

Title:

**Deprivation and Segregation in Ovarian cancer survival among African American Women: a mediated analysis**

## Authors:

Andrew B. Lawson<sup>1,2</sup>, Joanne Kim<sup>1</sup> Courtney Johnson<sup>4</sup>, Theresa Hastert<sup>5</sup>, Elisa V. Bandera<sup>12</sup>, Anthony J. Alberg<sup>6</sup>, Paul Terry<sup>11</sup>, Maxwell Akonde<sup>6</sup>, Hannah Mandle<sup>4</sup>, Michele L. Cote<sup>8</sup>, Melissa Bondy<sup>7</sup>, Jeffrey Marks<sup>10</sup>, Lauren Peres<sup>9</sup>, Joellen Schildkraut<sup>4</sup>, Edward S. Peters<sup>3</sup>

## Affiliations:

<sup>1</sup>Department of Public Health Sciences, College of Medicine, Medical University of South Carolina, Charleston, SC, 29425, USA

<sup>2</sup>Usher Institute, Centre for Population Health Sciences, Edinburgh Medical School, University of Edinburgh, UK

<sup>3</sup> Department of Epidemiology, College of Public Health, University of Nebraska Medical Center, Omaha, Nebraska, USA

<sup>4</sup> Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

<sup>5</sup> Department of Oncology, Wayne State University School of Medicine, Population Studies and Disparities Research Program, Karmanos Cancer Institute, Detroit, Michigan, USA

<sup>6</sup> Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

<sup>7</sup> Department of Epidemiology and Population Health, Stanford University, College of Medicine, Stanford, California, USA

<sup>8</sup> Indiana University Melvin and Bren Simon Comprehensive Cancer Center, Indianapolis, Indiana, USA

<sup>9</sup> Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

<sup>10</sup> Department of Surgery, Duke University School of Medicine, Durham, NC, USA

<sup>11</sup> Department of Medicine, University of Tennessee Medical Center-Knoxville, Knoxville, TN, USA

<sup>12</sup> Cancer Epidemiology and Health Outcomes, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA

**Keywords:** ovarian cancer; AACES1 cohort; survival; diagnostic delay; deprivation indices; Yost; ICE; causal mediation; Bayesian; Structural equation; Mixed effect models.



## **Abstract**

### Background:

Deprivation and segregation indices are often examined as possible explanations for observed health disparities in population-based studies. In this study we assessed the role of recognized deprivation and segregation indices specifically as they affect survival in a cohort of self-identified Black women diagnosed with ovarian cancer who enrolled in the African American Cancer Epidemiology Study (AACES).

### Methods:

Mediation analysis was used to examine the direct and indirect effects between deprivation or segregation and overall survival, via a Bayesian structural equation model with Gibbs variable selection.

### Results:

The results suggest that high SES -related indices have an association with increased survival, ranging from 25%-56%. In contrast, ICE-race does not have a significant impact on overall survival. In many cases, the indirect effects have very wide credible intervals, consequently the total effect is not well estimated despite the estimation of the direct effect.

### Conclusions:

Our results show that black women living in higher SES neighborhoods are associated with increased survival with ovarian cancer, using area-level economic indices such as Yost or ICE-income. In addition, the KOLAK urbanization index has a similar impact, and the highlights importance of area-level deprivation and segregation as potentially modifiable social factors in ovarian cancer survival.

## INTRODUCTION

It is widely recognized that despite having a lower incidence of ovarian cancer black women experience poorer survival than white women. Although the reasons for this disparity are not clear, studies<sup>1</sup> have suggested that multiple social factors may play a role including socioeconomic factors. In recent years, a variety of area level deprivation indices (DIs) have been widely used in population-based epidemiologic studies as novel tools to examine the impact of neighborhood SES on a variety of health outcomes. Mechanisms by which neighborhood SES influences health could be indirectly through limited health care access or directly from cumulative biological effects of chronic stress. A recent extensive survey of the use of deprivation indices in cancer studies has been conducted.<sup>2</sup> In that review, a number of commonly used indices (24) of deprivation were found to be associated with a range of cancer types in terms of incidence and survival or mortality. A few of these indices were most frequently used, such as the Yost, Concentrated Disadvantage Index (CDI), Area Deprivation Index (ADI), and Diez-Roux. In addition, the effects of structural racism and racial discrimination are increasingly being examined as a predictive factors in health disparities research. Segregation indices (SIs) can be formed based on race or SES and focus on the disparity between ethnicities or income at a local geographic level.

In the current study, we have sought to assess the effect of DIs and SIs on the survival of Black women diagnosed with ovarian cancer. In particular, we examined a cohort of Black epithelial ovarian cancer patients from the African American Cancer Epidemiology Study (AACES).<sup>3</sup> It has been observed that contextual areal effects, such as the effects of residing in a particular locale or neighborhood, however, defined, can have an impact on cancer outcomes.<sup>4</sup> In a previous analysis of patient characteristics (including histology), a disadvantage index measured at the census tract level was observed to be associated with ovarian cancer outcomes.<sup>5</sup> In subsequent work,<sup>6</sup> a disadvantage index at the census tract level was associated with poorer survival and later-stage outcomes in an ethnically mixed population. In addition, a significant relationship has been found between DI and ovarian cancer survival at the state and national levels.<sup>7</sup> Previous studies in breast cancer, lung cancer, and prostate cancer have also found consistent relationships between disadvantage or deprivation and survival.<sup>8,9,10,11</sup> For this report, we have selected a set of DIs and SIs which represent different aspects of neighborhood SES context. Our goal is to examine the association of differing attributes of DIs and SIs in relation to ovarian cancer survival metrics, including histologic subtype and diagnosis delays, and to consider the role of confounding and potential mediation in black women enrolled in AACES.

