Published in final edited form as:
Cancer Causes Control. 2016 March ; 27(3): 445–452. doi:10.1007/s10552-015-0711-4.

Latent class model characterization of neighborhood socioeconomic status

Aimee Palumbo[1], **Yvonne Michael**[1], and **Terry Hyslop**[2,3]

Aimee Palumbo: aimee.palumbo@glink.drexel.edu

[1]Department of Epidemiology and Biostatistics, Drexel University, Nesbitt Hall, 3215 Market St, 5th Floor, Philadelphia, PA 19104, USA

[2]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

[3]Duke Cancer Institute, Durham, NC, USA

Abstract

Purpose—Neighborhood-level socioeconomic status (NSES) can influence breast cancer mortality and poorer health outcomes are observed in deprived neighborhoods. Commonly used NSES indexes are difficult to interpret. Latent class models allow for alternative characterization of NSES for use in studies of cancer causes and control.

Methods—Breast cancer data was from a cohort of women diagnosed at an academic medical center in Philadelphia, PA. NSES variables were defined using Census data. Latent class modeling was used to characterize NSES.

Results—Complete data was available for 1,664 breast cancer patients diagnosed between 1994 and 2002. Two separate latent variables, each with 2-classes (LC2) best represented NSES. LC2 demonstrated strong associations with race and tumor stage and size.

Conclusions—Latent variable models identified specific characteristics associated with advantaged or disadvantaged neighborhoods, potentially improving our understanding of the impact of socioeconomic influence on breast cancer prognosis. Improved classification will enhance our ability to identify vulnerable populations and prioritize the targeting of cancer control efforts.

Keywords

Breast cancer; Methodology; modeling; biostatistics; Health disparities; Neighborhoods

Introduction

Breast cancer is the most common cancer in women and a significant source of mortality [1]. Although black women have lower overall incidence rates of breast cancer in comparison to white women, their mortality rates are substantially higher. Differences in

Correspondence to: Aimee Palumbo, aimee.palumbo@glink.drexel.edu.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

mortality rates are attributed to the greater likelihood that black women develop aggressive forms of cancer lacking key receptors often used as targets for treatment. These tumors are referred to as estrogen receptor (ER) negative and "triple-negative" (TN), or ER negative tumors that are also progesterone receptor and human epidermal growth factor receptor-2 negative [2-5].

While cancer phenotypes may be altered by the extra-cellular environment [6], the factors influencing tumor biology are not well understood. TN breast cancer is associated with younger age at diagnosis, higher grade and larger size at the time of diagnosis, and higher risk of recurrence [2, 7]. Basal-like tumors, which are often TN [5], are associated with factors such as reproductive history, breastfeeding, adiposity and weight gain, suggesting that tumor biology is influenced by external risk factors [8].

Neighborhoods provide exposure to social, environmental, and structural conditions that influence health [9, 10]. Neighborhood-level socioeconomic status has been shown to play a key role in health, with poorer health outcomes observed in deprived neighborhoods even after controlling for individual-level socioeconomic status [11, 12]. Neighborhood factors can influence everything from the presence of environmental toxins to the quality of public spaces and the availability of resources such as quality food choices [13, 14].

Neighborhoods may be an appropriate target for interventions [15]. Improvements in health behaviors can be made through neighborhood interventions [16]. For outcomes such as breast cancer mortality, neighborhood characteristics that have important influences on treatment, survival, and relevant risk factors are not clearly defined. In order to elucidate those risk factors, we must first adequately measure important neighborhood characteristics.

Neighborhood-level socioeconomic status (NSES) combines multiple indicators, including both economic and social neighborhood characteristics [10, 17]. Multiple indicators of NSES cannot be included simultaneously in models due to high correlations between variables. NSES indices combine the multidimensional concept into a single variable for use in quantitative analysis [18]. While such indices simplify the analysis, results are difficult to interpret because they are unit-less. As an alternative to a single, continuous measure of NSES we propose to use latent class models [19]. These models allow for characterization of NSES into meaningful classifications, easily described by relevant indicators of SES, and for estimation of the effects of those neighborhood characteristics on cancer outcomes. The purpose of this analysis was to use latent class models to identify neighborhood characteristics relevant for evaluating multi-level influences on breast cancer prognosis.

