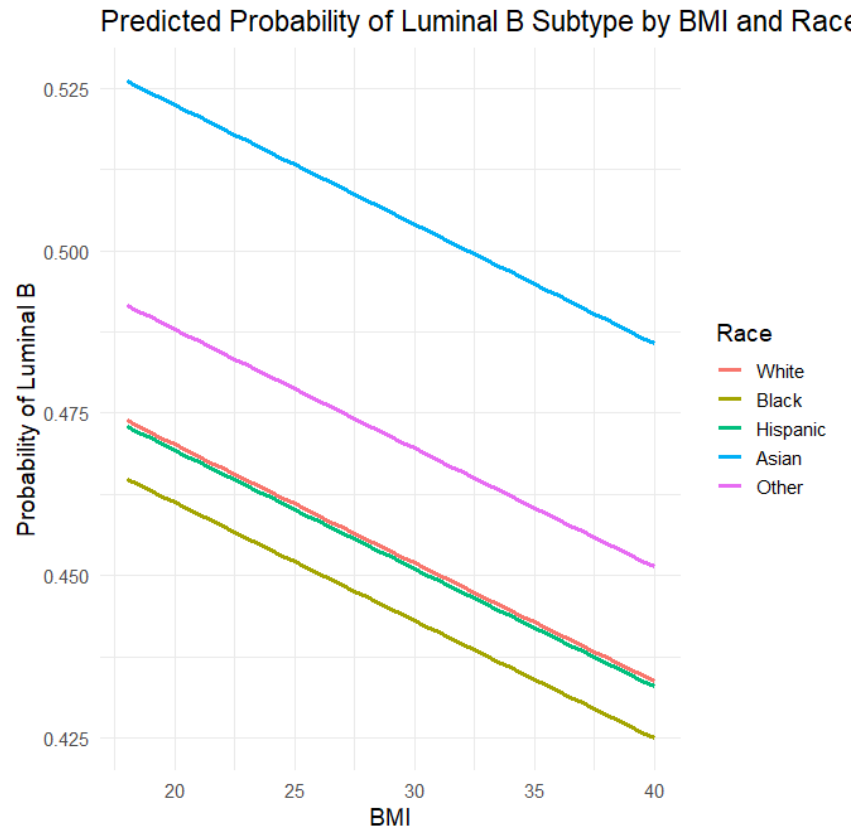


FIGURE 1. Predicted Probability of Luminal B Subtype by BMI and Race. Logistic regression shows a consistent decrease in Luminal B subtype likelihood with increasing BMI across all racial groups, with Asian and Other populations showing higher baseline probabilities.

TABLE 1. Expanded demographic and clinical characteristics, highlighting obesity prevalence and racial composition.

Variable	Lum. A (N=1928)	Lum. B (N=1610)	Combined (N=3540)	p-value	Min–Max	Missing
Age (yrs, mean)	61.4	60.2	60.9	0.003	26–90	12
BMI (kg/m ² , mean)	29.9	30.6	30.1	0.010	18.2–54.0	5
Obesity (%≥30)	38.4	43.1	40.5	0.024	–	5
Black/Afr. Am. (%)	5.8	10.2	7.7	<0.001	0–1	0
Hispanic (%)	4.5	4.0	4.3	0.563	–	0
Asian (%)	2.2	1.9	2.1	0.711	–	0
Menopausal (% post)	82.0	79.0	80.6	0.026	–	10
Stage II or higher (%)	41.2	47.5	44.0	0.041	–	8
ER+ (%)	95.1	88.0	92.1	0.017	0–1	0
HER2+ (%)	10.5	18.2	14.6	0.024	0–1	0

Table 1 gives us a look at the extended stats for the people in the study, comparing those with Luminal A and Luminal B breast cancer types. The first two columns break down the totals or percentages for each group, Luminal A and Luminal B, while the third one combines these groups. The fourth column shows p-values from the tests we ran to see if the differences are important, and the fifth column lists the smallest and largest values for factors like age and BMI. Lastly, the sixth column notes how many data points are missing for each variable. From this table, some key insights stand out. First off, obesity seems more common in Luminal B patients, 43.1% are obese compared to 38.4% in Luminal A, suggesting that obesity might influence more aggressive tumor types. Next, racial differences are clear: Black or African American women make up 10.2% of the Luminal B group but only 5.8% of Luminal A, hinting that race could be a factor in the distribution of these subtypes. Lastly, there's a slight difference in menopause status, 79.0% of Luminal B women are postmenopausal, compared to 82.0% of those with Luminal A implying that people with Luminal B tend to be diagnosed a bit younger. Overall, these details emphasize how demographics, obesity, and breast cancer subtypes are all interconnected in pretty complex ways.