Our major hypotheses concerning DI and SIs are that these composite measures representing complex socio-economic status and social segregation neighborhood features have an important influence on the survival experience of women diagnosed with ovarian cancer whilst being mediated by important clinical and diagnostic factors.

To address these hypotheses, a suitable spatial scale must be assumed for the DI and SIs, which defines their context. We have chosen the census tract as that scale, and this is further discussed in the methods section below. In this paper, we pose the question: does living in a census tract with high deprivation or segregation lead to worsened survival from ovarian cancer in Black women?

## **METHODS**

### **Study population.**

Phase 1 of the AACES study includes 11 US sites that were selected based on geographic regions with large African-American populations and to provide geographic diversity. These included Alabama, Georgia, Louisiana, metropolitan Detroit, Michigan, New Jersey, North Carolina, Ohio, South Carolina, Tennessee, Illinois, and Texas. Women diagnosed with epithelial ovarian cancer were eligible to participate if they self-identified as African American or Black, were 20-79 years of age at diagnosis, and could complete an interview in English. Eligible cases had histologically confirmed epithelial ovarian cancer (EOC) and were centrally reviewed by an expert study pathologist. Rapid-case-ascertainment (RCA) was used to identify the cases through state cancer or Surveillance, Epidemiology, and End Results (SEER) registries, and through gynecologic oncology departments at individual hospitals. Participants were newly diagnosed with EOC between December 2010 and December 2015. The average and median survival time for all phase 1 cases, as of 2021, is 4.79 years and 4.81 years, respectively. The survival rates for this cohort are consistent with the observed survival rates from national SEER data,<sup>12</sup> when conditioning on surviving at least 10 months past diagnosis. This accounts for the fact that rapidly fatal cases do not make it into the study, even when employing RCA. Approximately 50% of the phase 1 cases have participated in at least one follow-up survey. Altogether, 592 epithelial ovarian cancer cases were enrolled in phase 1. Of 1,720 cases attempted to contact, 1,199 (70%) were actively reached. Of these, 592 (49%) were successfully interviewed, and 388 (32%) actively refused. We are currently recruiting additional epithelial ovarian cancer cases in AACES phase 2, but these are excluded from the current analysis due to the brief follow-up time.

### **Deprivation and Segregation indices.**

Deprivation indices are composite measures, usually based on socio-economic indicators. They are usually available within a range of small area units such as block groups (BGs), census tracts (CTs), or zip codes. In this study, we limit our focus to census tracts only, allowing for consistency across the available measures. Segregation measures address the division within an area between different population groups based on some key characteristics. These characteristics could be race-based or based on socio-economic status of the area. The following is a description of the chosen census tract DIs and SIs and their rationale for use.

### **Yost index.**

The Yost index has been constructed as an SES index via principal components analysis of seven variables.<sup>13</sup> This index is widely used in cancer studies and is available at several geographic aggregation levels via the SEER database program (<https://seer.cancer.gov/seerstat/databases/census-tract/index.html>). This index was found to be significantly related to breast cancer incidence for a race stratified/adjusted analysis in the developers' original study<sup>13</sup> and other cancers in subsequent studies.<sup>14</sup> A higher Yost value represents a higher SES level. We have used the index computed for a range of years: 2006–2010 consistent with our case diagnosis years. The index is in standardized units.

### **Kolak urbanization (URB) measure**

A recent study by Kolak of the social determinants of health<sup>15</sup> proposed using four dimensions that encapsulate the variation in these determinants across spatial units within the US. The measures were derived from dimension reduction. The URB core dimension addresses urbanicity and reflects the degree to which an area has high rent, high income, and transport infrastructure and high costs for low-income residents. Unlike the Yost index, this measure does not directly reflect socio-economic status, but reflects a mix of factors related to urbanicity and hence urban and rural differences. The index is in standardized units.

### **ICE measures**

The index of concentration at the extremes (ICE) are segregation measures (SIs) that can be computed at a given small area unit such as a block group or census tract.<sup>16,17</sup> The indices estimate the differences in proportions of one group compared to an alternate group within an area relative to the total population, such as the proportion of the population that is Black compared to the White proportion. ICE-race is based on this comparison. ICE- income compares the lowest SES group in an area with the highest SES group. The calculation is based on the formula:

$$ICE = \frac{U - L}{T}$$

where  $U$  is the number of persons in the uppermost category,  $L$  is the number in the lowest, and  $T$  is the total population in the region in question. The ICE measures were computed from the 2010 census and the ACS 2006-2010. The indices range from -1 to +1 and a unit change represents a shift of 50% in the population structure.