Methods

We used data from a cohort of African-American and Caucasian women of unknown ethnicity diagnosed with breast cancer at a teaching hospital in Philadelphia, PA between 1995 and 2002 with at least 5 years of follow-up. This population is comparable to populations in the National Cancer Institute SEER database with respect to tumor grade, stage, and expression of key receptors [20]. Information on age, race, survival, and tumor characteristics were available for each woman. Census 2000 information at the Zip Code

Tabulation Area (ZCTA) level was used to obtain NSES variables. Twenty-four census characteristics related to poverty (10.7 % below poverty line), income (median: $49,800), education (17.5 % less than high school education), housing (29.4 % in rental housing), occupation (3.9 % unemployed), race and family structure identified in previous studies as being related to health were considered for analysis [18, 21]. To determine factors relevant to NSES in our population, we conducted exploratory factor analysis (EFA) of available ZCTA-level NSES variables. Variables were transformed to be approximately normal and on the same scale. We examined scree plots for natural cut-points and Eigenvalues greater than 1 to determine factors for consideration in confirmatory factor analysis (CFA) [22]. In CFA, variables from rotated factor patterns that explained a substantial proportion of the variance were considered for inclusion in the latent class model.

Variables with loadings >0.50 on selected factors were used as indicator variables in latent class analysis (LCA) using MPlus v7 (MPlus, Los Angeles, CA) [23, 24]. Various numbers of classes were considered and compared using model fit statistics [25]. Latent class membership was compared with a traditional, continuous NSES index (NSI) created including the same NSES variables using methods adapted from Messer et al. [18]. The NSI was created by multiplying each of the normalized NSES variables by their respective variable weights from principal components analysis and summed to create an NSI. The NSI was categorized using quartiles to enable comparison with the categorical NSES classes created from LCA. We tested for associations between the two categorical measures of NSES and prognostic indicators of cancer aggressiveness defined by the AJCC Cancer Staging Manual including, tumor size, subtype, histologic grade, and overall stage using Chi-square tests [26, 27]. All data analysis except the LCA was completed in SAS version 9.2 (SAS Institute, Cary, NC). Circos plots are a graphical tool used to visually represent complex data that highlights similarities or differences between groups [28]. Circos plots were created to represent the relationship between neighborhood advantage and disadvantage with an indicator of cancer aggressiveness and the prevalence of those characteristics among black and white women.

Results

Complete data was available for 1,664 breast cancer patients from 320 ZCTAs in 3 states. Patients ranged in age from 22 to 92 years at diagnosis, with a median age of 58. Twenty-five percent of patients died during follow-up; survival ranged from <1 month to 13 years. The majority of patients were white (87 %) and most had early stage and low-grade tumors (Table 1).

Confirmatory factor analysis was conducted on the 320 ZCTAs. Two factors explained 77 % of the variance. The variables from Factor 1 were related to poverty, housing and family structure and represented the latent construct of disadvantage; the variables from Factor 2 were related to education, income and occupation and represented the latent construct of advantage. Each latent variable was best categorized into 2-classes (LC2).When the 2 class memberships for the 2 variables were combined into a 4-level categorical variable, the first category represented the combination of high-advantage and low-disadvantage, or lowest neighborhood deprivation, and the last category represented the combination of low-advantage and high-disadvantage, or highest deprivation. The remaining two categories represented neighborhoods with contrasting assets and limitations, for example high-advantage and high-disadvantage neighborhoods may be high in poverty despite high white-collar employment and low advantage and low disadvantage may include stable family structure but low education levels. These neighborhoods were rare but present in our study sample. The LC3 model had a lower AIC than LC2 (−76,938.2 vs −69,528.9) and performed better than LC2 based on the Lo–Mendell–Rubin likelihood ratio test (LRT) and the Boostrap LRT [25]. However, a 3-class solution resulted in a 9-level categorical variable of which 2 were not present in our study sample; no women lived in neighborhoods of least advantage and least disadvantage or of most advantage and most disadvantage, so the LC2 models were deemed a better fit (see “Appendix”).