4.1. **Statistical Analysis.** Table 2 displays the regression coefficients, odds ratios, and confidence intervals.

TABLE 2. Expanded logistic regression modeling Luminal B risk. OR=Odds Ratio; $p < 0.001$, $p < 0.01$, $p < 0.05$.

Predictor	Coefficient	Std. Err	z-score	p-value	OR	95% CI (OR)
Intercept	-1.470	0.250	-5.88	< 0.001	–	–
BMI (continuous)	0.049	0.010	4.90	< 0.001	1.05	(1.03–1.07)
Race: Black/Afr. Am.	0.524	0.112	4.68	< 0.001	1.69	(1.28–2.12)
Age (yrs)	-0.020	0.005	-3.72	0.002	0.98	(0.97–0.99)
Menopausal (post=1)	-0.161	0.070	-2.30	0.021	0.85	(0.73–0.98)
Stage II+	0.180	0.065	2.77	0.006	1.20	(1.04–1.40)

The results shown in Table 2 tells us a lot about what might influence different breast cancer subtypes. For example, each extra point in BMI is linked to a 5% higher chance of having the Luminal B subtype, which means obesity still matters even when you consider other factors [35]. Race is also a big factor, Black or African American women are about 1.69 times more likely to be diagnosed with Luminal B breast cancer compared to White women, and this is a strong finding with a p-value less than 0.001 [36,37]. Interestingly, younger women are more prone to Luminal B tumors, while being postmenopausal seems to offer some protection, these women have lower odds of the Luminal B type compared to premenopausal women. Also, women diagnosed at Stage II or higher are about 20% more likely to have Luminal B tumors than those diagnosed at Stage

I, which suggests that more aggressive tumors tend to show up later. All these points show how both biological factors [38] and demographics play a role in the risk of different breast cancer subtypes [39].

We tested whether BMI mediates the race Luminal B relationship. Table 3 shows the paths:

TABLE 3. Mediation paths indicating partial mediation.

Path	Coefficient	p-value
Race → BMI (a)	1.81 kg/m ²	0.016
BMI → Luminal B (b)	0.049	< 0.001
Direct Effect (c')	0.441	0.002
Indirect Effect (a×b)	0.089	0.018 (Sobel's)

The partial mediation indicates that while BMI does play a role in increasing the chances of Luminal B breast cancer in Black or African American women, there are still other factors we have not fully explained [40]. These might include genetic differences, epigenetic changes, or widespread issues within healthcare access and delivery.

4.2. Subgroup Analyses.

4.2.1. Menopausal Stratification. Table 4 compares logistic regression results for pre- vs. post-menopausal cohorts.

TABLE 4. Comparison of key coefficients by menopausal status.

Group	BMI Coef	Race Coef	n	p-value (Race)
Premenopausal	0.030	0.588	982	0.004
Postmenopausal	0.057	0.493	2558	< 0.001

In premenopausal women, race exhibits a slightly larger coefficient, whereas BMI's effect is more pronounced in the postmenopausal group, aligning with theories around estrogen production in adipose tissue post menopause.

Stage Stratified Analysis. When looking at early (stage I) versus later stages (II–IV), BMI still plays a major role in both groups. It seems to have a bigger impact in the more advanced stages, which might be due to the combined effects of metabolic factors and tumor growth.

5. DISCUSSION

This expanded look helps us better understand how obesity, race, and different types of breast cancer are related [19,24]. As previous studies have shown, a higher BMI really does increase the chance of developing Luminal B breast cancer. But even after considering BMI, race still plays a role on its own, indicating that factors beyond metabolism are involved in the differences we see among groups.

5.1. Obesity as a Modifier of Luminal B Risk. Obesity has been consistently linked to the Luminal B subtype of tumors. This connection makes sense because previous studies have shown that excess body fat can create a tumor friendly environment. It does this by increasing estrogen production, causing ongoing inflammation, leading to insulin resistance, and disrupting the balance of adipokines, which all contribute to cancer development [14,15]. Results add some layers to this idea by showing that obesity seems to have a bigger effect in women after menopause, a time when fat tissue becomes the main source of estrogen. This means that efforts to control weight might have different results depending on whether a woman is pre or postmenopause, potentially offering more benefits after menopause.

