Published in final edited form as:

Health Place. 2009 September ; 15(3): 855–864. doi:10.1016/j.healthplace.2009.02.007.

Economic, Racial and Ethnic Disparities in Breast Cancer in the U. S.: Toward a More Comprehensive Model

Richard T. Campbell^{a,d,*}, Xue Li^{a,d}, Therese A. Dolecek^{a,d}, Richard E. Barrett^{b,d}, Kathryn E. Weaver^c, and Richard B. Warnecke^d

^a Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL. USA

^b Department of Sociology, University of Illinois at Chicago, Chicago, IL. USA

^c Cancer Prevention Fellowship Program, Office of Preventive Oncology, National Cancer Institute, National Institutes of Health

^d Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL. USA

Abstract

Using cancer registry data, we focus on racial and ethnic disparities in stage of breast cancer diagnosis in Cook County IL. The county health system is the “last resort” health care provider for low income persons. Socioeconomic status is measured using empirical Bayes estimates of tract-level poverty, specific to non-Hispanic whites, non-Hispanic Blacks or Hispanics in one of three age groups. We use ordinal logistic regression with non-proportional odds to model stage. Blacks and Hispanics are at great risk for regional and distant stage diagnosis, but the disparity declines with age. Women in high poverty areas are at substantially greater risk for late stage diagnosis. The effects of poverty do not differ by age or across racial and ethnic groups.

Keywords

Breast cancer; Racial and Ethnic Disparities; Stage; Socioeconomic Indicators

INTRODUCTION

While some recent evidence suggests that incidence rates are declining in the United States, breast cancer remains a serious threat to the health of adult women, affecting perhaps one in eight women over the course of a lifetime (Ries et al., 2007). The risk of breast cancer, the stage at which the disease is diagnosed, and subsequent outcomes are all subject to disparities with regard to socioeconomic status (SES) and race/ethnicity. In this paper, we attempt to ascertain the effects of race/ethnicity (white non-Hispanic, black non-Hispanic, and Hispanic) and SES on stage at diagnosis. Our work focuses on a particular geographic entity – Cook County, Illinois. There, access to care for the medically indigent population is provided at the county level, enabling us to examine disparities both with respect to a well defined population at risk and in light of particular health care policies and service delivery efforts. We use ordinal

*Corresponding author. Tel.: + 312-413-0480, E-mail address: dcamp@uic.edu (R. T. Campbell).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

regression methods to model all four diagnostic stages – *in situ*, localized, regional, distant – simultaneously, a technique which, to the best of our knowledge, has not been used in previous research, and we propose a new method of estimating individual level SES information from aggregate census data.

Within recent years, a number of papers have appeared that investigate race/ethnic disparities in stage at breast cancer diagnosis. Many of these studies have concentrated on non-Hispanic black/white differences, excluding women of Hispanic and other ethnic minority backgrounds. While it is clearly the case that white non-Hispanic women are at greater risk for breast cancer (Ries et al. 2007), a number of studies have documented that black non-Hispanic women and women of low SES are more likely to be diagnosed with an advanced stage of the disease than are non-Hispanic white or more economically privileged women (e.g., Bradley et al., 2002; Gumpertz et al., 2006; Mandelblatt et al., 1991; Merkin et al., 2002; Miller et al., 2002; Schwartz et al., 2003; Wells and Horm, 1992). The few studies that have included Hispanic and Asian/Pacific Islander women have suggested that these women may also be at increased risk of more advanced breast cancer at diagnosis. (Li et al., 2003; Mandelblatt et al., 1995; Richardson et al., 1992).

Hirshman et al. (2007) studied trends in breast cancer incidence, stage at diagnosis and mortality in Chicago between 1986 and 2002. Comparing blacks and whites, they found that although mortality rates for whites have been steadily decreasing, rates for blacks have remained constant, leading to an increased racial disparity over time. They also found that black women were at a disadvantage with respect to early detection, defined as either *in situ* or localized stage. Although the proportion of cases found at the early stages increased for both white and black women, black women were more likely to be diagnosed at late stages (regional and distant) than whites at any given time point. Unfortunately, this paper did not provide data on Chicago's large and rapidly growing Hispanic population.

To the best of our knowledge, there is no research in which SES is the sole factor of interest, ignoring race. Attempts to study the effects of race/ethnicity and SES simultaneously have resulted in somewhat inconsistent outcomes. Some researchers found that both SES and race/ethnicity indicators remained statistically significant (Barry & Breen, 2005; Richardson et al., 1992). Others found that the effect of race/ethnicity was no longer significant with SES in the model for at least some outcomes (Breen and Figueroa, 1996; Mandelblatt et al., 1991). Mandelblatt et al. (1995) concluded that African-American race but not Hispanic ethnicity remained as a statistically significant predictor of late stage breast cancer diagnosis when area SES was included in the model. Bradley et al. (2002) concluded that neither African-American race nor area SES remained as statistically significant predictors when Medicaid insurance status was included in the model. With regard to survival, Du et al. (2008), using data from eleven SEER areas, found no black-white disparity in overall survival after controlling for tract-level SES measures and treatment. However, their results were restricted to early stage diagnoses involving women ≥ 65 . Newman et al., (2008) in a meta analysis of survival data from more than 20 studies found that black women suffered a higher rate of poor outcomes after controlling for SES.