Women were divided equally among high and low advantaged neighborhoods; high advantaged neighborhoods had more men and women in professional and managerial occupations, whereas low advantage neighborhoods had more people with less than a high school education (Fig. 1a). Most women (72 %) lived in low disadvantage neighborhoods, which are characterized by higher proportions of people single with dependents, below the poverty line, using public assistance, having no access to a vehicle, and who are non-Hispanic black (Fig. 1b).

A comparison of the latent class variables with the continuous NSES index (NSI) revealed high correlation (r = 0.87), but the NSI varied within each latent class, particularly in the neighborhoods classified as high-disadvantage and low-advantage (Fig. 2). Additionally, subjects classified similarly by continuous NSI were placed in different latent classes, demonstrating lack of concordance. Compared to NSI, LC2 demonstrated a stronger association with stage at diagnosis and tumor size than the NSI quartiles, a similar association with race, and a slightly weaker association with tumor subtype (Table 2). Circos plots of the association between LC2 with tumor stage, the variable most strongly related to neighborhood class, by race provide a visual demonstration of the differential impact of socioeconomic indicators in the minority population (Fig. 3). A large portion of black patients (A) live in low-advantage, high-disadvantage (LAHD) neighborhoods while very few white women (B) live in those neighborhoods.

Discussion

This analysis suggests that neighborhood characteristics are better represented by multiple latent class variables than by a single index. The neighborhoods identified by the latent class analysis as most deprived were represented by a wide range of values on the continuous NSES index, suggesting the index might be less useful in identifying women in the highest risk neighborhoods. Further, the strong associations of neighborhood classification with both race and prognostic outcomes support the hypothesis that neighborhood characteristics may be contributing to the health disparities seen between black and white women. Black women are far more likely to live in neighborhoods of high-disadvantage and low-advantage, and previous evidence has shown that exposure to such neighborhoods influence health and biology [9, 14, 29].

Our goal was to identify an alternative way to account for neighborhood characteristics in analyses of breast cancer outcomes, addressing limitations inherent in using a single, continuous index. US census data are the most commonly used data for estimating neighborhood SES [10, 30]. While a continuous neighborhood index such as that developed by Messer et al. [18] overcomes the limitations of using individual variables for estimating NSES, there are still drawbacks to that approach. One limitation is that a continuous index assumes a constant linear relationship between the indicator variables and the outcome variable of interest, breast cancer outcome. By contrast, our model acknowledges that different factors of neighborhood SES (advantage and disadvantage) may have different effects on cancer outcomes and would allow for such differences (and divergence between advantage and disadvantage) while still reducing the complexity of the model. Furthermore, although a continuous neighborhood SES index has been used in multiple studies, these studies acknowledge that use of such an index is only one way to measure neighborhood SES and there is no widely accepted standard [18, 31-33]. Finally, the ability to identify highly deprived neighborhoods across multiple factors, rather than relying on an arbitrary cutoff of a continuous index, may allow us to better target interventions in these neighborhoods. An additional advantage to our approach is that we can extend our latent class models to incorporate covariates that may influence classifications, including race, age, and other potential modifiers [24, 34].

We cannot yet make conclusions about the influence of these neighborhood characteristics on breast cancer outcomes based on the current analysis. Furthermore, ZCTA-level data may not be the ideal indicator of neighborhood when trying to understand women's exposures; however, census-level variables can still provide insight into neighborhood health effects and the development of useful methodologies [14]. Additionally, neighborhood characteristics change over time and using indicators from a single point in time may misrepresent some neighborhoods particularly in women diagnosed earlier in the study.

A recent review on the impact of neighborhood environment on the cancer continuum called for better approaches for understanding the multidimensional features of neighborhoods and their influence on cancer outcomes [35]. The multidimensionality of neighborhoods implies that the neighborhood elements that influence health are often clustered geographically and highly correlated with each other [9]. However, indicators of neighborhood SES will never be perfectly measured nor perfectly correlated; latent class analysis is a useful tool for dealing with these issues and should be considered as a measurement tool for neighborhood exposure in studies of cancer outcomes [36]. Future research will be to provide an overall structural latent model that uses a multilevel approach to incorporate information on tumor biology, prognostic variables and survival outcomes [37]. This will aid in the identification of populations in the most vulnerable neighborhoods and to help understand which neighborhood characteristics more strongly influence disparities in breast cancer outcomes.