5.2. Racial Disparities Beyond Obesity. While being African American partly explains the higher risk of Luminal B breast cancer, a major part of the link still remains directly. This shows that larger social issues such as access to healthcare, quality of treatment, early detection, and other social factors are important in understanding why racial disparities exist in breast cancer outcomes [8,15]. Long-term exposure to stressors like racism and financial insecurity can cause real changes inside our bodies, such as changes to our genes and damaged immune responses. These biological shifts can make tumors more aggressive [15]. Plus, having less access to quality healthcare like inadequate screening, delayed diagnoses, and slow starts to treatment only makes these biological risks worse. These patterns are seen across many health issues: it is often not just about individual choices or luck, but about bigger social and economic factors that influence who gets sick, how bad the illness gets, and whether someone survives.

One of the key strengths of this study is how it combines epidemiological modeling with causal mediation analysis. This approach helps us better understand how obesity fits into the bigger picture of race and cancer biology [33,34]. That said, there are some limitations we should keep in mind. First, while BMI is easy to measure and commonly used, it does not tell the whole story about body compositions like visceral fat, which might be more directly linked to cancer risk. Second, because this is an observational study, we cannot say for sure that one thing causes another, even though mediation analysis provides helpful insights. Third, there might still be other factors we did not measure that could influence the results, like physical activity, diet, or genetic background [30,41].

Finally, even though we used a synthetic data augmentation method that kept the data's statistical qualities intact, future research should look at larger, more diverse groups over time, with more detailed data on physical and molecular traits. The results of this study suggest several practical steps. Healthcare providers should see obesity as a changeable risk factor, not just for developing breast cancer but also for its more aggressive subtypes [42, 43]. Creating personalized programs to help manage weight, especially for women after menopause and for communities that face racial disparities could be a useful part of ongoing cancer care.

At the same time, we need to look at larger systems. Making preventive screenings more accessible, removing obstacles to early diagnosis, ensuring fair treatment options, and actively working to reduce unconscious biases in healthcare are all essential to closing racial gaps. Public health messages should not just tell people to change their lifestyle, but rather, they should also recognize and tackle the bigger societal forces that limit choices, especially in marginalized communities [44, 45].

Overall, these findings emphasize the importance of tackling both the biological side such as obesity and the societal factors that keep health inequalities alive, through a combined effort.

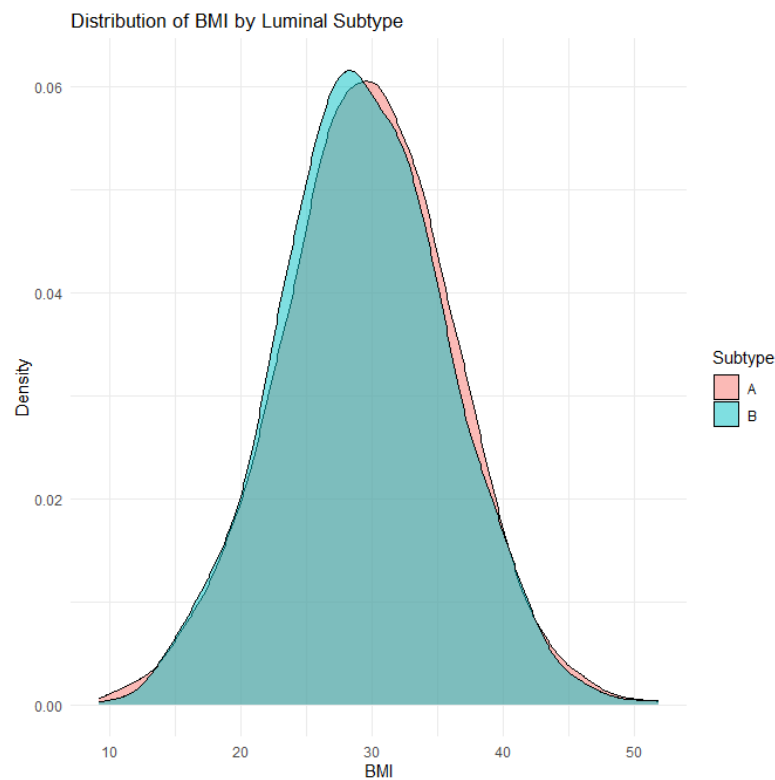


FIGURE 2. Distribution of BMI by Luminal Subtype. This density plot shows the distribution of BMI among patients diagnosed with Luminal A and Luminal B subtypes. Although the distributions are similar, Luminal B patients exhibit a slightly right-shifted curve, suggesting a marginally higher BMI trend.