Virtually all of these studies used indicators of SES derived from census data at the county, zipcode, census tract or block group level because individual-level SES indicators were not available in the cancer incidence databases under study. SES was assessed using area-based indicators such as income, education, occupation, percentage in poverty and unemployment. Most researchers utilized combinations of indicators to derive either dichotomous (e.g., “underclass” or “disadvantaged”; see Barry & Breen, 2005) or ordinal measures (e.g., see Schwartz et al., 2003) of SES. One study (Grann et al., 2006) used county level race-specific

SES data, e.g. proportion of white and non-Hispanic black women with less than a high school education.

Relatively few researchers have attempted to examine interactions between race/ethnicity and SES. The interaction issue is important. The conventional way to conceptualize a disparity is to show that members of a particular minority group are at a disadvantage relative to the minority, holding SES constant. However, if the *effect* of SES is more serious for members of minority groups than for the white non-Hispanic majority, the nature of the disparity is more severe. Merkin and colleagues (2002) concluded that while SES, defined as residential area education level, seemed to impact white non-Hispanic women more than their black counterparts, there was no interaction between race/ethnicity and SES when race-specific educational quintiles were used. Schwartz and colleagues (2003) also concluded that there was no significant interaction between SES and race. In contrast, Gumpertz et al. (2006) found that most census tract socioeconomic indicators were significantly related to advanced disease for African-American, but not Asian, Hispanic, or white non-Hispanic women diagnosed with breast cancer.

In general then, research in this area has suffered from several problems including:

- *Inconsistency in the definition of outcomes.* Some studies have dichotomized distant stage versus all other stages, while others have focused on regional and distant versus *in situ* and localized stage. Some studies have dropped *in situ* cases while others have retained them.
- *Arbitrary definitions of geographical areas.* Previous studies have focused on a wide range of geographic entities – New York City boroughs, specific counties, entire states, etc. While these studies are obviously informative, from a policy perspective it is desirable to focus on an area where the effects of available health care alternatives and existing policies can be ascertained with respect to a well defined population.
- *Reliance on area-based socioeconomic measures (ABSM).* Because cancer registries and other sources of case information typically do not contain person-level information on SES, most researchers have used ABSM's, typically derived from census data. As noted, the SES measures have varied but so have the geographic areas on which the measures are based. Some studies use census tracts, others use zip codes and still others use counties.
- *Substantial variation in the measurement and definition of SES.* Some studies have focused on a single measure while others have included several SES measures in the same analysis. The particular SES measure(s) used have also varied across studies, making comparisons difficult.
- *Failure to recognize within-area heterogeneity.* Almost all studies that use ABSM have failed to analyze the data in such a way as to account for the potential heterogeneity of SES measures for different age, race/ethnicity and gender within a given geographic area such as a census tract. For example, the aggregate poverty rate in a given census tract may conceal substantial variation in poverty across age or racial/ethnic groups.

Our study deals with several of these issues. First, rather than arbitrarily dichotomize the outcome, we use ordinal logistic regression models to examine stage at diagnosis as a four-stage ordinal outcome; (1) *in situ*, (2) localized, (3) regional and (4) distant. This approach allows us to make use of the maximum amount of information in the data. Second, we focus on a specific geographical area – Cook County Illinois. – which serves as the health care provider of last resort of the medically indigent and uninsured population. Thus we can interpret observed disparities in light of specific services and policies in place to alleviate them. Third,

although we too are forced to use ABSM's, in line with recent recommendations (Krieger et al., 2003, 2005), we focus on a single measure – tract-level poverty – but we deal with within tract heterogeneity by using poverty estimates specific to groups based on age, race, ethnicity and sex. Finally, with this measure in hand, we are able to carry out a partial test of the interaction between race/ethnicity and SES in determining stage at breast cancer diagnosis.

METHODS

Study Population and Available Cases

Our evaluation begins with data for 30,130 breast cancer cases reported to the Illinois State Cancer Registry (ISCR) among Cook County, Illinois women diagnosed from January 1, 1994 through December 31, 2000. As noted above, we focus on Cook County as a whole rather than either the City of Chicago (which is entirely within and comprises a large fraction of Cook County) or the entire state of Illinois because the Cook County health care system is the provider of last resort for both Chicago and suburban residents. An analysis focusing only on the City of Chicago ignores the substantial suburban population, about 45% of the total, which has access to Cook County health services for the poor and which now includes more younger, higher-income and better-educated minority residents.

Of the original sample, approximately 5% were excluded due to missing stage data. An additional 3.4% were excluded because our analysis is restricted to white non-Hispanic, black non-Hispanic cases and cases reporting Hispanic origin regardless of race. All other racial categories were excluded due to small numbers. For about two percent of cases, tract-level poverty estimates were unavailable and those cases were excluded due to suppression by the U.S. Bureau of the Census when a race/ethnic group has less than 100 observations. Finally, in order to obtain reasonable estimates of age-specific probabilities, we excluded cases younger than 30 or older than 89ⁱ. After these and other minor exclusions, we were left 25,900 cases on which to base the analysis.

Variables

Each ISCR record contains standard information on age at diagnosis, race, Hispanic ethnicity and general summary stage. In addition, ISCR case residential addresses are routinely geocoded.

Stage Classification—As noted above, we use the SEER general summary stage categories that include *in situ*, localized, regional and distant stages (Havener and Thorton, 2008). Approximately six percent of the ISCR breast cancer cases did not have a stage classification. Preliminary analysis indicated that these cases did not differ significantly from cases for which stage was available and they were eliminated from further analysis.

Geocodes—In the registry data, each case's residential address at the time of diagnosis had been geocoded to the block level. For the study period of 1994 – 2000, ISCR was able to obtain complete and valid address information for more than 98 percent of the reported Cook County breast cancer cases. Based on the geocoded data, each case was located in a Cook County census tract. A small proportion of cases that could not be geocoded were excluded from the analysis.ⁱⁱ

ⁱResults to be reported below are essentially the same when these cases are included.