### **The Spatial Context**

In this study, we focus on a set of DIs as well as more recently developed segregation measures (SIs). It is important to consider the spatial level of resolution at which these indices are studied. The spatial context of deprivation is important as the influence of disadvantage on health could operate differently across various spatial scales. The EOC participants enrolled in phase 1 of AACES have been geo-coded to the census tract (CT) level. The social determinants of Health (SDOH) index (from now on Kolak indices), are only available at CT level.<sup>15</sup> For these reasons we have consider census tract measures as most relevant in our study.

AACES 1 consists of a sample of population-based registries. These registries are sparsely distributed across a selection of US states, and thus the spatial context is restricted. Our approach includes using observed individual predictors (such as age body mass index, physical activity self-reported SES) and also unobserved contextual effects within the spatial unit which allow for extra variation. In the modeling which follows we exploit a rich class of Bayesian hierarchical models (BHM) which include predictors and random effects to allow for contextual confounding.

### **Mediation Methodology**

To assess the impact of DIs and SIs on survival outcomes we apply a Bayesian causal mediation approach.<sup>18,19,20</sup> In this approach, a structural equation model is assumed whereby any mediator is modelled jointly with the outcome. In these joint models, the outcome is assumed to be a function of an exposure, a mediator, potential measured confounders, and a random effect that represents unobserved confounding. In this case, a Weibull end-point distribution is assumed where the log hazard is modelled with a linear predictor. In the typical mediator model, a dependence on exposure, confounders, and a random term representing unobserved extra variation is included. One of the significant advantages of Bayesian hierarchical models (BHM) is the ability to incorporate unobserved variation via random effects.<sup>21</sup> Note that the same random effect is shared and scaled in each of the outcome and mediator models. The mediation model structure<sup>22,232</sup> allows for the estimation of direct and indirect effects of predictors and we examine the paths within the directed dependence graph. The direct and indirect effects are estimated initially on the log scale and the relation between these



is assumed to be  $total\ effect = direct\ effect + mediator * mediator$  parameter. The result is then exponentiated to give the odds estimate.

In the above model, a linear predictor for confounders is usually assumed to be fixed. However, we also consider an extension whereby we examine a range of putative models for confounders during our estimation procedure. A Bayesian method that allows the sampling of all models within the computational method of Markov chain Monte Carlo (MCMC) is Gibbs variable selection.<sup>24</sup> By allowing the sampling of different combinations of confounders, it is possible to estimate inclusion probability for any given variable. When the estimation process is complete, any confounder achieving an inclusion probability greater than 0.5 is selected for a final model.<sup>25</sup> The final model is fitted with only those confounders reaching this threshold. This avoids the use of stepwise selection methods and allows for all possible combinations of confounders to be assessed. The models are fit using the Nimble R package.<sup>26</sup> Additional details of the statistical approach can be found in the Appendix.

### **Mediation analysis Applied in AACES.**

Outcomes: We consider as our outcome the survival of the cohort participants. Here, survival is defined as the time in days from diagnosis to the date of death or last contact, updated through 2021.

Independent Variables: Our focus is on the effect of deprivation and segregation on EOC survival. As DIs or SIs are usually included as single adjustments in models, we assumed that, for our evaluation purposes, each index should be treated as the primary independent variable and fitted separately with mediators and confounders.

#### Mediators:

Potential mediators were chosen *a priori* based on existing knowledge of their importance in EOC survival. The mediators included in this study were histology and diagnostic delay. Histology is regarded as an important prognostic indicator in EOC survival.<sup>27</sup> Epithelial ovarian cancer histotype was determined through expert pathology review and was collapsed into high-grade serous or other for analysis. Additional mediators considered initially in this study included stage at diagnosis and residual disease (based on debulking status). Stage at diagnosis, is often used to reflect the severity of the disease.<sup>27</sup> However, histology and stage are closely correlated and display similar mediation results, so we only discuss histology for brevity.

Diagnostic delay is included as a mediator as it could impact the severity of the diagnosed cancer and may be related to the socio-cultural environment of the locale or neighborhood. Diagnostic delay was defined in terms of the length of symptoms before the

diagnosis. Ten symptoms were included in the baseline questionnaire. Specifically, participants were asked about symptoms in the year prior to diagnosis including pelvic or abdominal discomfort, changes in bowel movements, urinary issues, distended or hard abdomen, lump in the abdomen, fatigue, side or back pain, bleeding not related to menstruation, weight gain, and nausea. Diagnostic delay was summarized as the longest symptom duration before diagnosis, ranging from 1 to 12 months.

### Confounders:

We examined a range of both outcome and mediator confounders relevant in the analysis of survival outcomes: body mass index (BMI), smoking status, physical activity, age at diagnosis, and self-reported gross annual family income. BMI was calculated as  $\text{kg/m}^2$  from self-reported height and weight at a year prior to diagnosis. Smoking status (never smoker, former smoker, or current smoker) physical activity (amount of weekly physical activity (mild, moderate, and strenuous) reported in the year prior to diagnosis) was further categorized to capture inactivity (participants were considered inactive if their total weekly physical activity was less than 2 hours, and active if their total weekly physical activity was 2 hours or more), and family income (less than \$10,000, \$10,000-\$24,999, \$25,000-\$49,999, \$50,000-\$74,999, \$75,000-\$100,000, and more than \$100,000).