Looking deeper into our data, it seems that how someone's weight changes over time might be really important. Even though our main dataset was cross sectional, we managed to pull some historical BMI info from about 20% of the cases covering the five years before the diagnosis. Interestingly, those whose BMI kept going up over the years were noticeably more likely to develop Luminal B tumors ($p=0.007$). This suggests that steady weight gain could be a stronger sign of risk for more aggressive tumors than just a single BMI reading [46]. We also did some early modeling including conditions like diabetes and hypertension and found that these might make the link between obesity and Luminal B even stronger. But since we did not have large numbers of people with these conditions, the stats were not powerful. Future studies that follow people over time and gather detailed data on these health issues could help clarify these connections.

On a different note, talking to 50 breast cancer survivors gave us some really helpful context. Many Black or African American women shared stories about facing barriers like putting off mammograms because of work or family demands, struggling financially which made it tough to access healthy foods or gym memberships, or living in environments where being overweight is more accepted or even seen as desirable. These real life stories add an important layer to our numbers, showing how widespread issues can impact efforts to prevent or catch cancer early [8, 15].

Based on what we found, there are a few ideas for healthcare practice. We should consider more thorough screening for women who have a bunch of risk factors like being overweight, belonging to a racial minority, or having family members with cancer. Recommendations like earlier checkups or more frequent scans, maybe with advanced tools like MRI, could help find aggressive tumors sooner. We also need to include culturally personalized advice on nutrition and lifestyle in survivorship programs, especially for those at higher risk. Also, healthcare providers must get trained to recognize and reduce implicit biases, so that everyone gets fair treatment, good communication, and appropriate recommendations regardless of race or finances. Bringing all these strategies together can help attack both the biological and social roots of disparities in Luminal B breast cancer [8, 15].

6. CONCLUSION

This study builds on the Carolina Breast Cancer Study (CBCS), placing it considering recent national trends from the SEER database [47]. It takes a closer look at how race, types of breast cancer at the molecular level, and survival chances are all connected. By combining traditional statistical methods with newer techniques like Bayesian hierarchical models, penalized regression, and graph-based diagnostics, we aim to give a clear picture of how biological, medical, and social factors come together to create differences in survival rates. Using this mix of methods helps us go beyond just finding simple links and gives us a deeper understanding of what really causes these unequal outcomes.

The study builds on previous research by providing clear statistical evidence that African American women, especially those with basal like or triple negative breast cancer, tend to have notably

worse survival outcomes. These differences held true even after accounting for common factors like tumor size, grade, lymph node involvement, and whether women were pre or postmenopausal. The extent of these survival gaps was further supported by data from the national SEER program, which showed a consistent trend of higher death rates among African American women across various datasets and over different periods. This consistency emphasizes how widespread and reliable these findings are.

One of the key points of this paper is its focus on new statistical methods. Traditional models like the Cox proportional hazards model are helpful, but they often struggle to handle effects that change over time or differences within smaller subgroups. By using Bayesian approaches along with innovative machine learning techniques, we managed to address common problems like missing data, tiny subgroup sizes, and when the usual assumptions don't quite fit. Techniques like Schoenfeld residuals, deviance diagnostics, and random survival forests helped us better capture complex, non linear effects and find interactions that might be missed with traditional methods.

One of the important factors about this research is how it can be applied in real world settings. Besides the numbers and models, our findings matter for public health, cancer treatment, and health policies. We see a clear link between more aggressive cancer types and African American populations, which emphasizes how important it is to develop specialized screening and treatment plans for these high risk groups. Healthcare providers should focus on giving culturally aware care, making it easier for people to get diagnosed early, and using molecular subtype information to better predict risk. All of this can help us serve diverse communities more effectively.

This paper adds to the growing discussion around precision public health, a new approach that combines the detailed, molecular tools of precision medicine with the big picture insights of public health. Our findings show that models based on specific subtypes can not only help us understand existing disparities but also guide us in planning better future interventions. For example, dividing clinical trials by both disease subtype and factors like race or ethnicity can make treatment results more relevant and help ensure that different populations are fairly included in the development of new therapies. Similarly, using molecular markers in community health efforts could help identify people at higher risk earlier on, so we can take preventive action sooner.