ⁱⁱGeocoding of these addresses was done either by a commercial firm or, in more recent years, at the registry using Map Info and Map Marker software programs.

Race/ethnicity—Race/ethnic classification of breast cancer cases was based on two data elements on the ISCR database; race and Hispanic ethnicity defined according to the North American Association of Cancer Registries Hispanic identification algorithm (NHIA) (NAACCR Latino Research Work Group 2005). Only white non-Hispanic, black non-Hispanic and Hispanic breast cancer cases defined from these two data elements were included in the analyses.

Socioeconomic status—In an influential series of papers, Krieger and her colleagues (Krieger et al., 2002, 2003, 2005) have argued that studies which use ABSM should use just one SES measure – the proportion of persons in the census tract below the federally defined poverty level -- and should focus on census tracts as opposed to block groups or other geographic aggregations. We agree with the first point but not the second. We agree in that aggregate-level SES measures tend to be very highly correlated and multiple-indicator models are sometimes difficult to interpret. When more than one measure is used, it is very difficult to interpret the independent effects of each, and analyses tend to suffer from the effects of collinearity. Linear composites, based on education, income, proportion of owner-occupied housing and the like, tend to obscure the meaning of the effect. While using the proportion of households below a somewhat arbitrarily defined poverty line has its disadvantages, principally in terms of ignoring information on income at the upper end of the distribution, the advantages of a simple single measure outweigh the disadvantages.

With regard to the appropriate level of ABSM, tract level measures are most interpretable when there is little within-tract heterogeneity, (i.e. when the tract-level measure represents all persons in the tract equally well). In that case, in effect, the tract-level measure is being used as a missing data imputation. In Cook County, however, there is substantial variation in poverty levels within many census tracts. For example, Figure 1 displays a scatter plot of the proportion of households below the poverty level for white non-Hispanics and black non-Hispanics in the 515 Cook County tracts where there are at least 100 cases of each race/ethnic group. City and suburban tracts are shown separately.ⁱⁱⁱ

In tracts where white non-Hispanic poverty is zero or near zero there are black non-Hispanic persons with poverty rates well above ten and even twenty percent, with more variation in city than in suburban tracts. In general, same-tract poverty rates for white non-Hispanics are lower than for black non-Hispanics. For the tracts shown in Figure 1, white non-Hispanic poverty rates are lower than black non-Hispanics in 70 percent of the tracts. But the reverse is true in the other 30 percent of tracts. On average, the proportion of black non-Hispanics below the poverty line in a given tract is nine percentage points higher than white non-Hispanics. All of this suggests that analyses that treat all persons in a tract as having the same poverty rate will produce biased results, and is unnecessary since data for these subpopulations are readily available. Poverty also tends to vary by age. Table 1 shows poverty rates by age among race/ethnic groups in Cook County.

There are, of course, substantial overall differences in poverty by race/ethnicity, but within each group, younger and older persons have higher poverty rates than those aged 45–64. The trend is more pronounced among black non-Hispanics and least dramatic for white non-Hispanics. Ignoring such variation by assuming that a given census tract is homogenous with regard to poverty within race/ethnic or age groups will also tend to give a biased picture of the effects of poverty. Hence, the poverty measure to be used here will be at the tract level; however

ⁱⁱⁱResearcher who use aggregate poverty rates for entire tracts as a measure of SES appear to be relying on Summary File (SF) 3 for 2000 (or STF 3 for 1990), both of which are available for all U.S. census tracts from the Census Bureau website. However, Summary File 4 for both censuses (SF 4 for 2000, STF 4 for 1990), also available on this website, breaks out the number of persons in and out of poverty for every census tract by broad age groups and sex for major race/ethnic categories.

it will be specific to the female poverty rate for the age and race/ethnic group of the breast cancer case under study.

Empirical Bayes Estimates of Poverty Status by Age, Race/Ethnicity and Sex

In a given census tract, there are varying numbers of persons in a particular age-race/ethnic group, ranging from as few as one or two persons to many hundreds. Of course some groups will not be represented. Thus, tract-specific poverty estimates for race/ethnic groups are based on widely varying sample sizes and the reliability of poverty estimates varies accordingly. We cannot take poverty estimates for each of the observed groups in a given tract and use them as though they were equally reliable regardless of sample size. While one could potentially weight the cell estimates inversely to sample size, a more effective approach is based on empirical Bayes (EB) estimation (Raudenbush and Bryk, 2002). Using this approach, we obtained poverty estimates within each tract for women cross classified by three age categories (25–44, 45–64, 65+) and race/ethnicity (white non-Hispanic, black non-Hispanic and Hispanic). The correlation between the observed estimates and the empirical Bayes estimates was .99, suggesting that raw estimates might be used directly. A more complete analysis of this issue will be the topic of a subsequent paper. A detailed explanation of this method is contained in the Appendix.

