Disparities of Invasive Cervical Cancer Incidence and Related Factors in Ohio: An Integrated Approach

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jesse John Plascak, M.P.H.
Graduate Program in Public Health

The Ohio State University
2013

Dissertation Committee:
Electra D. Paskett, Ph.D., Advisor
Catherine A. Calder, Ph.D.
Amy K. Ferketich, Ph.D.
Ningchuan Xiao, Ph.D.
Copyright by
Jesse John Plascak
2013
Abstract

Despite steady reductions overall, disparities of invasive cervical cancer incidence (ICCI) exist across socioeconomic and geographic gradients. ICCI within the Central Appalachian region of the United States has remained higher than adjacent non-Appalachian regions for decades. Several social, economic, political, and ecologic factors are believed to have contributed to the lower socioeconomic conditions of Central Appalachia. The goals of this research were to theoretically integrate current knowledge of the causes of socioeconomic disparities with the causes of ICCI disparities within Central Appalachia, to describe the multilevel and spatial characteristic of ICCI in Ohio, and to investigate relationships between sociopolitical, economic, and biobehavioral factors together with the prevalence of cervical abnormalities.

Ecosocial theory was used to develop a more comprehensive understanding of the relationships affecting disparities of ICCI within Central Appalachia. The theory posits that population patterns of health are consequences of historically interacting, social, economic and political processes. Literature searches of various databases, directories, and websites was conducted to integrate the current understanding of the processes contributing to both socioeconomic and ICCI disparities within Central Appalachia. Individual-level demographic and ICCI data from the Ohio Cancer Incidence Surveillance System were used to describe the multilevel and spatial characteristics of ICCI in Ohio. Controls were sampled from a synthetic dataset geographically
representative of the Ohio female population. Area-level data were gathered for census tract- school district- and county-levels from various Federal and State sources. Hierarchical logistic regression models were used to investigate and explain variation of ICCI at individual- and area-levels. Spatial data analysis techniques described the spatial variation of ICCI at each level and quantified the degree and location of clustering.

Individual-level data from the Community, Awareness, Resources and Education (CARE) Project were used to investigate the relationships between sociopolitical, economic, and biobehavioral factors together with the prevalence of cervical abnormalities. The goal of CARE project 3 was to investigate the social, behavioral, and biologic variables that may contribute to increased risk of abnormal cervical cytology among women of Appalachia Ohio. Area-level data were gathered for census tract- and school district-levels from various Federal and State sources. Structural equation modeling was used to investigate relationships between the latent variables taxable land value per a pupil, area-level deprivation, individual-level socioeconomic position (SEP), age, tobacco use, risky sexual behavior, human papillomavirus (HPV) infection and persistence, and cervical abnormalities. Mediating effects were also investigated.

Application of Ecosocial theory suggests that relationships may exist between historically dependent sociopolitical (i.e., land-use/land-tax policy, public education funding system), environmental (i.e., rurality), built environment (i.e., physical activity facilities and resources), socioeconomic factors (i.e., community resources, individual-level SEP), and ICCI or ICCI-related factors (i.e., HPV infection and persistence, biologic stress, tobacco use, Pap smear utilization) in Central Appalachia.
Several statistically significant findings of the multilevel and spatial description of ICCI in Ohio are of note. Individual-level age and race interacted such that African Americans (compared to Whites) younger than 39 years have reduced odds of ICCI, while African Americans older than 53 have an increased odds. County-level teen birth rate and percent without health insurance, and school district-level childhood poverty were associated with an increased odds of ICCI, while county-level Chlamydia rate, school district-level taxable land value per a pupil, and census tract-level SEP were associated with a decreased odds of ICCI. Model covariates accounted for area-level global spatial clustering of county- and census tract-level models. The spatial distribution of the school district-level source of variation was considerably changed with adjustment for model covariates. The percent of ICCI variation possibly due to county-, school district-, or census tract-level factors was 1.6%, 3.2%, and 4.3%, respectively.

The final structural equation model contained relationships between individual-level SEP, age, tobacco use, risky sexual behavior, HPV infection and persistence, and cervical abnormalities. Increased risky sexual behavior and HPV infection and persistence were inversely associated with increased cervical abnormalities (P < 0.05). Risky sexual behavior and tobacco use were inversely associated with age (P < 0.05). There was marginal evidence for the inverse relationships between individual-level SEP and risky sexual behavior and HPV infection and persistence (both P = 0.07). Tobacco use decreased with increased individual-level SEP (P < 0.01).

The application of Ecosocial theory leads to the conclusion that explanations of factors contributing to the development or persistence of ICCI disparities that do not include explanations of socioeconomic disparities may be incomplete. Moreover,
emphasizing the possibility that present-day factors and processes may have future consequences for ICCI disparities leads to the conclusion that the prevention of ICCI disparities may be addressed, in part, by addressing socioeconomic disparities.