Acknowledgments

This research was supported by Grants from Susan G. Komen for the Cure,© Investigator-Initiated Grant No. KG110710 (A.P., Y.M., T.H.) and Promise Grant No. KG091116 from Komen for the Cure (T.H).

Appendix

See Table 3.

Table 3

Model fit parameters for latent class analysis

Classes	4 (2 adv, 2 disadv)	6 (2 adv, 3 disadv)	6 (3 adv, 2 disadv)	9 (3 adv, 3 disadv)
Free parameters	57	70	66	80
Loglikelihood	34,821.483	36,757.465	36,707.574	38,549.102
AIC	−69,528.965	−73,374.93	−73,283.148	−76,938.204
BIC	−69,220.197	−72,995.742	−72,925.628	−76,504.846
Entropy	0.957	0.965	0.958	0.961
D		3,871.964	3,772.182	3,583.274
LRT		0	0	0
p value	Ref	< 0.05	<0.05	<0.05
		1 combination class empty	2 combination classes empty	2 combination classes empty

References

1. Jemal A, Simard EP, Dorell C, Noone A-M, Markowitz LE, et al. Annual Report to the Nation on the Status of Cancer, 1975–2009, featuring the burden and trends in human papillomavirus (HPV)-associated cancers and HPV vaccination coverage levels. J Natl Cancer Inst. 2013; 105:175–201. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3565628&tool=pmcentrez&rendertype=abstract. [PubMed: 23297039]
2. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, proges-terone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer. 2007; 109:1721–1728. http://www.ncbi.nlm.nih.gov/pubmed/17387718. [PubMed: 17387718]
3. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006; 295:2492–2502. http://www.ncbinlm.nih.gov/pubmed/16757721. [PubMed: 16757721]
4. Hudis CA, Gianni L. Triple-negative breast cancer: an unmet medical need. Oncologist. 2011; 16:1–11. http://www.ncbi.nlmnih.gov/pubmed/21278435.
5. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. J Clin Oncol Off J Am Soc Clin Oncol. 2008; 26:2568–2581. http://www.ncbi.nlm.nih.gov/pubmed/18487574.
6. Krieger N. History, biology, and health inequities: emergent embodied phenotypes and the illustrative case of the breast cancer estrogen receptor. Am J Public Health. 2013; 103:22–27. www.ncbi.nlm.nih.gov/pubmed/23153126. [PubMed: 23153126]
7. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res. 2007; 13:4429–4434. http://www.ncbi.nlm.nih.gov/pubmed/17671126. [PubMed: 17671126]
8. Millikan RC, Newman B, Tse C-K, Moorman PG, Conway K, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008; 109:123–139. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2443103&tool=pmcentrez&renderty pe=abstract.
9. Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health. 2001; 91:1783–1789. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1446876&tool=pmcentrez&rendertype=abstract. [PubMed: 11684601]