Statistical Methods

Having obtained non-parametric empirical Bayes estimates of the proportion of cases below the poverty line in all clusters, we then estimated a partial-proportional odds multilevel ordinal regression model in order to estimate the odds (and subsequently the probability) that a given breast cancer case was at or below a particular stage. Such models are sometimes called “cumulative logit models” meaning the model describes the odds as a series of cumulative steps ranging in this instance from (1) *in situ* to (2) localized or below to (3) regional or below through (4) distant or below. The conventional ordered logistic regression model assumes “proportional odds” meaning that the model takes the form:

$$\ln \left[\frac{p(y_{ij} \leq c)}{1 - (p(y_{ij} \leq c))} \right] = \gamma_c - [\beta_0 + x_{ijk}\beta_k + v_{0i}] \quad (i=1, \dots, 1260; c=1, \dots, 3; k=1, \dots, K),$$

where c refers to the set of categories defined above, γ_c is the threshold for the c^{th} logit, β_0 is the intercept, x_{ijk} is a vector of k observed variables for the j^{th} breast cancer case in the i^{th} tract, β_k is the k^{th} regression coefficient and v_{0i} is a normally distributed random effect. For identification, γ_1 is fixed to 0. This model would assume, for example, that the effect of age on stage of diagnosis is the same regardless of the cumulative cut point. Such a model would produce a set of parallel straight lines for plotted log odds at each cumulative stage and a set of proportional odds such that across levels of a given covariate the odds of being at or above a particular stage are in constant proportion. A partial-proportional odds model, on the other hand, allows the effect of a given covariate to differ across the various thresholds. We have estimated non-proportional odds random effect models using MIXOR, a program that provides tests of the proportional odds assumption for each covariate and allows estimation of non-proportional effects where appropriate (Hedeker and Gibbons, 1996). Prior to estimation, age and poverty were centered at their mean and rescaled such that age = (age–61.8)/10 and poverty = (NEBP estimate–9.4)/10.

Considering non-linearities, interactions, non-proportional odds and their various combinations leads to a huge number of models which might be considered. Our goal was to find the most parsimonious model that provided a reasonable fit to the data while at the same time minimizing the number of alternative models fit. The sequence of models is necessarily

arbitrary to a degree but we tried to begin with simple models and gradually elaborate them as necessary. The following provides a brief summary. More detailed information is available from the authors.

Our general procedure was to consider non-linearities first, and then interactions before considering non-proportional odds. We began with a simple linear “main effects” proportional odds model containing age, poverty and indicator variables for African-American and Hispanic with white as the reference category. We then tested for nonlinear effects of poverty and race based on significance tests of second and third order polynomials finding non-linear effects for age but not poverty. We then investigated differences in age and poverty effects by race and ethnicity, i.e. interactions. The results led us to include interactions between the race/ethnicity indicators and linear age only but not poverty. We then tested for an interaction between age and poverty and found that it was significant. These various tests for interactions are correlated, of course, and we ran various within group models not detailed here to assess stability of findings.

Convinced that we had obtained a maximally parsimonious proportional odds model, we tested for non-proportional odds. For each variable, MIXOR provides a global test of equality of logits across the three thresholds defined by the four diagnostic stages. For non-nested partial proportional odds models, Akaike (AIC) information criteria were used for model comparison. The final partial proportional odds model has non-proportional fits for age, age², age³, black non-Hispanic, black non-Hispanic by age, and proportional fits for Hispanic, Hispanic by age. See Table 3 for details. The comparison of the final partial proportional odds model (AIC=59165.3) with the most parsimonious portioned odds model (AIC=58751.8) shows a highly significant p-value < 0.0001 with 10 degrees of freedom. The partial proportional odds model outperforms the proportional odds model. The final model takes the form:

$$\log \left[\frac{p(y_{ij} \leq c)}{1 - (p(y_{ij} \leq c))} \right] = \gamma_c - \left[\beta_0 + \beta_{1c} \text{age}_{ij} + \beta_{2c} \text{age}_{ij}^2 + \beta_{3c} \text{age}_{ij}^3 + \beta_{4c} \text{BNH}_{ij} + \beta_5 \text{HISP}_{ij} + \beta_{6c} \text{poverty}_{ij} + \beta_{7c} \text{BNH}_{ij} \times \text{age}_{ij} + \beta_{8c} \text{HISP}_{ij} \times \text{age}_{ij} + \nu_{0i} \right]. \quad (i=1, \dots, 1260; c=1, \dots, 3)$$

As described above, we estimated several more complex models involving nonlinear specifications of the poverty effects and interactions involving poverty and the race/ethnic classification. Although there was marginal evidence for some of these effects, we selected the model above based on the Bayesian Information Criterion (BIC). None of the marginally significant effects would have changed the substantive interpretation of the results.

The results shown below are “population averaged” as opposed to “subject specific” meaning that they show the effect of a one unit difference in a given covariate, controlling all others, averaged over all census tracts. “Subject specific,” on the other hand, means the effect of a one unit difference in a given covariate for a given person, or for different persons with equal values on all other covariates including the random effect, from the same census tract (Allison, 1999). The two estimates differ as a function of the variance of the random effect (Hu et al, 1998). Because the degree of clustering, as indexed by the intra-class correlation coefficient of .012 is quite low, subject-specific and population-averaged results are closely similar. Therefore, we show only the latter.

RESULTS

Descriptive statistics

Basic descriptive statistics are shown in Table 2. For each group, the table shows the percentage of cases diagnosed at each stage along with the mean age and proportion below the poverty

line for each group at each stage. Overall, the results show that black non-Hispanics are at a disadvantage with respect to regional and distant stage diagnosis. Also, black non-Hispanic women diagnosed at later stages tend to be younger than whites but Hispanics are even younger.

Given the complexity of the model, our interpretation of results will focus on both the coefficients and significance tests shown in Table 3 and the graphs of estimated and observed probabilities shown in Figures 2 and 3.