Results of the multilevel and spatial description of ICCI suggests that a small percentage of the total ICCI variation in Ohio is possibly due to county-, school district-, or census tract-level factors. However, taxable land value per a pupil, a present-day manifestation of historical sociopolitical processes, was inversely associated with ICCI. These results may be used by public health practitioners to target interventions, by cancer researchers to motivate future studies, or by those wishing to address disparities through policy action.

Results of the structural equation modeling suggest that age and individual-level SEP were related to cervical abnormalities and several ICCI-related factors within a sample of Appalachian women that may be at an increased risk of ICCI, including: older age was positively associated with individual-level SEP, which, in turn, was significantly associated with cigarette use; individual-level SEP may have mediated the relationship between age and HPV infection and persistence; and individual-level SEP may be associated with risky sexual behavior and HPV infection and persistence. If true, efforts to reduce ICCI disparities in Appalachia Ohio guided by these results may benefit by targeting younger women of lower SEP for risk reduction behavior and HPV vaccine uptake.

This approach taken to investigate disparities of ICCI in Central Appalachia was broad and integrative in scope and methodology. Advantages of this approach include: 1) a deeper understanding of factors and processes that may contribute to the cause and
persistence of ICCI disparities in Central Appalachia, 2) empirical evidence of a previously unconsidered association with ICCI disparities that has theoretical connections to historically relevant sociopolitical processes, and 3) detailed descriptions of the distribution of ICCI, and factors associated with ICCI, measured at multiple geographic levels. While the possible compromise of discipline-specific depth for the sake of multidisciplinary breadth is an important disadvantage of this integrative approach, the insight gained and future avenues of research identified may offset this disadvantage. Persistent ICCI disparities in Central Appalachia may be the result of complex relationships between historically-dependent social, economic, political, cultural, and environmental processes. Interventions may need to consider several of these processes in order to successfully eliminate ICCI disparities in Central Appalachia.
To Megan, Jacob and Leah.
Acknowledgments

This work would not have been possible without the guidance and support of several individuals. Foremost, I must thank Doctors Electra Paskett, Catherine Calder, Amy Ferketich, and Ningchuan Xiao who graciously agreed to be a member of my dissertation committee. Their mentorship, guidance, support, and influence have shaped numerous facets of this research.

I am also grateful for the time and patience that several have given. Namely, Erinn Hade’s and Cecelia Degraffenreid’s detailed explanations of the data, Georgette Haydu’s explanations of data quality, and Dr. Mack Ruffin’s insights on some of the research questions addressed in this work. I am especially grateful for the enumerable conversations with Dr. Jay Fisher that have positively shaped much of this research. Each of these individuals has helped to get this work off the ground and proceed smoothly.

I am forever indebted to my family for the values that they instilled, support and love that they provided, and confidence that they maintained. This undertaking would not have occurred without the constant examples of diligence learned from my parents. The support of my sisters has eased many of the challenging times. Lastly, I have been fortunate to have the support, love, and confidence of my wife, Megan, and children, Jacob and Leah. Your understanding and support has enabled and continually motivated me.
Data used in this research were from a study funded by the National Cancer Institute (CA105632).

Cancer incidence data used in these analyses were obtained from the Ohio Cancer Incidence Surveillance System (OCISS), Ohio Department of Health (ODH), a cancer registry partially supported by the National Program of Cancer Registries at the Centers for Disease Control and Prevention (CDC) through Cooperative Agreement Number 1U58DP003936-01. Use of these data does not imply that ODH or CDC agrees or disagrees with the analyses, interpretations or conclusions in this report (or publication/presentation).
Vita

Education

2009..........................................................M.P.H. Epidemiology, The Ohio State University

2006..........................................................B.S. Microbiology, The Ohio State University

Professional Experience

2011-Present ..............................................Research Analyst, Comprehensive Cancer Center, The Ohio State University

2008-2011 .....................................................Graduate Research Associate, Comprehensive Cancer Center, The Ohio State University

Peer-Reviewed Publications


Peer-Reviewed Publications (Cont’d)


Peer-Reviewed Manuscripts Under Review


2. Plascak JJ, Fisher JL, Paskett ED. Availability of Primary Care Physicians, Type of health insurance and disparities in late-stage cancer diagnosis. *JHCPU.* (First Submission)

Fields of Study

Major Field: Public Health, Epidemiology

Minor Field: Geospatial Data and Analysis
Table of Contents

Abstract ........................................................................................................................................... ii
To Megan, Jacob and Leah. ............................................................................................................... vii
Acknowledgments ........................................................................................................................... viii
Vita ................................................................................................................................................ x
List of Tables ................................................................................................................................ xvi
List of Figures ................................................................................................................................ xvii