## **RESULTS**

Basic descriptive summaries of the selected clinical characteristics of the cohort appear in Table 1, and summaries of the DIs and Sis, and confounders are shown in Table 2. To frame the discussion, we note that significantly longer survival times are reflected as hazard ratios less than one. Any effect that significantly increases the hazard will show a ratio bounded above one. The original hazard is measured on the log scale, so the parameter estimates reported in Tables 3 and 4 have been exponentiated to yield odds. The estimates are reported with 95% credible intervals.

Table 3 displays the result of fitting models separately to the four indices as exposures with histology as mediator. The final model with selected confounders included only age for the URB and Yost indices, whereas the ICE indices had age and BMI. Notably, neither self-reported gross annual family income, smoking, nor physical activity are selected when histology is the mediator.

The major results in Table 3 illustrate that URB, Yost, and ICE-income are all associated with improved survival. ICE-race was also associated with improved survival but not statistically

significant (OR = 0.736; 95% credible interval: 0.509-1.049). The results show that the Yost index, representing lower levels of deprivation, is associated with a 25% higher overall ovarian cancer survival per unit index (indirect effect: 1.014, 95% CI: 0.65, 1.79). The URB index, representing lower levels of deprivation and higher urbanicity, is associated with a 33% higher overall survival (indirect effect: 1.00, 95% CI: 0.64, 1.50). ICE-income, which represents higher levels of income segregation and a larger high income population, is associated with a 56% increase in ovarian cancer survival (indirect effect: 0.93, 95% CI: 0.0,  $1.47 \times 10^{14}$ ). ICE-race, representing higher levels of segregation by race, showed no significant association with ovarian cancer survival. None of the indices we examined were mediated by histology when controlling for age and BMI. The indirect effects are also weakly estimated and centered on one, in some cases with wide credible intervals, thus total effects have wide intervals because of this. This suggests that histology has a limited effect as a mediator with the selected Dis and SIs. Note that the average effect of histotype is also not significant. In a separate analysis with the stage at diagnosis (data not shown), stage was also not found to be significant as a mediator. In summary, older age is associated with worse survival, whereas increases in the Dis and SIs are associated with longer survival, although the association for ICE-race was not statistically significant.

Table 4 displays the results of fitting models separately to the four indices as exposures with diagnostic delay as a mediator. Note that diagnostic delay is a continuous variable. In this case, age and BMI were selected for the final model for all indices. The direct and total estimated effects are more precisely estimated as the indirect effects have very narrow credible intervals, despite the indirect effects lacking statistical significance. URB, Yost, and ICE-income all have reduced odds and associated longer survival. ICE-race again reduces odds but does not achieve significance (OR = 0.88, 95% credible interval: 0.718-1.074). In addition, under this mediator, the ICE-income achieves the biggest reduction in survival compared with the other indices (0.385). However, a unit change in ICE represents a large population change (50%). Examination of the distribution of diagnostic delay times demonstrates that shorter delays are often associated with early mortality, and longer delays are found in the group with longer survival. Hence, this mediator displays a marginally enhanced effect on survival time. Similar interpretations for the four indices, as cited above in Table 3, can be made for the diagnostic delay mediator in Table 4.

## DISCUSSION

Our findings here suggest several important results for research on neighborhood effects on ovarian cancer survival. We observe that DIs and SIs do impact survival, although the degree of their effects differs. URB, Yost, and ICE-income all have a positive association with longer survival. This suggests that there is a 'bootstrap' (or health improvement) effect of living within a highly urbanized and high SES area with diverse income levels. The most marked direct effect observed is for ICE-income when a diagnostic delay is used as the mediator. However, this index also has an important impact when histotype is used as a mediator, demonstrating a nuanced relation between deprivation and survival, perhaps because early medical intervention often occurs when diagnosed at an advanced stage which is associated with worse survival even when the diagnosis is delayed.

In general, the different segregation indices differ in their association with survival. We observed that the SES-related index (ICE-income) displays a significant association with survival, with greater segregation income associated with improved survival. In contrast, the racial index (ICE-race) does not show any significance. This contrasts with other studies where ICE-race was observed to relate to racial disparity in an ethnically diverse population.<sup>27</sup> However, within our cohort, there was limited evidence of racial segregation (as measured by ICE-race) associated with survival. This could reflect the ethnic focus of the AACES cohort, while survival differences are often displayed when Black people are compared to White people in other studies.<sup>28</sup> This suggests that even concentrated wealth regardless of race has an effect on survival. The ranking of indices in terms of the greatest reduction in odds differs by a mediator. Yost and ICE-Income are lowest for diagnostic delay, and for histology, URB and ICE-income are lower.

Ross et al.<sup>14</sup> used the Yost index for SES to explore racial disparities in ovarian cancer survival and found that while SES may be associated with survival disparities by race, these disparities remained when adjusting for the Yost index. Hodeib et al.<sup>29</sup> found that the Yost index was also a predictor of treatment guideline adherence for women with early-stage ovarian cancer. Our study shows that Yost SES at the census level has a significant direct effect overall effect on survival regardless of the mediator.