Age Effects

Overall, the effect of age on stage at diagnosis is non-linear, requiring a cubic polynomial for adequate fit (see age, age² and age³ in Table 3.) Age effects also differ by race/ethnicity as indicated by the significant interaction terms, and they differ across the cumulative logits (see black non-Hispanic*age and Hispanic*age in Table 3.). Given this complexity, we show fitted probabilities by age in Figures 2a and 2b holding poverty constant at the median. Figure 2a compares black non-Hispanics to white non-Hispanics and Figure 2b makes a similar comparison for white non-Hispanics to Hispanics. In order to assess goodness of fit, each graph also shows observed probabilities cumulated across ten-year age groups.^{iv}

The key finding is that although at any age, black non-Hispanic women are more likely to be diagnosed at later stages (regional or distant) and less likely to be diagnosed at earlier stages (*in situ* or localized) than white non-Hispanic women, the differences are more pronounced at younger ages, that is, the age effect is different for each of the three race-ethnic groups.. For example, at age 40, holding poverty constant at the race/ethnic group median, the estimated probability that a white non-Hispanic women's diagnosis will be at the regional stage is approximately .33, but for a black non-Hispanic woman of a comparable age the estimated probability is .41. Thus, the relative advantage of white non-Hispanic women holds throughout the age range, but declines with increasing age. The same pattern is found when distant stage diagnosis is examined. Among younger women, black non-Hispanics are about twice as likely to be diagnosed at the distant stage as white non-Hispanic women, but the difference gradually declines with advancing age.

A similar pattern holds when we compare Hispanic women to white non-Hispanic women, again holding poverty at the race/ethnic group-specific median. In general, Hispanics are more likely to be diagnosed at later stages relative to white non-Hispanics. As figure 2b shows, at age 40 the probability of being diagnosed at the regional stage is about .33 for white non-Hispanics and about .38 for Hispanics. Yet by about age 68, the estimated probabilities of diagnosis for white non-Hispanic and Hispanic women at any given stage are about the same. Unlike Black non-Hispanics, the difference between Hispanics and white non-Hispanics is not significant (Table 3).

Poverty Effects

The effects of poverty on stage of diagnosis are shown in Figures 3a and 3b. As with age, we show comparisons of white non-Hispanics and black non-Hispanics in Figure 3a and white non-Hispanics and Hispanics in Figure 3b. Each figure also shows observed probabilities. The fitted probabilities were obtained holding age constant at the race/ethnic group median. The findings are somewhat less complex than they are for age. In particular, poverty does not interact with race/ethnicity and age. Regardless of race/ethnicity, poverty has a strong effect on the probability of being diagnosed at the later stages. As poverty increases by ten percentage

^{iv}The non-linear nature of logistic regression models means that different plots for fitted probabilities with respect to age would be obtained if poverty were held at some other value than the median, say the 75th percentile. We have experimented with other values and find that our interpretation would be the same.

points, the odds of being diagnosed at a regional or distant stage increase by a factor of approximately 1.07, an effect that does not differ by race/ethnicity. The proportional odds model holds, such that this increment is the same at each of the thresholds.

Interaction Analysis

The previously reported analyses assume that the effects of poverty are the same regardless of race and ethnicity. Testing the interaction between poverty and indicators for race and ethnicity proved difficult. After some trial and error we concluded that there was not enough information on Hispanics to carry out a test. With regard to white non-Hispanics and black non-Hispanics, there is not much overlap in the distribution of poverty. Almost all whites live in low poverty areas and a very high proportion of black non-Hispanics live in relatively high poverty areas. For example, about four-fifths of the white population lives in areas where poverty is 8% or below, but only about a fifth of black non-Hispanics live in such areas. Thus, there is relatively little data on which to base tests of interactions. However, it is possible to obtain a test for that portion of the poverty distribution where there is an overlap using a piecewise regression model in which the model is fit separately for specific parts of the poverty distribution. For white non-Hispanics, we used a two-piece model with poverty equal to 15% as the cut-off point, which is the 95th percentile of the white poverty distribution. For black non-Hispanics, we used a three-piece model with cut points equal to 15% and 45%. These values correspond roughly to the 40th and 95th percentiles of the distribution of poverty for black non-Hispanics. The interaction could then be tested in the first segment of the model corresponding to approximately 95% and 40% of white non-Hispanic and black non-Hispanic cases respectively. That test was non-significant indicating that the effect of poverty on stage at diagnosis was the same within that segment of the poverty distribution where such a test could legitimately be carried out.

Goodness of Fit

There is no good single measure for evaluating the degree to which an ordinal regression model fits the data. In Figures 2 and 3, in addition to showing fitted probabilities for each stage, we show observed probabilities within ten-year age groups and seven-point poverty intervals, respectively. Observed poverty is grouped into seven intervals corresponding to the 20th, 40th, 60th and 80th quintiles with an additional point at the 95th percentile only for black non-Hispanics and Hispanics. The additional point does not include white non-Hispanics due to lack of overlap in poverty distribution after the 80th quintile for this race/ethnic group. The observed proportion of stage at diagnosis stratified by race/ethnicity and poverty interval is plotted at the midpoint of each interval across poverty. Except at the extremes, where there are relatively few cases, the data fit the model well and the fitted values are quite close to the observed probabilities.

DISCUSSION

Our analysis is based on associations using the conventional approach in the field to measure stage at diagnosis of breast cancer. That is, it reflects the proportion of stage-specific cases to the total number of cases within the race/ethnic group. An alternative analysis might examine stage-specific breast cancer incidence rates where the denominator consists of the population at risk. Conceivably, we might reach different conclusions with regard to disparities if stage-specific incidence rates were observed to differ by race/ethnicity and age. A thorough analysis of potential differences in the two approaches is beyond the scope of the present paper.

