Bayesian Causal Mediation with Deprivation and Segregation Indices in Ovarian Cancer Survival

Andrew B Lawson (MUSC, UOE)

Yao Xin (MUSC)

Joanne Kim(OHSU)

AACES study consortium

Background

- The African American Cancer Epidemiology Study (AACES) consortium is funded to study a cohort of African American Ovarian cancer (EOC) patients in terms of survival.

- The study has been funded for over 10 years and the current version involves multilevel analyses including genetic and environmental factors

- A primary aim of the study is to examine deprivation and segregation in relation to the survival experience of EOC sufferers

Deprivation and Segregation

- To assess the impact of deprivation and segregation we examined indices that are commonly used to assess contextual neighborhood effects (census tract level)

- DIs: Yost, CDI, ADI, % under poverty line (POV), Kolak: URB, SES, MOB, MICA
  - Kolak M, et al (2020) Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. *JAMA Netw Open.*;3(1):e1919928. doi:10.1001/jamanetworkopen.2019.19928

- SIs: Index of Concentration at the extremes (ICE) measures (income and race)
  - Krieger, N., Singh, N. & Waterman, P.D. (2016) 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. https://doi.org/10.1007/s10552-016-0793-7

Data

- AACES1 includes data on cancer registry patients from a range of 11 states included in the study (Alabama, Georgia, Illinois, Lousiana, South Carolina, North Carolina, Texas, Michigan, New Jersey, Ohio, Tennessee)

- These are spatially separate and recruitment was performed between 2010 and 2015.

- There was a total of 558 participants included in the final analysis

- Exploratory analysis performed on block group and census tract indices : census tract (CT) level chosen for compatibility

Basic Characteristics

Characteristics		Mean (SD)
Days on study		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

DIs/SIs and confounders

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.

** Physical activity guidelines for Americans

Mediation models

- Assume a joint model for survival and mediator (s)
- Survival model ….with Linear predictor $\mathbf{C}$ and exposure X

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

- Mediator model

$$M \mid X, C, U \sim f(\alpha_0 + \alpha_x X + \alpha_C^T \mathbf{C} + \alpha_U U)$$

- Shared random effect

$$U \mid X, \mathbf{C} \sim g(\ldots\ldots X)$$

Prior distributions

- Regression parameters: $\beta_*, \alpha_* \sim N(0, \tau_*^{-1}), \quad \beta_C^{*T}\mathbf{C}$ where $\beta_C^* = \beta_c.\lambda_c^S$

- Precisions: $\tau_* \sim Ga(2, 0.5)$ weakly informative

- Random effects: $U \sim N(0, \tau_U^{-1})$

- Entry parameters: $\lambda_{cj}^S \sim Bern(p_{cj}^S); \; p_{cj}^S \sim beta(1,1)$

- Weibull parameter: $v \sim Ga(0.1, 0.1)$

Computation I

- Joint models result from the formulation described
- These can be implemented using McMC samplers
- Note that we have a two stage process:
  - Stochastic variable selection
    - Identifies the important predictors/confounders
  - Refitting with selected variables.
- Use Gibbs variable selection with entry parameters (Dellaportas et al , 2002)
- Use the threshold of 50 % inclusion (Barbieri and Berger, 2004)

Computation II

- Precision parameters: conditional conjugate samplers
- Regression parameters: zero mean Gaussian prior distributions: random walk samplers
- Random effects: zero mean Gaussian prior distributions: random walk samplers

- Computation set up in R package Nimble
- Single chain runs with Geweke diagnostics
- Use parallel package in R for multi-chain speedups