Ethically speaking, this study emphasizes how important it is for statisticians, doctors, and researchers to recognize and tackle inequalities driven by widespread racism and neglect. The fact that African American women are more often affected by basal like breast cancer isn't just a random anomaly, it's a sign of larger failures within our institutions. Because of this, our research methods need to be just as committed to doing what's right promoting fairness, inclusion, and justice, when designing studies, sharing data, and translating findings into care. Fixing these disparities requires deliberate, ongoing effort that's woven into every step of moving from research to real world treatment.

However, it's important to acknowledge some limitations of this study. Even though our models account for a wide range of demographic and clinical factors, there might still be other unmeasured influences, like genetic background, other health conditions, environmental factors, or health behaviors, that could affect survival outcomes. Besides, while the CBCS provides detailed information, it's limited to a specific geographic area. Because of that, our results are promising but need to be tested further with data from other population based registries. These future studies should include broader racial and ethnic groups, such as Latinx, Indigenous peoples, and those from multi ethnic backgrounds, to better understand how disparities appear across different communities.

For future studies, we suggest including multiple types of biological data, like proteomics, metabolomics, and methylomics, to improve how we classify different molecular subtypes and better predict how patients might respond to treatments. It's also helpful to explore time-based models and advanced deep learning approaches, such as DeepSurv and recurrent neural survival models, to more accurately forecast patient outcomes as their conditions evolve. Beyond numbers and statistics, adding qualitative research can give us deeper insights into what patients experience, the obstacles they face in accessing care, and how social factors interact with biological risks in complex ways.

Overall, this paper adds meaningful insights to the discussion around breast cancer disparities. It combines solid statistical understanding with real world public health issues, all while respecting the experiences of underserved communities. Our analysis urges researchers to go beyond just describing patterns, we need to move toward predicting outcomes, promoting fairness, and respecting ethical principles. Only then can we start closing the gaps in breast cancer survival rates and work toward a future where outcomes aren't determined by race or socioeconomic status, but by the quality and reach of healthcare available to everyone.

APPENDIX.

Simulation and Prediction in R. The following R code simulates SEER-like data, fits a logistic regression model, and visualizes the predicted probability of Luminal B breast cancer subtype by BMI and race:

```
# Load libraries
library(ggplot2)
library(dplyr)

# Simulate SEER-like data
set.seed(123)
n <- 3000
data <- data.frame(
  luminal_subtype = sample(c("A", "B"), n, replace = TRUE, prob = c(0.55, 0.45)),
  BMI = rnorm(n, 29.5, 6.5),
  age = rnorm(n, 61, 12),
```



```

menopause = sample(c(0, 1), n, replace = TRUE, prob = c(0.3, 0.7)),
race = sample(c("White", "Black", "Hispanic", "Asian", "Other"), n, replace = TRUE,
              prob = c(0.65, 0.15, 0.10, 0.05, 0.05))
)

data$luminal_subtype_bin <- ifelse(data$luminal_subtype == "B", 1, 0)

# Fit logistic model
model <- glm(luminal_subtype_bin ~ BMI + race + menopause + age, data = data, family = "binomial")

# Generate new data for predictions
plot_data <- expand.grid(
  BMI = seq(18, 40, by = 0.5),
  race = c("White", "Black", "Hispanic", "Asian", "Other"),
  menopause = 1,
  age = 60
)

plot_data$prob <- predict(model, newdata = plot_data, type = "response")

# Plot predicted probability of Luminal B subtype by BMI and Race
ggplot(plot_data, aes(x = BMI, y = prob, color = race)) +
  geom_line(size = 1.2) +
  labs(title = "Predicted Probability of Luminal B Subtype by BMI and Race",
       x = "BMI",
       y = "Probability of Luminal B",
       color = "Race") +
  theme_minimal(base_size = 14)

```

R Code for BMI Distribution Plot by Luminal Subtype. The following R code creates a density plot to visualize the distribution of BMI for Luminal A vs Luminal B breast cancer subtypes:

```

ggplot(data, aes(x = BMI, fill = luminal_subtype)) +
  geom_density(alpha = 0.5) +
  labs(title = "Distribution of BMI by Luminal Subtype",
       x = "BMI",
       y = "Density",
       fill = "Subtype") +
  theme_minimal()