10. Yen IH, Syme SL. The social environment and health: a discussion of the epidemiologic literature. Annu Rev Public Health. 1999; 20:287–308. http://www.ncbi.nlm.nih.gov/pubmed/10352860. [PubMed: 10352860]
11. Stafford M, Marmot M. Neighbourhood deprivation and health: does it affect us all equally? Int J Epidemiol. 2003; 32:357–366. http://www.ije.oupjournals.org/cgi. 10.1093/ije/dyg084 [PubMed: 12777420]
12. LeClere FB, Rogers RG, Peters KD. Ethnicity and mortality in the United states: individual and community correlates. Soc Forces. 1997; 76:169–198.
13. O'Neill MS, Jerrett M, Kawachi I, Levy JI, Cohen AJ, et al. Health, wealth, and air pollution: advancing theory and methods. Environ Health Perspect. 2003; 111:1861–1870. http://www.ehponline.org/ambra-doi-resolver. 10.1289/ehp.6334 [PubMed: 14644658]
14. Diez Roux AV, Mair C. Neighborhoods and health. Ann N Y Acad Sci. 2010; 1186:125–145. http://www.ncbi.nlm.nih.gov/pubmed/20201871. [PubMed: 20201871]
15. Mohnen SM, Vo¨lker B, Flap H, Groenewegen PP. Health-related behavior as a mechanism behind the relationship between neighborhood social capital and individual health—a multilevel analysis. BMC Public Health. 2012; 12:116. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3347984&tool=pmcentrez&rendertype&abstract. [PubMed: 22325740]
16. Emmons KM, Stoddard AM, Fletcher R, Gutheil C, Suarez EG, et al. Cancer prevention among working class, multiethnic adults: results of the healthy directions–health centers study. Am J Public Health. 2005; 95:1200–1205. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1449340&tool=pmcentrez&rendrt ype=abstract. [PubMed: 15933240]
17. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health. 1997; 18:341–378. http://www.ncbi.nlm.nih.gov/pubmed/9143723. [PubMed: 9143723]
18. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, et al. The development of a standardized neighborhood deprivation index. J Urban Health Bull N Y Acad Med. 2006; 83:1041–1062. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3261293&tool=pmcentrez&rendertype=abstract.
19. Sánchez BN, Budtz-Jørgensen E, Ryan LM, Hu H. Structural equation models. J Am Stat Assoc. 2005; 100:1443–1455.10.1198/016214505000001005
20. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the national cancer institute's surveillance, epidemiology, and end results database. Cancer. 2007; 110:876–884.10.1002/cncr.22836 [PubMed: 17620276]
21. Eibner C, Sturm R. US-based indices of area-level deprivation: results from HealthCare for Communities. Soc Sci Med. 2006; 62:348–359.10.1016/j.socscimed.2005.06.017 [PubMed: 16039764]
22. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. Organ Res Methods. 2004; 7:191–205. http://orm.sagepub.com/cgi. 10.1177/1094428104263675
23. Muthén, LK.; Muthén, BO. MPlus user's guide. 7. Muthén & Muthén; Los Angeles: 2012.
24. Muthén, BO. Latent variable mixture modeling. In: Mar-coulides, GA.; Schumacker, RE., editors. new developments and techniques in structural equation modeling. Lawrence Erlbaum Associates, Inc; Mahwah: 2001. p. 1-34.
25. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo Simulation study. Struct Equ Model. 2007; 14:535–569.
26. Greene, FL.; Page, DL.; Fleming, ID.; Fritz, AG.; Balch, CM.; Haller, DG.; Morrow, M. AJCC cancer staging manual. 6. Springer; New York: 2002.
27. Singletary SE, Connolly JL. Breast cancer staging: working with the sixth edition of the AJCC cancer staging manual. CA Cancer J Clin. 2006; 56:37–47.10.3322/canjclin.56.1.37 [PubMed: 16449185]

28. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. Circos: an information aesthetic for comparative geno-mics. Genome Res. 2009; 19:1639–1645.10.1101/gr.092759.109.19 [PubMed: 19541911]
29. Andaya AA, Enewold L, Horner M-J, Jatoi I, Shriver CD, et al. Socioeconomic disparities and breast cancer hormone receptor status. Cancer Causes Control. 2012; 23:951–958. http://www.ncbi.nlm.nih.gov/pubmed/22527173. [PubMed: 22527173]
30. Mair C, Diez-Roux AV, Galea S. Are neighbourhood characteristics associated with depressive symptoms? A review of evidence. J Epidemiol Community Health. 2008; 62:940–946. 8 p following 946. http://www.ncbi.nlm.nih.gov/pubmed/18775943. [PubMed: 18775943]
31. Robert SA, Strombom I, Trentham-Dietz A, Hampton JM, McElroy JA, et al. Socioeconomic risk factors for breast cancer. Epidemiology. 2004; 15:442–450. http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001648-200407000-00011. [PubMed: 15232405]
32. Webster TF, Hoffman K, Weinberg J, Vieira V, Aschengrau A. Community- and individual-level socioeconomic status and breast cancer risk: multilevel modeling on Cape Cod, Massachusetts. Environ Health Perspect. 2008; 116:1125–1129.10.1289/ehp.l0818 [PubMed: 18709175]
33. Palmer JR, Boggs DA, Wise LA, Adams-Campbell LL, Rosenberg L. Individual and neighborhood socioeconomic status in relation to breast cancer incidence in African-American women. Am J Epidemiol. 2012; 176:1141–1146. http://www.ncbi.nlmnih.gov/pubmed/23171873. [PubMed: 23171873]
34. McCutcheon, AL. Latent class analysis. Sage; Newbury Park: 1987.
35. Gomez SL, Shariff-Marco S, DeRouen M, Keegan THM, Yen IH, et al. The impact of neighborhood social and built environment factors across the cancer continuum: current research, methodological considerations, and future directions. Cancer. 201510.1002/cncr.29345
36. Muthén BO. latent variable modeling in epidemiology. Alcohol Health Res World. 1992; 16:286–292.
37. Lynch SM, Rebbeck TR. Bridging the gap between biologic, individual, and macroenvironmental factors in cancer: a multilevel approach. Cancer Epidemiol Biomark Prev. 2013; 22:485–495. http://cebp.aacrjournals.org/cgi. 10.1158/1055-9965.EPI-13-0010

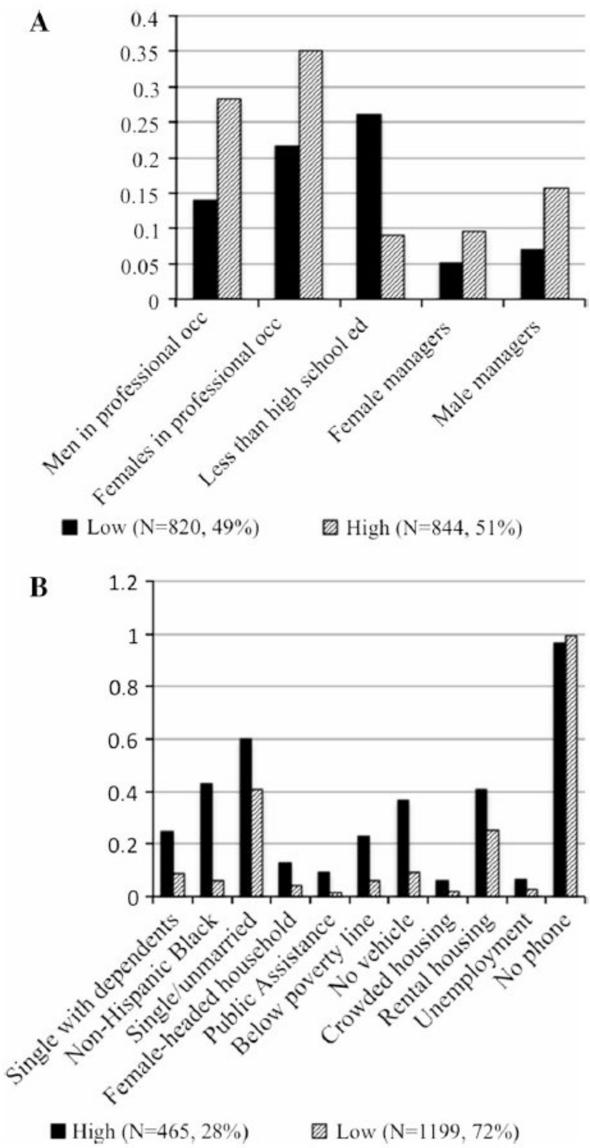

Fig. 1.