Direct and indirect effects
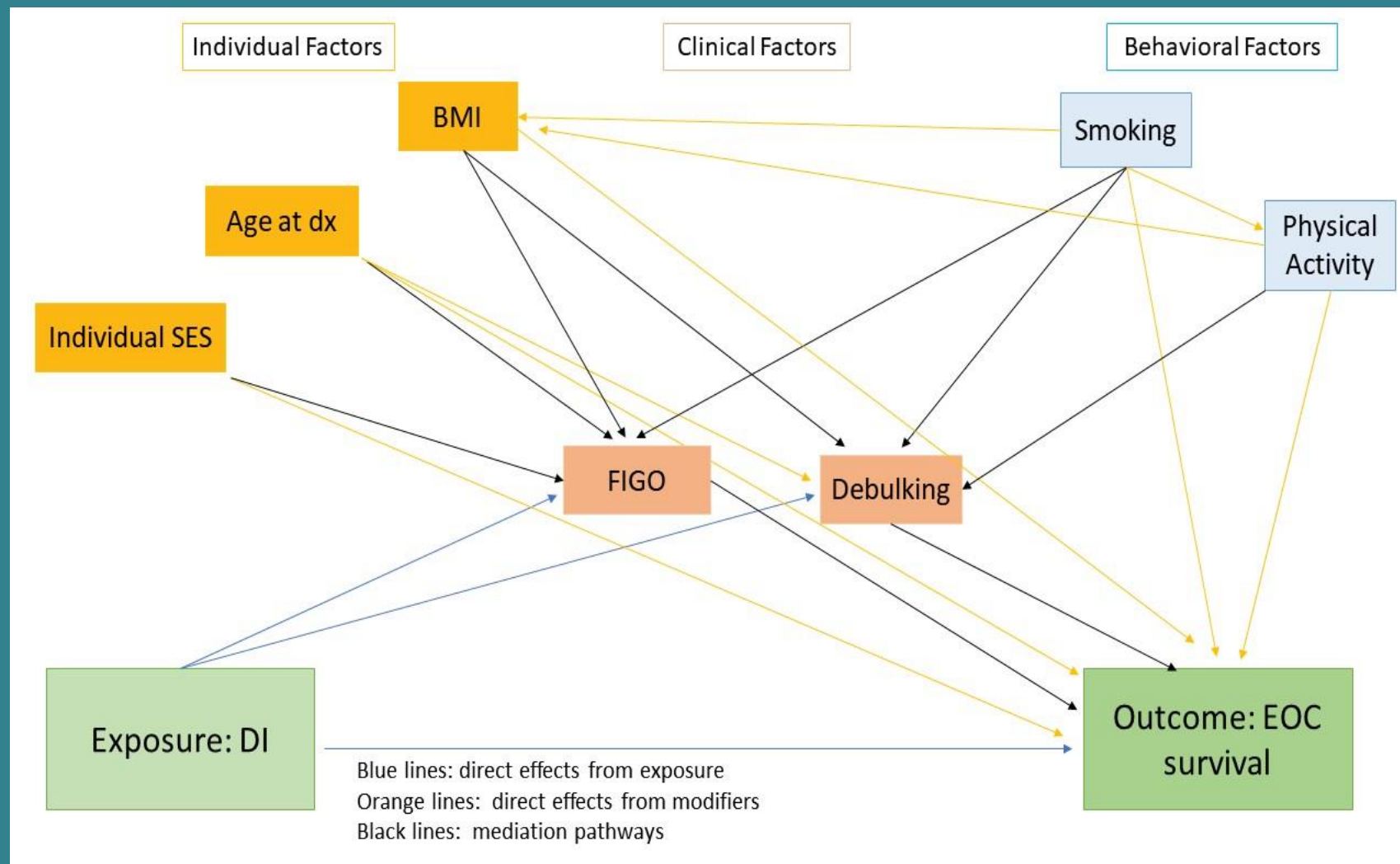
- Following McCandless and Somers (2019)

- We assumed that

- Direct effect $\approx \beta_x$

- Indirect effect $\approx \alpha_X \beta_M$

- Total effect: additive on original scale: $\beta_X + \alpha_X \beta_M$

DAG

Results: URB, Yost, ICE-income, ICE-race by histotype

	URB	Yost	ICE-income	ICE - race
Direct Effect of each DI/SI	0.67 (0.51,0.849)	0.752 (0.593,0.946)	0.439 (0.193,0.922)	0.736 (0.509,1.049)
Indirect effect of histotype[a]	1.002 (0.635,1.5)	1.014 (0.647,1.791)	0.927 (0, 1.47×10^{14})	0.988 (0.509,1.755)
Total effect	0.671 (0.397,1.046)	0.762 (0.437,1.415)	0.407 (0, 5.43×10^{13})	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	1.05 (1.035,1.068)	1.051 (1.035,1.07)	1.055 (1.035,1.071)	1.043 (1.032,1.055)
BMI	—	—	1.208 (0.938,1.598)	1.211 (0.958,1.562)