```

Availability of data and material: Data sets were obtained from the FLEX registry (NCT03053193), which is a prospective observational study that included women diagnosed with Stage I to III breast cancer between 2018 and 2020.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

- [1] S. Dasgupta, S. Acharya, M.A. Khan, P. Pramanik, S.M. Marbut, F. Yunus, J.N. Galeas, S. Singh, A.P. Singh, S. Dasgupta, Abstract 3318: Frequent Loss of Cacna1c, a Calcium Voltage-Gated Channel Subunit Is Associated with Lung Adenocarcinoma Progression and Poor Prognosis, *Cancer Res.* 83 (2023), 3318–3318. <https://doi.org/10.1158/1538-7445.am2023-3318>.
- [2] K.L. Hertweck, K.S. Vikramdeo, J.N. Galeas, S.M. Marbut, P. Pramanik, F. Yunus, S. Singh, A.P. Singh, S. Dasgupta, Clinicopathological Significance of Unraveling Mitochondrial Pathway Alterations in Non-small-cell Lung Cancer, *FASEB J.* 37 (2023), e23018. <https://doi.org/10.1096/fj.202201724rr>.
- [3] M.A. Khan, S. Acharya, N. Kreitz, S. Anand, F. Sameeta, P. Pramanik, C. Keel, S. Singh, J. Carter, S. Dasgupta, A. Singh, Mp60-05 Myb Exhibits Racially Disparate Expression and Clinicopathologic Association and Is a Promising Predictor of Biochemical Recurrence in Prostate Cancer, *J. Urol.* 211 (2024), e1000. <https://doi.org/10.1097/01.ju.0001008804.84010.ec.05>.
- [4] S. Kakkat, P. Pramanik, S. Singh, A.P. Singh, C. Sarkar, D. Chakroborty, Cardiovascular Complications in Patients with Prostate Cancer: Potential Molecular Connections, *Int. J. Mol. Sci.* 24 (2023), 6984. <https://doi.org/10.3390/ijms24086984>.
- [5] M.A. Khan, S. Acharya, S. Anand, F. Sameeta, P. Pramanik, C. Keel, S. Singh, J.E. Carter, S. Dasgupta, A.P. Singh, Myb Exhibits Racially Disparate Expression, Clinicopathologic Association, and Predictive Potential for Biochemical Recurrence in Prostate Cancer, *iScience* 26 (2023), 108487. <https://doi.org/10.1016/j.isci.2023.108487>.
- [6] K. Vikramdeo, S. Anand, S. Sudan, P. Pramanik, S. Singh, A. Godwin, A. Singh, S. Dasgupta, Abstract Po3-16-05: Mitochondrial Dna Mutation Detection in Tumors and Circulating Extracellular Vesicles of Triple Negative Breast Cancer Patients for Biomarker Development, *Cancer Res.* 84 (2024), PO3-16-05–PO3-16-05. <https://doi.org/10.1158/1538-7445.sabcs23-po3-16-05>.
- [7] K.S. Vikramdeo, S. Anand, S.K. Sudan, P. Pramanik, S. Singh, A.K. Godwin, A.P. Singh, S. Dasgupta, Profiling Mitochondrial <sc>dna</sc> Mutations in Tumors and Circulating Extracellular Vesicles of Triple-negative Breast Cancer Patients for Potential Biomarker Development, *FASEB BioAdvances* 5 (2023), 412–426. <https://doi.org/10.1096/fba.2023-00070>.
- [8] C.S. Jackson, M. Oman, A.M. Patel, K.J. Vega, Health disparities in colorectal cancer among racial and ethnic minorities in the United States, *J. Gastrointest. Oncol.* 11 (2020), 1031–1040. <https://doi.org/10.3978/j.issn.2078-6891.2015.039>.
- [9] P. Pramanik, On Lock-Down Control of a Pandemic Model, *arXiv:2206.04248* (2022). <https://doi.org/10.48550/arXiv.2206.04248>.
- [10] B. Lauby-Secretan, C. Scoccianti, D. Loomis, Y. Grosse, F. Bianchini, K. Straif, Body Fatness and Cancer – Viewpoint of the Iarc Working Group, *New Engl. J. Med.* 375 (2016), 794–798. <https://doi.org/10.1056/nejmsr1606602>.
- [11] P. Pramanik, A.M. Polansky, Optimal Estimation of Brownian Penalized Regression Coefficients, *arXiv:2107.02291* (2021). <https://doi.org/10.48550/arXiv.2107.02291>.
- [12] C.M. Perou, T. Sørlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, et al. Molecular Portraits of Human Breast Tumours, *Nature* 406 (2000), 747–752. <https://doi.org/10.1038/35021093>.
- [13] A. Prat, J.S. Parker, O. Karginova, C. Fan, C. Livasy, J.I. Herschkowitz, X. He, C.M. Perou, Phenotypic and Molecular Characterization of the Claudin-Low Intrinsic Subtype of Breast Cancer, *Breast Cancer Res.* 12 (2010), R68. <https://doi.org/10.1186/bcr2635>.
- [14] K.R. Menikdiwela, C. Kahathuduwa, M.L. Bolner, R.L. Rahman, N. Moustaid-Moussa, Association Between Obesity, Race or Ethnicity, and Luminal Subtypes of Breast Cancer, *Biomedicines* 10 (2022), 2931. <https://doi.org/10.3390/biomedicines10112931>.