Average population characteristics by latent class neighborhood advantage (a) and disadvantage (b) (model AIC: −69,259.0, entropy: 0.96)

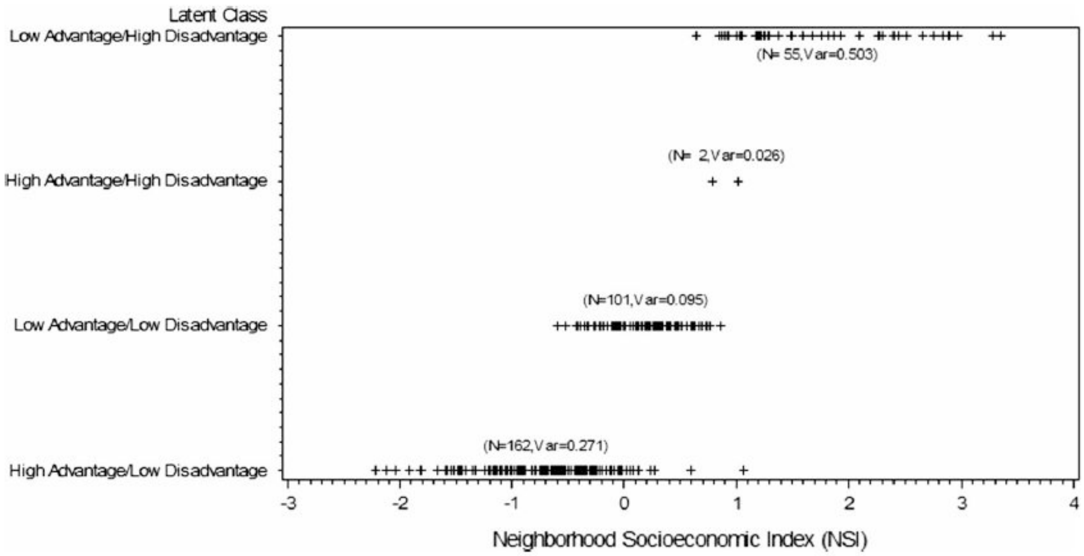

Fig. 2.

Combined latent class categories and continuous neighborhood socioeconomic index (NSI)