These results are broadly consistent with the existing literature in that minority and poor women are disadvantaged with more frequent advanced breast cancer at diagnosis. However, the analysis adds to our understanding of these disparities in several ways. We have a clearer

understanding of the complex relationship between race/ethnicity on one hand and stage at diagnosis on the other. By using an ordinal regression model, we were able to see how disparities at the various stages are interrelated. The interactions involving age, show that the disadvantage faced by black non-Hispanic women is most serious at younger ages. A similar but somewhat weaker result was apparent among Hispanic women. With regard to black-white comparisons, the finding of an age interaction parallels results by Grann et al., (2006) who found a similar interaction between race and age with respect to breast cancer mortality. It is probably no accident that racial disparities decline with age, particularly because after age 65 Medicare becomes available to a large proportion of the population.

With regard to poverty, we found that its effects are, within the limitations of the available data, very similar for all three race/ethnic groups. Our results suggest that the disparities found between middle class and poverty level white non-Hispanic women are also apparent among black non-Hispanic and Hispanic women. Given the relatively small degree of overlap in the distribution of wealth and poverty across race/ethnic groups in Cook County, this finding is tentative. All attempts to disentangle race and SES effects, and particularly to test an interaction between race and SES are plagued by this problem. However, particularly among younger women, a well-educated black non-Hispanic middle class is slowly emerging and with time more effective tests will eventually be possible.

We believe that the approach we have taken to arrive at a poverty estimate is useful. Future studies that link ABSM based on census tract data might well link cancer registry data (which is usually available by age, sex, race/ethnicity and tract) to the census tract subpopulation poverty data as we have done. Such data are available from public use census files, and result in a more accurate representation of patterns of social stratification within census tracts.^v On the other hand, there is an inherent tension between using ABSM's as though there were imputed estimates at the individual level and treating tract-level poverty as an indicator of some sort of ecological or neighborhood effect. Inevitably, there is a confound. We can not interpret the effect of poverty either as a property of individuals or as a characteristic of an ecological area. In this analysis, we have tried to find a poverty measure that is closest to a person-level imputation and have interpreted it as such. In principle, it ought to be possible to include both neighborhood level variables (remembering that census tracts are not necessarily "neighborhoods") and more disaggregated indicators as done here.

Stage at diagnosis of course is, to a great degree, a function of the availability and quality of screening mammography. Smith-Bindman et al., (2006), using data from a number of mammography registries around the country, show that black women receive less adequate mammography screening, in terms of both availability and quality, than white women. Blackman and Masi (2006) in assessing the "root causes" of black-white mortality differentials argue that disparities occur at each stage of care ranging from detection to the quality of treatment. From a policy perspective then, these results suggest that resources for early detection and prompt resolution of anomalous mammograms should be focused on younger minority women, particularly those residing in high poverty areas of Cook County. Although at present Cook County does not have a mammography registry, special tabulations of data from 1994–2000 Behavioral Risk Factors Survey data from Cook County show that non-white women are, overall, somewhat less likely to report ever having a mammogram (60.2% vs, 56.5%, OR 1.16, $p < .02$).^{vi}

^vBeginning in 2010 the traditional census "long form" will no longer be available, to be replaced by tract-level data, pooled across three annual surveys, from the American Community Survey. It remains to be seen if this data source will provide sufficient sample size to obtain race/ethnic and age-specific estimates at the tract level (Mather et al., 2005).

^{vi}We are indebted to Bruce Steiner of the Illinois Department of Public Health for making these data available.

Hirschman et al (2007) documented substantial disparities in breast cancer mortality in the City of Chicago between white non-Hispanics and black non-Hispanics. Our results extend their findings to a degree, by focusing on Cook County as a whole and bringing other factors in addition to race/ethnicity, such as age and residential poverty into play. An important point here is that health disparities of whatever kind are best understood within a well-defined medical care catchment area. For Cook County women, the county health care system is the provider of last resort. Women with adequate financial resources can, of course, obtain health care wherever they wish, within or outside the county, but women without resources are, with very few exceptions, unable to venture outside the county for screening, diagnosis and treatment. These are also the groups that experience the greatest disparities in mortality.

Acknowledgments

This research was funded by the U. S. National Cancer Institute grant (P-50 CA106743, to R. Warnecke, PI). The cancer data used in this study were supplied by the Illinois Department of Public Health. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions.

We are indebted to Young Ik Cho, Vincent Freeman, Donald Hedeker, Elizabeth Tarlov, Shannon Zenk and especially to Garth Rauscher for consultation and assistance with this paper. Yu-Li Hsieh, Kevin James and Mayumi Saegusa provided research assistance. We thank the reviewers for very helpful comments and suggestions.