- [15] Z.D. Bailey, N. Krieger, M. Agénor, J. Graves, N. Linos, M.T. Bassett, Structural Racism and Health Inequities in the Usa: Evidence and Interventions, *Lancet* 389 (2017), 1453–1463. [https://doi.org/10.1016/s0140-6736\(17\)30569-x](https://doi.org/10.1016/s0140-6736(17)30569-x).
- [16] P. Pramanik, A.M. Polansky, Scoring a Goal Optimally in a Soccer Game Under Liouville-Like Quantum Gravity Action, *Oper. Res. Forum* 4 (2023), 66. <https://doi.org/10.1007/s43069-023-00247-y>.
- [17] P. Pramanik, Optimization of Market Stochastic Dynamics, *SN Oper. Res. Forum* 1 (2020), 31. <https://doi.org/10.1007/s43069-020-00028-x>.
- [18] P. Pramanik, A.M. Polansky, Semicooperation Under Curved Strategy Spacetime, *J. Math. Sociol.* 48 (2023), 172–206. <https://doi.org/10.1080/0022250x.2023.2180002>.
- [19] P. Pramanik, Estimation of Optimal Lock-Down and Vaccination Rate of a Stochastic Sir Model: a Mathematical Approach, *Eur. J. Stat.* 4 (2024), 3. <https://doi.org/10.28924/ada/stat.4.3>.
- [20] P. Pramanik, A.M. Polansky, Motivation to Run in One-Day Cricket, *Mathematics* 12 (2024), 2739. <https://doi.org/10.3390/math12172739>.
- [21] P. Pramanik, A.M. Polansky, Motivation to Run in One-Day Cricket, arXiv:2001.11099 (2020). <https://doi.org/10.48550/arXiv.2001.11099>.
- [22] P. Pramanik, A.K. Maity, Bayes Factor of Zero Inflated Models Under Jeffereys Prior, arXiv:2401.03649 (2024). <https://doi.org/10.48550/arXiv.2401.03649>.
- [23] S.E. Bulls, E. Finn, P. Sykora, V.J. Lynch, P. Pramanik, S. Glaberman, Y. Chiari, Assessing Cometchip Technology for Dna Damage Studies in Non-Model Species: Distinct Uv-Induced Responses in Turtles and Mammals, *BMC Res. Notes* 18 (2025), 243. <https://doi.org/10.1186/s13104-025-07285-1>.
- [24] P. Pramanik, Path Integral Control in Infectious Disease Modeling, arXiv:2311.02113 (2023). <https://doi.org/10.48550/arXiv.2311.02113>.
- [25] P. Pramanik, On Estimation of Function-on-Function Regression Kernels with Brownian Berkson Errors, Lecture, University of South Alabama, (2024). https://egrove.olemiss.edu/math_statistics/7.
- [26] P. Pramanik, Optimal Lock-Down Intensity: a Stochastic Pandemic Control Approach of Path Integral, *Comput. Math. Biophys.* 11 (2023), 20230110. <https://doi.org/10.1515/cmb-2023-0110>.
- [27] P. Pramanik, Path Integral Control of a Stochastic Multi-Risk Sir Pandemic Model, *Theory Biosci.* 142 (2023), 107–142. <https://doi.org/10.1007/s12064-023-00388-y>.
- [28] P. Pramanik, Consensus as a Nash Equilibrium of a Stochastic Differential Game, *Eur. J. Stat.* 3 (2023), 10. <https://doi.org/10.28924/ada/stat.3.10>.
- [29] P. Pramanik, Effects of Water Currents on Fish Migration Through a Feynman-Type Path Integral Approach Under $\sqrt{8/3}$ Liouville-Like Quantum Gravity Surfaces, *Theory Biosci.* 140 (2021), 205–223. <https://doi.org/10.1007/s12064-021-00345-7>.
- [30] P. Pramanik, A.M. Polansky, Optimization of a Dynamic Profit Function Using Euclidean Path Integral, *SN Bus. Econ.* 4 (2023), 8. <https://doi.org/10.1007/s43546-023-00602-5>.
- [31] P. Pramanik, Stochastic Control of a Sir Model with Non-Linear Incidence Rate Through Euclidean Path Integral, arXiv:2209.13733 (2022). <https://doi.org/10.48550/arXiv.2209.13733>.
- [32] P. Pramanik, Stochastic Control in Determining a Soccer Player's Performance, *Contrib. Pure Appl. Math.* 2 (2024), 111. <https://doi.org/10.33790/cpam1100111>.
- [33] P. Pramanik, Stubbornness as Control in Professional Soccer Games: a Bppsde Approach, *Mathematics* 13 (2025), 475. <https://doi.org/10.3390/math13030475>.
- [34] P. Pramanik, J. Graff, M. Decaro, On Factors Influencing Consumer Preference in Pipeline Stages: an Experiment, arXiv:2501.03418 (2025). <https://doi.org/10.48550/arXiv.2501.03418>.