[a] Histotype categorized as high-grade serous vs other

Results: by diagnostic delay

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

Kolak indices and ADI with stage mediator

Gibbs Variable Selection for Covariates (Modifiers). Covariates with Values Higher Than 0.5 Will Be Included in the Analysis					
Mediator: Stage	URB	MOB	SES	MICA	ADI
Age	1	0.458	0.002	0.912	1
BMI	0.457	0.475	0.144	0.499	0.495
Self-reported SES	0.295	0.509	0.058	0.488	0.483
Smoking	0.428	0.491	0.033	0.506	0.489
Physical activity	0.530	0.516	1	0.504	0.511
Analysis result with selected modifiers. Each cell shows a posterior mean and its 95% credible interval.					
Direct effect (Deprivation indices)	0.731 (0.560, 0.914)	0.887 (0.78, 0.999)	0.895 (0.808, 0.991)	1.149 (0.916, 1.437)	1.873 (0.665, 5.456)
Indirect effect Through stage 1	$0.674 \ (0, 4.5 \times 10^{12})$	0.983 (0.221, 3.977)	0.996 (0.188, 4.571)	1.022 (0.258, 4.088)	0.979 (0.141, 6.51)
Indirect effect Through stage 2	$0.683 \ (0, 8.5 \times 10^{11})$	0.995 (0.493, 1.874)	1.012 (0.508, 2.013)	0.987 (0.381, 2.476)	0.980 (0.352, 2.595)
Indirect effect Through stage 3	$0.711 \ (0, 1.06 \times 10^{11})$	0.998 (0.120, 6.812)	1.028 (0.086, 18.340)	0.902 (0.037, 12.862)	1.024 (0.125, 8.055)
Total effect 1 (Direct + Indirect through stage 1)	$0.492 \ (0, 3.1 \times 10^{12})$	0.872 (0.198, 3.572)	0.891 (0.169, 4.112)	1.175 (0.292, 4.808)	1.833 (0.214, 16.031)
Total effect 2 (Direct + Indirect through stage 2)	$0.499 \ (0, 7.0 \times 10^{11})$	0.882 (0.438, 1.640)	0.906 (0.455, 1.783)	1.134 (0.443, 2.879)	1.835 (0.441, 7.707)
Total effect 3 (Direct + Indirect through stage 3)	$0.519 \ (0, 6.82 \times 10^{10})$	0.885 (0.103, 6.069)	0.92 (0.079, 16.689)	1.036 (0.046, 14.348)	1.919 (0.186, 20.697)

Yost, CDI and ADI mediated by stage

Gibbs Variable Selection for Covariates (modifiers). Covariates with Values Higher Than 0.5 Will Be Included in the Analysis			
Mediator: Stage	YOST	CDI	POV
Age	1	0.001	0.972
BMI	0.497	0.073	0.480
Self-reported SES	0.449	0.032	0.397
Smoking	0.485	1	0.491
Physical activity	0.507	0.029	0.511
Analysis result with selected modifiers. Each cell shows a posterior mean and its 95% credible interval			
Direct effect (Deprivation indices)	0.736 (0.566, 0.954)	1.036 (0.998, 1.075)	1.823 (0.777, 5.187)
Indirect effect Through stage 1	0.971 (0.073, 10.406)	0.971 (0.135, 7.717)	0.99 (0.242, 4.631)
Indirect effect Through stage 2	0.980 (0.292, 3.002)	0.989 (0.292, 3.369)	1.016 (0.413, 2.398)
Indirect effect Through stage 3	1.011 (0.101, 12.289)	1.01 (0.303, 3.504)	1.071 (0.124, 8.947)
Total effect 1 (Direct + Indirect through stage 1)	0.715 (0.050, 7.672)	1.006 (0.141, 7.987)	1.806 (0.345, 11.422)
Total effect 2 (Direct + Indirect through stage 2)	0.721 (0.211, 2.186)	1.025 (0.305, 3.474)	1.852 (0.598, 6.724)
Total effect 3 (Direct + Indirect through stage 3)	0.745 (0.073, 7.829)	1.047 (0.312, 3.616)	1.953 (0.154, 20.444)

Multiple Mediation

- Alternative approach whereby more than one mediator is considered.

- Question: can stage and histology mediate the effects of DIs and/or SIs on survival?

- Extension of the McCandless and Somers approach to multiple mediators whereby multiple joint models assumed

Multiple mediation models

- Survival model

$$T \mid X, M, \mathbf{C}, U \sim Weib(\exp(\beta_0 + \beta_x X + \beta_{M1}M_1 + \beta_{M2}M_2 + \beta_C^T \mathbf{C} + \beta_U U), \nu)$$

- Mediator models

$$M_k \mid X, C, U \sim f(\alpha_{0k} + \alpha_{xk}X + \alpha_{Ck}^T \mathbf{C} + \alpha_{Uk}U)$$
$$k = 1 \; or \; 2$$

- Direct effect β_x

- Indirect effect $\alpha_{x1}\beta_{M1} + \alpha_{x2}\beta_{M2}$

Stage and histotype mediation example

Gibbs Variable Selection for covariates. Covariates with values higher than 0.5 will be included.				
Exposure	**Yost**	**URB**	**ICE-income**	**ADI**
Age	**0.704**	**0.904**	**0.736**	**0.740**
Adjusted income	0.076	0.056	0.076	0.071
Smoking	0.159	0.197	0.133	0.166
Physical Activity	0.245	0.314	0.250	0.243
Diabetes	**0.779**	**0.794**	**0.791**	**0.795**
Hypertension	0.155	0.121	0.140	0.167