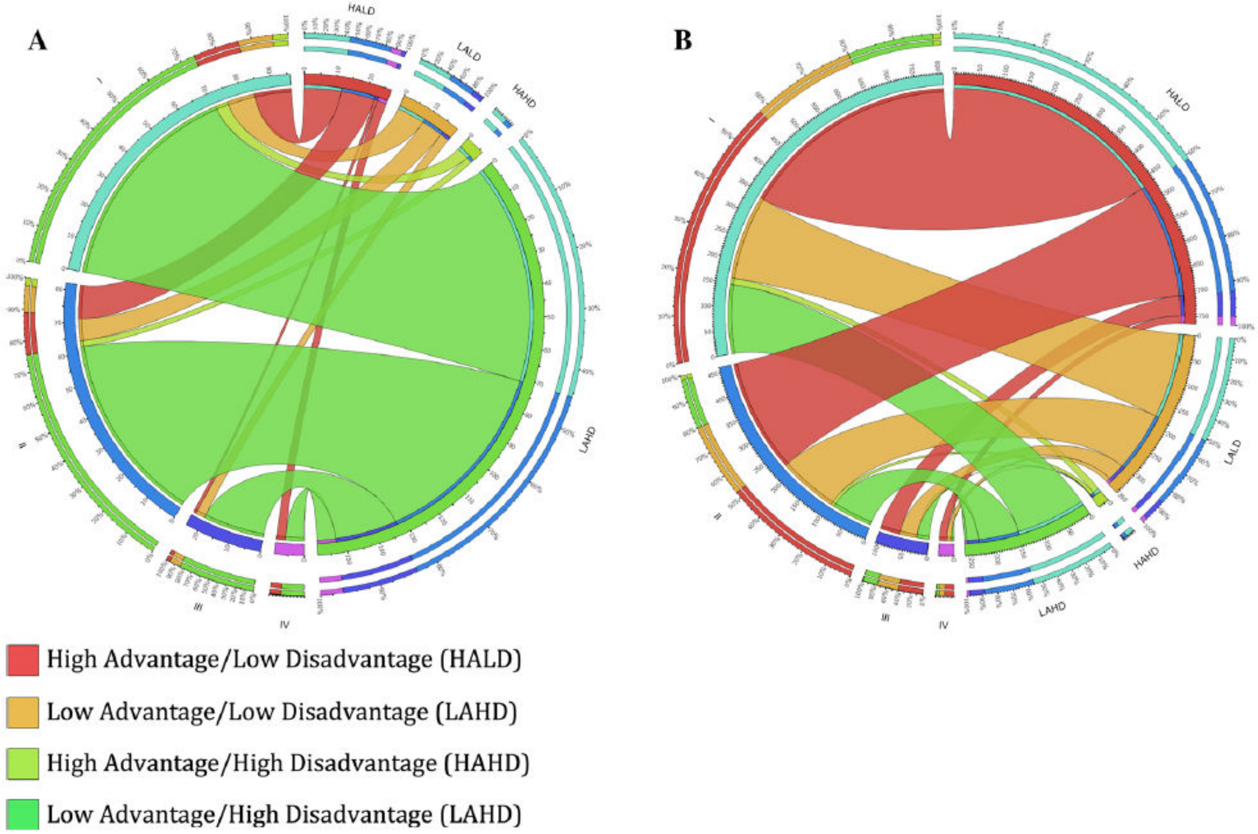

Fig. 3.

Associations of neighborhood advantage and disadvantage with tumor stage (I–IV) in black patients (a) and white patients (b)

Table 1

Patient and tumor characteristics of breast cancer patients (n = 1,664)

Characteristic	***n* (%)**
Race	
White	1,445 (87)
Black	219 (13)
Tumor subtype	
Basal	255 (15)
HER2	99 (6)
Luminal A	1,125 (68)
Luminal B	185 (11)
Overall stage	
I	904 (56)
II	543 (34)
III	129 (8)
IV	40 (2)
Tumor size category	
T1 (≤2 cm)	493 (32)
T2 (>2 to ≤5 cm)	956 (61)
T3 (>5 cm)	110 (7)
Histologic grade	
G1 (low—favorable)	167 (12)
G2 (intermediate—moderately favorable)	635 (44)
G3 (high—unfavorable)	647 (45)

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Categorical associations of neighborhood class and prognostic indicators in women with breast cancer based on different classifications of neighborhood socioeconomic status (n = 1,664)

Characteristic	Latent class membership *n* (row %)				Latent class association		NSI quartile association	
	HALD	LALD	HAHD	LAHD	χ^2 value	*p* value	χ^2 value	*p* value
Race					334.4	<0.001	372.7	< 0.001
White	788 (54.5)	365 (25.3)	23 (1.6)	269 (18.6)				
Black	27 (12.3)	19 (8.7)	6 (2.7)	167 (76.3)				
Subtype					15.6	0.076	19.6	0.021
Basal	112 (43.9)	53 (20.8)	6 (2.4)	84 (32.9)				
HER2	58 (58.6)	24 (24.2)	1 (1.0)	16 (16.2)				
Luminal A	548 (48.7)	261 (23.2)	20 (1.8)	296 (26.3)				
Luminal B	97 (52.4)	46 (24.9)	2 (1.1)	40 (21.6)				
Overall stage					22.7	0.007	16.2	0.063
I	479 (53.0)	189 (20.9)	18 (2.0)	218 (24.1)				
II	252 (46.4)	134 (24.7)	9 (1.7)	148 (27.3)				
III	43 (33.3)	40 (31.0)	1 (0.8)	45 (34.9)				
IV	19 (47.5)	9 (22.5)	1 (2.5)	11 (27.5)				
Size category					19.4	0.004	9.3	0.160
T1	273 (55.4)	99 (20.1)	11 (2.2)	110 (22.3)				
T2	454 (47.5)	220 (23.0)	15 (1.6)	267 (27.9)				
T3	40 (36.4)	34 (30.9)	1 (0.9)	35 (31.8)				
Histologic grade					0.43	0.999	1.3	0.971
G1	82 (49.1)	39 (23.4)	2 (1.2)	44 (26.4)				
G2	312 (49.1)	146 (23.0)	12 (1.9)	165 (26.0)				
G3	316 (48.8)	148 (22.9)	11 (1.7)	172 (26.6)				

HALD high-advantage low-disadvantage, *LALD* low-advantage low-disadvantage, *HAHD* high-advantage high-disadvantage, *LAHD* low-advantage high-disadvantage