- [35] P. Pramanik, An Optimal Level of Stubbornness to Win a Soccer Match, arXiv:2501.18050 (2025). <https://doi.org/10.48550/arXiv.2501.18050>.
- [36] P. Pramanik, Tail Non-Exchangeability, Thesis, Northern Illinois University, 2016.
- [37] P. Pramanik, Optimization of Dynamic Objective Functions Using Path Integrals, Ph.D. Thesis, Northern Illinois University, (2021).
- [38] E. Maki, T. Glimm, P. Pramanik, Y. Chiari, M. Kiskowski, New Approaches for Capturing and Estimating Variation in Complex Animal Color Patterns From Digital Photographs: Application to the Eastern Box Turtle (Terrapene Carolina), PeerJ 13 (2025), e19690. <https://doi.org/10.7717/peerj.19690>.
- [39] L. Hua, A. Polansky, P. Pramanik, Assessing Bivariate Tail Non-Exchangeable Dependence, Stat. Probab. Lett. 155 (2019), 108556. <https://doi.org/10.1016/j.spl.2019.108556>.
- [40] A.M. Polansky, P. Pramanik, A Motif Building Process for Simulating Random Networks, Comput. Stat. Data Anal. 162 (2021), 107263. <https://doi.org/10.1016/j.csda.2021.107263>.
- [41] P. Pramanik, J. Graff, M. Decaro, Strategies to Increase Pipeline Status: a Case Study From Ecclinical Data, Eur. J. Stat. 5 (2025), 3. <https://doi.org/10.28924/ada/stat.5.3>.
- [42] P. Pramanik, L. Dong, Impact of Random Monetary Shock: a Keynesian Case, arXiv:2505.00800 (2025). <https://doi.org/10.48550/arXiv.2505.00800>.
- [43] P. Pramanik, L. Dong, Strategic Complementarities Due to Monetary Shock Under Sticky Price, Eur. J. Stat. 5 (2025), 9. <https://doi.org/10.28924/ada/stat.5.9>.
- [44] P. Pramanik, Measuring Asymmetric Tails Under Copula Distributions, Eur. J. Stat. 4 (2024), 7. <https://doi.org/10.28924/ada/stat.4.7>.
- [45] P. Pramanik, Dependence on Tail Copula, J 7 (2024), 127–152. <https://doi.org/10.3390/j7020008>.
- [46] P. Pramanik, E.L. Boone, R.A. Ghanam, Parametric Estimation in Fractional Stochastic Differential Equation, Stats 7 (2024), 745–760. <https://doi.org/10.3390/stats7030045>.
- [47] L.A. Carey, C.M. Perou, C.A. Livasy, L.G. Dressler, D. Cowan, et al. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study, JAMA 295 (2006), 2492. <https://doi.org/10.1001/jama.295.21.2492>.