Analyses result with selected confounders. Posterior hazard ratio estimates with 95% credible interval.					
Exposure		**Yost**	**URB**	**ICE-income**	**ADI**
Direct effect		0.909(0.799-1.027)	**0.843(0.739-0.960)**	0.896(0.794-1.009)	1.103(0.983-1.2
Indirect effect through stage		1.004(0.320-3.221)	1.002(0.303-3.096)	0.977(0.313-2.716)	0.994(0.316-2.8
Indirect effect through histology		0.994(0.569-1.728)	0.999(0.574-1.666)	1.003(0.637-1.623)	0.999(0.546-1.9
Stage	Distant vs. localized	**5.122(3.423-8.049)**	**5.122(3.386-7.937)**	**5.186(3.451-7.945)**	**5.212(3.376-8.2**
	Regional vs. localized	1.732(0.968-3.055)	1.716(0.966-2.982)	**1.760(1.013-3.080)**	1.740(0.950-3.0
Histology	High grade serous vs. other	1.041(0.767-1.403)	1.019(0.747-1.379)	1.036(0.768-1.412)	1.042(0.757-1.4
Age at diagnosis (years)		**1.020(1.008-1.032)**	**1.022(1.009-1.035)**	**1.020(1.008-1.033)**	**1.020(1.008-1.0**
Diabetes	Have been diagnosed vs. never	**1.392(1.045-1.847)**	**1.428(1.060-1.900)**	**1.389(1.038-1.859)**	**1.405(1.044-1.8**

Stratification and population adjustment

- To evaluate the sensitivity to the effect of population imbalance we have examined the addition of inverse probability weights (IPWs) in the analyses

- Their inclusion did not impact the overall single mediator conclusions .

- We also considered IPWs in the multiple mediation and they also did not show an impact on the major conclusions

Conclusions: single mediation

- DIs: direct effect of Dis mostly significant in that SES-based indices related to improved survival

- SIs: ICE-income often significant but ICE-race not so

- Kolak's URB index and the Yost index show the most marked positive impact on survival

- Poverty is not significantly associated

- Age is often significantly associated with worsened survival, but other predictors (BMI, Physical activity, smoking, individual SES) less often.

- Stage has a significant effect in worsening survival, whereas histology does not show clear significance

Conclusions: multiple mediation

- Most SES-based Dis and SIs (Yost, ADI, ICE-income) are not significant when multiple clinical mediators are present.

- ICE-income is no longer significant BUT

- Kolak's URB index remains as a significant measure which yields improved survival outcome.

- Indirect effects of stage and histology are not significant

- Stage , age and diabetes are all significant as confounders in worsening survival.

Caveats

- We looked at single and multiple mediators only
- Also only looked at CT level and so these relations may not hold at BG or B level.
- This work didn't explore effect modification or other interactions
- Future work will explore computational efficiencies

References

- Schildkraut JM, Johnson C,…., Lawson AB,….. (2023) Survival of epithelial ovarian cancer in Black women: a society to cell approach in the African American cancer epidemiology study (AACES). *Cancer Causes Control.* Mar;34(3):251-265.

- Lawson, A. B. , Kim, J., Johnson, C., Hastert, T., Bandera, E. V., Alberg, A., Terry, P., Akonde, M., Mandle, H., Cote, M., Bondy, M., Marks, J., Peres, L., Ratnapradipa, K. L., Schildkraut, J., Peters, E. (2023) Deprivation and Segregation in Ovarian cancer survival among African American Women: a mediation analysis *Annals of Epidemiology* *https://doi.org/10.1016/j.annepidem.2023.07.001* .

- Lawson, A.B.; Kim, J.; Johnson, C.; Ratnapradipa, K.L.; Alberg, A.J.; Akonde, M.; Hastert, T.; Bandera, E.V.; Terry, P.; Mandle, H.; et al. (2023) The Association between Mediated Deprivation and Ovarian Cancer Survival among African American Women. *Cancers*, *15*, 4848. https://doi.org/10.3390/cancers15194848

References

- **Lawson, A. B.;** Xin, Y.,; Peters, E. S.; Johnson, C.; Hastert, T.; Bandera, E. V.; Alberg, A. J.; Collin, L.; Terry,P.; Akonde, M.; Mandle, H.; Cote, M.; Bondy, M.; Marks, J.; Peres, L.; Ratnapradipa, K. L.; Schildkraut, J. M. (2024) The Role of Multiple Mediation with contextual Neighborhood measures in Ovarian Cancer Survival. (submitted)

- 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

Thank you!