References

- Allison, P. Logistic Regression Using the SAS System: Theory and Applications. SAS Institute Inc; Cary, N.C.S: 1999.
- Barry J, Breen N. The importance of place of residence in predicting late-stage diagnosis of breast or cervical cancer. *Health and Place* 2005;11:15–29. [PubMed: 15550353]
- Blackman DJ, Masi CM. Racial and ethnic disparities in breast cancer mortality: Are we doing enough to address the root causes? *Journal of Clinical Oncology* 2006;24:2170–2178. [PubMed: 16682736]
- Bradley CJ, Given CW, Roberts C. Race, socioeconomic status, and breast cancer treatment and survival. *Journal of the National Cancer Institute* 2002;94:490–496. [PubMed: 11929949]
- Breen N, Figueroa JB. Stage of breast and cervical cancer diagnosis in disadvantaged neighborhoods: a prevention policy perspective. *American Journal of Preventive Medicine* 1996;2:319–326. [PubMed: 8909640]
- Du XL, Fang S, Meyer TE. Impact of treatment and socioeconomic status on racial disparities in survival among older women with breast cancer. *American Journal of Clinical Oncology* 2008;31:125–132. [PubMed: 18391595]
- Grann V, Troxel AB, Zojwalla N, Herschman D, Glied SA, Jacobson JS. Regional and racial disparities in breast cancer-specific mortality. *Social Science & Medicine* 2006;62:347–357.
- Gumpertz ML, Pickle LW, Miller BA, Bell BS. Geographic patterns of advanced breast cancer in Los Angeles: associations with biological and sociodemographic factors (United States). *Cancer Causes and Control* 2006;17:325–339. [PubMed: 16489540]
- Havener, L.; Thornton, M., editors. Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Version 11.3. Vol. 13. Springfield, IL: North American Association of Central Cancer Registries; 2008.
- Hedeker DRD, Gibbons RD. MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods Programs Biomed* 1996;49:157–76.
- Hirschman J, Whitman S, Ansell D. The black:white disparity in breast cancer mortality: the example of Chicago. *Cancer Causes and Control* 2007;18:323–33. [PubMed: 17285262]
- Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz MA. Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology* 1998;147:694–703. [PubMed: 9554609]
- Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic

measures--the public health disparities geocoding project. *American Journal of Public Health* 2003;93:655–671.

Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *American Journal of Public Health* 2005;95:312–323. [PubMed: 15671470]

Krieger N, Waterman P, Chen JT, Soobader MJ, Subramanian SV, Carson R. Zip code caveat: bias due to spatiotemporal mismatches between ZIP codes and US census-defined geographic areas--the Public Health Disparities Geocoding Project. *American Journal of Public Health* 2002;92:1100–1102. [PubMed: 12084688]

Li C, Malone I, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Archives of Internal Medicine* 2003;163:49–56. [PubMed: 12523916]

Mandelblatt J, Andrews H, Kao R, Wallace R, Kerner J. Impact of access and social context on breast cancer stage at diagnosis. *Journal of Health Care for the Poor and Underserved* 1995;6:342–351. [PubMed: 7548491]

Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. *American Journal of Public Health* 1991;81:646–649. [PubMed: 2014871]

Mather, M.; Rivers, KL.; Jacobsen, LA. *Population Bulletin*. Vol. 60. Washington, DC: Population Reference Bureau; 2005. The American Community Survey.

Merkin SS, Stevenson L, Powe N. Geographic socioeconomic status, race, and advanced-stage breast cancer in New York City. *American Journal of Public Health* 2002;92:64–70. [PubMed: 11772763]

Miller BA, Hankey BF, Thomas TL. Impact of sociodemographic factors, hormone receptor status, and tumor grade on ethnic differences in tumor stage and size for breast cancer in US women. *American Journal of Epidemiology* 2002;155:534–545. [PubMed: 11882527]

NAACCR Latino Research Work Group. NAACCR Guideline for Enhancing Hispanic/Latino Identification. Revised NAACCR Hispanic/Latino Identification Algorithm (NHIA v2). North American Association of Central Cancer Registries; Springfield, IL: 2005.

Newman LA, Griffith KA, Jatoi I, Simon MS, Crowe JP, Colditz GA. Meta-analysis of survival in African American and white American patients with breast cancer: Ethnicity compared with socioeconomic status. *Journal of Clinical Oncology* 2008;24:1342–1349. [PubMed: 16549828]

Rabe-Hesketh S, Pickles AA, Skrondal A. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modeling* 2003;3:215–232.

Raudenbush, SW.; Bryk, AS. *Hierarchical Linear Models*. Sage; Thousand Oaks, C.A: 2002.

Richardson J, Langholz LB, Bernstein L, Burciaga C, Danley K, Ross RK. Stage and delay in breast cancer diagnosis by race, socioeconomic status, age and year. *British Journal of Cancer* 1992;65:922–926. [PubMed: 1616865]

Ries, L.; Melbert, D.; Krapcho, M.; Mariotto, A.; Miller, B.; Feuer, E.; Clegg, L.; Horner, MJ.; Eisner, MP.; Reichman, M.; Edwards, BK. *SEER Cancer Statistics Review, 1975–2004*, based on November 2006 SEER data submission. [accessed on Dec. 15 2007]. Available at: http://seer.cancer.gov/csr/1975_2004/

Schwartz K, Crossley-May LH, Vigneau FD, Brown K, Banerjee M. Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes and Control* 2003;14:761–766. [PubMed: 14674740]

Smith-Bindman R, Miglioretti DL, Lurie N, Abraham L, Barbash RB, Strzelczyk J, Dignan M, Barlow WE, Beasley CM, Kerlikowske K. Does utilization of screening mammography explain racial and ethnic differences in breast cancer? *Annals of Internal Medicine* 2008;144:541–53. [PubMed: 16618951]

Wells BL, Horm JW. Stage at diagnosis in breast cancer: race and socioeconomic factors. *American Journal of Public Health* 1992;82:1383–85. [PubMed: 1415866]