There are few additional studies investigating the role of other deprivation indices in ovarian cancer survival. An exception is Hufnagel et al.<sup>7</sup> where an increase in area deprivation index

(ADI) was found to affect ovarian cancer survival negatively. A statistically significant association has been found between ovarian cancer-specific survival and Cagney's concentrated disadvantage factor (CDI),<sup>5,30</sup> which attenuates the disparity in survival between Black women and White women. A higher concentrated disadvantage was associated with poorer survival. Neither histology nor stage was found to have significant mediation effects with the chosen indices, perhaps because there is no effective screening protocol for ovarian cancer<sup>31</sup>.

Mediation effects differ according to the mediator we tested but generally do not add information to the relationship between these indices and survival. The direct effects dominate the total effect of the deprivation and segregation indices. Both histology and diagnostic delay have relatively narrow indirect credible intervals, and diagnostic delay has well-estimated total effects. Notably, ICE-income shows a considerable odds reduction of the hazard rate when mediated by diagnostic delay. This may suggest that a 'bootstrap' effect is in play, whereby differences in higher SES neighborhoods engender more prolonged survival. The URB index also shows a significant total effect with this mediator, which may reflect a similar bootstrap effect due to high urbanicity.

It should be noted however, that the mediation results cited are conditional on the presence of selected confounders. For example, while the inclusion of age and BMI with histology as a mediator led ICE-income to have much reduced odds (0.439), translating to a 56% increase in survival, the same confounders with diagnostic delay as a mediator led the ICE-income index to further improve survival (0.385), with a 62% increase in survival. Age shows a significant reduction in survival with histology, whereas it only shows a marginal improvement in survival under diagnostic delay. BMI is not selected under histology as a mediator, whereas with diagnostic delay, it is selected and shows a significant worsening of survival for URB and ICE-income. The importance of BMI and age in relation to the role of inflammation has been noted for this cohort.<sup>32</sup> A previous report using ICE-income/race was for breast cancer and found a significant association with the presence of ER+ status at the county level.<sup>16</sup> It is also notable that ICE-race are not found to be very well estimated under any mediator in our study. For ICE-income the biggest change was under diagnostic delay.

## Conclusions

This mediated analysis of DIs and SIs has wide-ranging implications for both neighborhood effects on ovarian cancer survival and the relative effectiveness of these indices when used in conjunction with different mediators and sets of confounders. First, commonly used indices such

as Yost, ICE-income, and URB at the CT level, all significantly impact the survival experience of Black women diagnosed with epithelial ovarian cancer.

If the focus is to adjust for factors that could confound relations in ovarian cancer survival studies, then deprivation or segregation indices should be considered important general adjusters. In terms of confounders, age, BMI, or age alone were often selected within our variable selection strategy. However, self-reported SES, smoking, or physical activity did not appear as important confounders. Also, different mediators can affect the adjustment achieved by the use of DIs and SIs. While histology is important in explaining survival variation,<sup>28</sup> the different mediators behaved differently. The indices most strongly associated with survival in our analyses are Yost and ICE-income when mediated by diagnostic delay and URB and ICE-income when histology is present. ICE-race does not provide significant adjustment in survival.

There are a number of caveats to these conclusions which should be stressed. First, all analyses were based on census tract-level measures, so results could vary if other spatial units, such as block group or county, were chosen. Our cohort is purely self-identified Black women with ovarian cancer, and so disparities between races are not our focus, although there are results reported previously that found the ethnic disparity in survival for ovarian cancer was significantly associated with neighborhood disadvantage.<sup>30</sup> Our mediated analysis did not examine multiple mediation as we confined our focus to single mediators. While multiple mediation is certainly feasible within our strategy, our focus on evaluating DIs and SIs limited our approach. In addition, our results are conditional on the inclusion of selected important confounders (mainly age and BMI) but also included random effects that allowed for unobserved confounding. Any analysis that did not adjust for these predictors would likely have extra noise. The inclusion of random effects is therefore crucial in allowing for this additional variation.

Finally, the most important outcomes are in validating contextual indices and their essential role in affecting survival among Black women diagnosed with ovarian cancer. The census tract appears to yield significant survival associations for both DIs and SIs. We have confirmed that the much used Yost index does indeed provide a validated indicator for extended survival, as does the Kolak urbanization dimension. The ICE indices also provide important stratification associations. These results attest to the utility of the Bayesian mediation strategy used and also to the relevance of the census tract as an important contextual spatial unit.

## 9 References

- [1] Karanth S, Fowler ME, Mao X, Wilson LE, Huang B, Pisu M, Potosky A, Tucker T, Akinyemiju T. Race, Socioeconomic Status, and Health-Care Access Disparities in Ovarian Cancer Treatment and Mortality: Systematic Review and Meta-Analysis. *JNCI Cancer Spectr.* 2019 Oct 9;3(4):pkz084. doi: 10.1093/jncics/pkz084. PMID: 31840133; PMCID: PMC6899434.
- [2] Sorice KA, Fang CY, Wiese D, Ortiz A, Chen Y, Henry KA, Lynch SM. Systematic review of neighborhood socioeconomic indices studied across the cancer control continuum. *Cancer Med.* 2022 May;11(10):2125-2144. doi: 10.1002/cam4.4601.
- [3] Schildkraut, J. M., Alberg, A. J., Bandera, E. V., Barnholtz-Sloan, J., Bondy, M., Cote, M. L., Funkhouser, E., Peters, E., Schwartz, A. G., Terry, P., Wallace, K., Akushevich, L., Wang, F., Crankshaw, S., & Moorman, P. G. (2014). A multi-center population-based case-control study of ovarian cancer in African-American women: the African American Cancer Epidemiology Study (AACES). *BMC cancer*, 14, 688.
- [4] Hastert TA, Beresford SAA, Sheppard L, et al (2015) Disparities in cancer incidence and mortality by area-level socioeconomic status: a multilevel analysis. *J Epidemiol Community Health* 2015;69:168-176.
- [5] Peterson, C.E., Rauscher, G.H., Johnson, T.P. et al. The association between neighborhood socioeconomic status and ovarian cancer tumor characteristics. *Cancer Causes Control* 25, 633–637 (2014). <https://doi.org/10.1007/s10552-014-0357-7>
- [6] Peterson C.E, Rauscher G.H, Johnson T.P, et al. The effect of neighborhood disadvantage on the racial disparity in ovarian cancer-specific survival in a large hospital-based study in cook county, Illinois. *Front Public Health.* 2015;3:8. Published 2015 Jan 22.
- [7] Hufnagel DH, Khabele D, Yull FE, Hull PC, Schildkraut J, Crispens MA, Beeghly-Fadiel A. Increasing Area Deprivation Index negatively impacts ovarian cancer survival. *Cancer Epidemiol.* 2021 Oct;74:102013. doi: 10.1016/j.canep.2021.102013.
- [8] Hossain F, Danos D, Prakash O, et al. Neighborhood Social Determinants of Triple Negative Breast Cancer. *Front Public Health.* 2019;7:18. Published 2019 Feb 18.
- [9] Roseland ME, Pressler ME, Lamerato LE, et al. Racial differences in breast cancer survival in a large urban integrated health system. *Cancer.* 2015;121(20):3668-3675.
- [10] Singh GK, Jemal A. Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950-2014: Over Six Decades of Changing Patterns and Widening Inequalities. *J Environ Public Health.* 2017; 2017:2819372. doi:10.1155/2017/2819372
- [11] Madhav KC, Oral E, Rung AL, et al. Neighborhood deprivation and risk of mortality among men with prostate cancer: Findings from a long-term follow-up study. *Prostate.* 2022;82(7):783-792. doi:10.1002/pros.24320
- [12] Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](http://www.seer.cancer.gov)) SEER\*Stat Database: Incidence - SEER Research Plus Data, 17 Registries, Nov 2021 Sub (2000-2019) - Linked To County Attributes - Total U.S., 1969-2020 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2022, based on the November 2021 submission



- [13] Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control*. 2001 Oct;12(8):703-11. doi: 10.1023/a:1011240019516.
- [14] Ross J, Braswell KV, da Silva LM, et al. Unraveling the etiology of ovarian cancer racial disparity in the deep south: Is it nature or nurture? *Gynecol Oncol*. 2017; 145(2): 329-333.
- [15] Kolak M, Bhatt J, Park YH, Padrón NA, Molefe A. Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. *JAMA Netw Open*. 2020;3(1):e1919928. doi:10.1001/jamanetworkopen.2019.19928
- [16] Krieger, N., Singh, N. & Waterman, P.D. Metrics for monitoring cancer inequities: residential segregation, the Index of Concentration at the Extremes (ICE), and breast cancer estrogen receptor status (USA, 1992–2012). *Cancer Causes Control* **27**, 1139–1151 (2016). <https://doi.org/10.1007/s10552-016-0793-7>
- [17] 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. *Am J Public Health*. 2003;93(10):1655-1671. doi:10.2105/ajph.93.10.1655
- [18] Yuan Y and MacKinnon DP. Bayesian mediation analysis. *Psychol Methods* 2009; 14: 301–322.
- [19] McCandless LC, Gustafson P and Levy AR. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med* 2007; 26: 2331–2347.
- [20] McCandless LC, Somers JM. Bayesian sensitivity analysis for unmeasured confounding in causal mediation analysis. *Statistical Methods in Medical Research*. 2019;28(2):515-531. doi:[10.1177/0962280217729844](https://doi.org/10.1177/0962280217729844)
- [21] Greenland S. Multiple bias modelling for analysis of observational data. *J Roy Stat Soc Ser A* 2005; 168: 267–306.
- [22] VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology* 2011; 22: 582–585.
- [23] Greenland S. Bayesian perspectives for epidemiologic research, part III. *Int J Epidemiol* 2009; 38: 1662-73.
- [24] Dellaportas, P. & Forster, J. & Ntzoufras, I.. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*. 12. 27-36. 10.1023/A:1013164120801.
- [25] Barbieri, M. M & Berger, J. O. (2004) Optimal predictive model selection. *Ann. Statist.* 32 (3) 870 - 897, <https://doi.org/10.1214/009053604000000238>
- [26] de Valpine, P., D. Turek, C.J. Paciorek, C. Anderson-Bergman, D. Temple Lang, and R. Bodik. 2017. Programming with models: writing statistical algorithms for general model



structures with NIMBLE. *Journal of Computational and Graphical Statistics* 26:403-413. <https://doi.org/10.1080/10618600.2016.1172487>