Appendix

Appendix

Empirical Bayes Estimation

We began with a simple multilevel logistic regression model of the following form:

$$\log \left[\frac{P(y_{ij}=1)}{P(y_{ij}=0)} \right] = \beta_0 + \theta_j.$$

where $y_{ij} = 1$ if the i^{th} person in the j^{th} group is below the poverty line and 0 otherwise. Age in years is categorized in three groups: 25–44, 45–64 and 65+. We restricted the analysis to women. There are a total of 7,591 groups classified by tract by age by race/ethnic group. As formulated here, this is a standard “empty” model in which the intercept, β_0 , represents the overall mean and θ_j represents the deviation from the mean for the j^{th} group. The conventional model assumes that the θ_j are normally distributed, from which one can obtain empirical Bayes estimates which “shrink back” the estimate toward the grand mean relative to the amount of information available, i.e., the sample size. However, for problems of this kind, the usual normality assumption, implying that each θ_j is a random draw from a normal distribution, is untenable because in 2,540 groups (33.5%) all women are above the poverty line and 159 groups (2.2%) all women are below the poverty line. Because of this, we used a non-parametric empirical Bayes method described by Rabe-Hesketh et al. (2003) and implemented in Stata. Using this approach, we were able to obtain poverty estimates for all 7,591 groups.

Figure 1.
Proportion of whites and blacks below poverty line in Cook County census tracts

Figure 2.

Figure 2a Fitted and observed stage at breast cancer diagnosis by age for women aged 30–89^a

^aNon-Hispanic Black (red lines) and non-Hispanic white (black lines), poverty held at median for each group.

Figure 2b Fitted and observed stage at breast cancer diagnosis by age for women aged 30–89^a

^aHispanic (red lines) and non-Hispanic whites (black lines), poverty held at median for each group.

Figure 3.
Figure 3a Fitted and observed stage at breast cancer diagnosis by poverty level for women aged 30–89^a

^aNon-Hispanic Black (red lines) and non-Hispanic white (black lines), poverty held at median for each group.

Figure 3b Fitted and observed stage at breast cancer diagnosis by poverty level for women aged 30–89^a

^aHispanic (red lines) and non-Hispanic whites (black lines), poverty held at median for each group.

Table 1

Proportion of Female Population Below Poverty Level by Race/Ethnicity and Age Group, Cook County, Illinois, Decennial Census 2000

Age Group		Black non-Hispanic	Hispanic	White non-Hispanic	Other non-Hispanic	Total
25–44	n	212,774	169,580	400,293	80,227	862,874
	Poverty rate (%)	24.2	17.3	4.9	12.5	12.8
45–64	n	148,393	66,957	321,603	44,969	581,922
	Poverty rate (%)	19.1	12.9	4.7	9.1	9.6
65+	n	73,124	19,588	252,727	16,389	361,828
	Poverty rate (%)	22.8	17.3	7.7	20.8	11.8
Total	n	434,281	256,125	974,623	141,585	1,806,624
	Poverty rate (%)	22.2	16.2	5.5	12.4	11.6

Source: United States Bureau of the Census

Table 2
Descriptive Statistics for Stage at Diagnosis of Breast Cancer by Race/Ethnicity and Age, Cook County, Illinois, 1994 – 2000

Stage at Dx	Variable	Race/Ethnicity			Total
		Black non-Hispanic	Hispanic	White non-Hispanic	
0 (<i>in situ</i>)	Frequency (%)	773 (14.0)	215 (16.0)	3131 (16.5)	4119 (15.9)
	Mean age	59.5	54.5	60.8	60.2
	Poverty (%)	19.0	12.1	5.4	8.3
1 (Localized)	Frequency (%)	2,533 (45.9)	640 (47.6)	10,217 (53.7)	13,390 (51.7)
	Mean age	60.1	55.0	64.5	63.2
	Poverty (%)	19.9	14.8	5.9	9.0
2 (Regional)	Frequency (%)	1,754 (31.8)	432 (32.1)	4,729 (24.8)	6,915 (26.7)
	Mean age	56.4	51.7	61.5	59.6
	Poverty (%)	21.3	13.5	6.0	10.3
3 (Distant)	Frequency (%)	457 (8.3)	57 (4.3)	962 (5.0)	1,476 (5.7)
	Mean age	59.3	55.2	66.5	63.8
	Poverty (%)	23.1	16.4	6.8	12.2
Total	Frequency (%)	5,517 (100.0)	1,344 (100.0)	19,039 (100.0)	25,900 (100.00)
	Mean age (yrs)	58.8	53.9	63.2	61.8
	Poverty (%)	20.5	14.0	5.9	9.4

Source: U. S. Bureau of the Census and Illinois Department of Public Health

Table 3

Results for Non-Proportional Ordered Logistic Random Effect Regression Model

	(Population average estimates)			
	logit	logit-1	logit-2	logit-3
Intercept	1.530 (β_0)			
Threshold 2	2.393 (γ_2)			
Threshold 3	4.527 (γ_3)			
Age [‡]		0.248 (β_{11})	-0.053 (β_{12})	0.219 (β_{13})
Age ²		0.073 (β_{21})	0.033 (β_{22})	0.018 (β_{23})
Age ³		-0.025 (β_{31})	-0.005 (β_{32})	-0.009 (β_{33})
Black non-Hispanic *		0.084 (β_{41})	0.279 (β_{42})	0.492 (β_{43})
Hispanic *	0.016 (β_5)			
Poverty [‡]	0.067 (β_6)			
Black non-Hispanic * age		-0.172 (β_{71})	-0.076 (β_{72})	-0.142 (β_{73})
Hispanic * age	-0.086 (β_8)			

Bold indicates significant at $\alpha/2=0.05$ level. Threshold 1 (γ_1) is defined to be 0 for identification.

* White non-Hispanic is the comparison group.

[‡] Age is centered and rescaled.

[‡] Poverty is the Empirical Bayes estimate for the proportion below poverty line, centered and rescaled.