[27] Westrick, AC, Bailey, ZD, Schlumbrecht, M, Hlaing, WM, Kobetz, EE, Feaster, DJ, Balise, RR. Residential segregation and overall survival of women with epithelial ovarian cancer. *Cancer*. 2020; 126: 3698- 3707. <https://doi.org/10.1002/cncr.32989>

[28] Westrick A, Schlumbrecht M, Hlaing WM, Kobetz EK, Feaster D, Balise R. Racial and ethnic disparities in the overall survival of women with epithelial ovarian cancer in Florida, 2001-2015. *Cancer Causes Control*. 2020 Apr;31(4):333-340. doi: 10.1007/s10552-020-01276-2. Epub 2020 Feb 13. PMID: 32052218

[29] Hodeib M, Chang J, Liu F, et al. Socioeconomic status as a predictor of adherence to treatment guidelines for early-stage ovarian cancer. *Gynecol Oncol*. 2015;138(1):121-127. doi:10.1016/j.ygyno.2015.04.01

[30] Brewer KC, Peterson CE, Davis FG, Hoskins K, Pauls H, Joslin CE. The influence of neighborhood socioeconomic status and race on survival from ovarian cancer: a population-based analysis of Cook County, Illinois. *Ann Epidemiol*. 2015 Aug;25(8):556-63.

[31] Menon, U., Gentry-Maharaj, A. et al. (2021) Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial *Lancet*, 397,10290, 2182-2193

[32] Schildkraut, J., Johnson, C., Dempsey, L. et al (2022) Survival of epithelial ovarian cancer in Black women: a society to cell approach in the African American Cancer Epidemiology Study (AACES) (submitted)

Table 1 Selected clinical characteristics of the cohort (N=558). The table count reflects the inclusion only of participants whose geocoding was complete.

Characteristics	Mean (SD)	
Days	1774.50 (962.86)	
Delayed diagnosis (1-12 months)*	6.98(4.34)	
	N	Percentage (%)
Survival Status		
Alive (censored)	221	39.6
Deceased	337	60.4
Histology		
High grade serous	375	67.2
Others	183	32.8
Stage		
1-Localized	125	23.9
2-Regional	51	9.8
3-Distant	346	66.3
Unknown	36	-
Debulking status – after imputation (residual disease)		
1=Optimal debulking (or CA125 after adjuvant <35)	391	70.1
2=Suboptimal debulking (or CA125 after adjuvant ≥35)	167	29.9

\*Delayed diagnosis variable has 64 missing observations

Table 2. Selected characteristics of the Dis , SIs and confounding variables. (N=558)

Characteristics	Mean (SD)	
Age	58.04(10.90)	
BMI	32.82(8.42)	
Yost 2010	9493.66(921.62)	
ICE race	-0.10(0.57)	
ICE income	-0.19(0.27)	
Kolak URB	-0.40(0.85)	
Characteristics	N	Percentage (%)
Self-reported income		
less than \$10,000	113	22.2
\$10,000-\$24,999	120	23.6
\$25,000-\$49,999	125	24.6
\$50,000-\$74,999	76	14.9
\$75,000-\$100,000	44	8.6
More than \$100,000	31	6.1
NA	49	-
Smoking status		
Never smoker	309	55.4
Ever Smoker (Before/current)	249	44.6
PAGA		
1 = Yes	130	25.1
2 = NO	387	74.9
NA	41	-

\*Number of missing observations for the continuous variables

Kolak URB : 1, Yost: 13.

Table 3: Mediation analysis for ovarian cancer survival with the histology as the mediator variable. (Binary histology variable: 1. High grade Serous, 2. Others) . The table reports the odds of increased (decreased) survival as influenced by the DI or SI reported. Well estimated odds shown in bold. The total effect is estimated from the direct effect and indirect effect estimated at each level of histology. In the final model the selected confounders were age alone (URB and Yost), and age and BMI (ICE indices).

	<b>URB</b>	<b>Yost</b>	<b>ICE-income</b>	<b>ICE - race</b>
Direct Effect of each DI (each column)	<b>0.67 (0.51,0.849)</b>	<b>0.752 (0.593,0.946)</b>	<b>0.439 (0.193,0.922)</b>	0.736 (0.509,1.049)
Indirect effect	1.002 (0.635,1.5)	1.014 (0.647,1.791)	0.927 (0, 1.47x10 <sup>14</sup> )	0.988 (0.509,1.755)
Total effect	0.671 (0.397,1.046)	0.762 (0.437,1.415)	0.407 (0,5.43x10 <sup>13</sup> )	0.727 (0.37,1.427)
Average histotype	1.043 (0.61,1.951)	0.991 (0.456,2.056)	0.901 (0.366,1.896)	0.714 (0.184,1.575)
Age	<b>1.05 (1.035,1.068)</b>	<b>1.051 (1.035,1.07)</b>	<b>1.055 (1.035,1.071)</b>	<b>1.043 (1.032,1.055)</b>
BMI	—	—	1.208 (0.938,1.598)	1.211 (0.958,1.562)

Table 4: Mediation analysis for ovarian cancer survival with the delayed diagnosis of the cancer as the mediator variable. The table reports the odds of increased (decreased) survival as influenced by the DI or SI reported Well estimated odds shown in bold. In the final model the selected confounders were age and BMI for all the indices.

	URB	Yost	ICE-Income	ICE - race
Direct Effect of each DI (each column)	<b>0.785</b> ( <b>0.678,0.904</b> )	<b>0.704</b> ( <b>0.542,0.937</b> )	<b>0.385</b> ( <b>0.124,0.901</b> )	0.88 (0.718,1.074)
Indirect effect	1.001 (0.994,1.009)	1 (0.995,1.006)	1.01 (0.972,1.057)	1.007 (0.995,1.022)
Total effect	<b>0.785</b> (0.679,0.907)	<b>0.704</b> (0.542,0.932)	<b>0.389</b> (0.128,0.916)	0.886 (0.725,1.076)
Diagnostic delay	<b>0.900</b> (0.833,0.974)	0.995 (0.873,1.12)	<b>0.829</b> (0.729,0.926)	<b>0.892</b> (0.826,0.961)
Age	<b>1.032</b> (1.023,1.041)	<b>0.929</b> (0.874,0.962)	<b>1.078</b> (1.062,1.094)	<b>1.032</b> (1.023,1.041)
BMI	<b>1.175</b> (1.046,1.315)	0.773 (0.483,1.15)	<b>1.354</b> (1.067,1.717)	<b>1.182</b> (1.052,1.34)

## Appendix

In this Appendix the detail of the Bayesian joint mediation models is provided.

For a single mediator the following describes the model assumed: for the  $i$  th participant:

the survival time is  $T_i$ , and the censoring indicator is  $\delta_i$ . The mediator is defined to be  $M_i$ , the exposure is  $X_i$ , and the confounder set is  $\{C_i\}$ . In addition, we assume that a random effect at the participant level will be included in the models:  $U_i$ . For a survival outcome, we assume a conventional Weibull model, conditional on the mediator, exposure and observed and unobserved confounding, suppressing the individual subscript for convenience:

$$T | X, M, C, U \sim Weib(\exp(\beta_0 + \beta_x X + \beta_M M + \beta_C^T C + \beta_U U), \nu)$$

This choice of model allows for a flexible form of hazard. In this formulation, the term  $\beta_C^T C$  is a linear predictor as a function of the confounder set.

For the mediation model, conditional on exposure and confounders :

$$M | X, C, U \sim f(\alpha_0 + \alpha_x X + \alpha_C^T C + \alpha_U U)$$

where  $f(\cdot)$  is a chosen mediator distribution. Note that the same random effect is shared and scaled in each of the outcome and mediator models.

Finally, the random effect model, which, in general, can depend on exposure and confounders, is

$$U | X, C \sim g(\dots X)$$

where  $g(\cdot)$  is a suitable random effect distribution. This will be assumed to be continuous and zero mean Gaussian unless otherwise stated, and in our study will not be directly dependent on exposure or confounding. The above models are fitted jointly so that the random effect is shared but scaled. Note that the confounders are assumed to have different linear predictors in the outcome and mediator models.

In this model structure, the direct and indirect effects of an exposure is defined as: Direct effect:

$$\approx \beta_x, \text{ Indirect effect: } \approx \alpha_x \beta_M \text{ and Total effect: } \beta_x + \alpha_x \beta_M.$$

### Variable Selection

In the above model, the linear predictor for confounders is assumed to be fixed. However, it is useful to consider an extension whereby we examine a range of putative models

during our estimation procedure. A Bayesian method which allows the sampling of all models within the computational method of MCMC is Gibbs variable selection.<sup>23</sup>

Within the design matrix  $\mathbf{C}$  we can have multiple predictors and may wish to perform variable selection on these. Hence, we could replace  $\beta_c^T \mathbf{C}$  and  $\alpha_c^T \mathbf{C}$  with  $\beta_c^{*T} \mathbf{C}$  where  $\beta_c^* = \beta_c \cdot \lambda_c^S$  and  $\alpha_c^{*T} \mathbf{C}$  where  $\alpha_c^* = \alpha_c \cdot \lambda_c^M$  respectively. The  $\lambda$  parameters are inclusion or entry parameters and allow the predictors to be swapped in and out of the model during estimation. It is assumed that the ‘entry’ parameters have Bernoulli prior distributions:  $\lambda_{cj}^S \sim \text{Bern}(p_{cj}^S)$ ,  $\lambda_{cj}^M \sim \text{Bern}(p_{cj}^M)$ . This allows the inclusion parameters to range between [0,1]. When the estimation process is complete, the inclusion parameters are averaged and any predictor where  $\text{ave}(\lambda) > 0.5$  is selected for a final model.<sup>24</sup> The final model is then fitted with only those predictors reaching this threshold.

### Computation and Prior Distributions

To fit the above model, and all the model variants we discuss here, we have implemented the models using a MCMC procedure in the R package Nimble.<sup>25</sup> In general, we assume regression parameters to have conventional zero mean Gaussian prior distributions with precisions which have  $Ga(a,b)$  prior distributions. We assume weakly informative gamma priors with  $a = 2.0$  and  $b = 0.5$ .

Nimble chooses the most appropriate sampling algorithms to use to generate samples from the posterior distribution.