Chapters

Chapter 1: Introduction .................................................................................................................. 1
Chapter 2: Background .................................................................................................................. 6
  Appalachia .................................................................................................................................... 6
  Disparities of Appalachia .............................................................................................................. 8
    Economic, Social, and Demographic ......................................................................................... 8
    Health ....................................................................................................................................... 12
  Ohio Appalachia .......................................................................................................................... 15
  Disparities of Ohio Appalachia ....................................................................................................... 15
Cervical Cancer .............................................................................................................................. 38
  Factors Related to Cervical Cancer Incidence .......................................................................... 40
    Human Papillomavirus .............................................................................................................. 40
    Human Papillomavirus-Cofactors ............................................................................................. 49
  Pap Smear Test ............................................................................................................................ 61
Socioenvironmental Influences of Health ...................................................................................... 63
  Individual-level, Absolute Socioenvironmental Measures ......................................................... 65
  Individual-level, Relative Socioenvironmental Measures ........................................................... 67
| Group-level Socioenvironmental Measures                                                                 | 69 |
| Social Epidemiologic Theoretical Frameworks                                                                | 73 |
| Social Determinants of Health                                                                              | 74 |
| The Centers for Population Health and Health Disparities Multilevel Model of Population Health and Health Disparities | 75 |
| Ecosocial Theory                                                                                            | 76 |
| Chapter 3: Methods                                                                                         | 80 |
| Datasets                                                                                                   | 80 |
| Community Awareness Resources and Education Study                                                         | 80 |
| Study Design and Data Collection                                                                           | 81 |
| Geocoding Participants                                                                                     | 85 |
| Ohio Cancer Incidence and Surveillance System                                                              | 87 |
| Federal and State Databases                                                                               | 90 |
| A Synthesized U.S.Geospatial Database                                                                      | 90 |
| Area-level Covariates                                                                                      | 90 |
| Data Use and Institutional Review Board Approval                                                          | 93 |
| Data Analyses                                                                                              | 93 |
| Research Aim 1: An *Ecosocial* Perspective of Disparities of Invasive Cervical Cancer Incidence in Central Appalachia | 93 |
| Research Aims 2 & 3: Multilevel and Spatial Characteristics of Invasive Cervical Cancer Incidence in Ohio   | 94 |
| Study Sample                                                                                               | 94 |
| Statistical Modeling                                                                                        | 100 |
| Spatial Analysis                                                                                           | 104 |
| Software                                                                                                   | 107 |
| Research Aim 4: Pathways from Socioeconomic and Biobehavioral Factors to Cervical Abnormalities Among Women in Appalachia Ohio | 108 |
| Study Sample                                                                                               | 108 |
| Latent Variables                                                                                            | 113 |
| Statistical Modeling                                                                                        | 115 |
| Software                                                                                                   | 117 |
Summary .................................................................................................................................................. 117

Chapter 4: An *Ecosocial* Perspective of Invasive Cervical Cancer Incidence in Central
Appalachia .................................................................................................................................................. 119

Abstract .................................................................................................................................................. 119

Introduction ............................................................................................................................................ 120

Biomedical Approach to Invasive Cervical Cancer Incidence Risk and Disparity
Research ................................................................................................................................................... 123

*An Ecosocial* Perspective of Population Health Disparities ................................................................. 124

Central Appalachian Cervical Cancer Disparities from an *Ecosocial* Perspective............................. 127


Reconceptualization of Invasive Cervical Cancer Incidence: A Biologic Consequence of Socioecologic Disparities ................................................................................................................ 130

Interdependent Pathways Connecting Invasive Cervical Cancer Incidence to Ecologic, Social, Economic and Political Historical Conditions within Central Appalachia .................................................................................................................. 131

Discussion and Conclusions .................................................................................................................. 141

Chapter 5: Multilevel and Spatial Characteristics of Invasive Cervical Cancer Incidence in Ohio .................................................................................................................................................. 145

Abstract .................................................................................................................................................. 145

Introduction ............................................................................................................................................ 146

Methods .................................................................................................................................................. 149

Study Sample ......................................................................................................................................... 149

Study Design ......................................................................................................................................... 149

Measures ............................................................................................................................................... 150

Statistical Analysis ................................................................................................................................. 153

Variable Classification .......................................................................................................................... 153

Statistical Modeling ............................................................................................................................... 154

Spatial Analysis .................................................................................................................................... 157

Results .................................................................................................................................................... 159

Principal Component Analysis ............................................................................................................ 159

Multilevel Modeling .............................................................................................................................. 160
Chapter 6: Pathways Leading from Socioeconomic and Biobehavioral Factors to Cervical Abnormalities Among Women in Appalachia Ohio

Abstract .................................................................................................................. 208
Introduction .............................................................................................................. 209
Methods ..................................................................................................................... 211
Study Sample .......................................................................................................... 211
Measured Variables Considered for Structural Equation Modeling ................. 213
Latent Variables Considered for Structural Equation Modeling ......................... 216
Statistical Analysis ................................................................................................. 217
   Structural Equation Modeling ............................................................................... 217
Results ..................................................................................................................... 219
Sample Description ............................................................................................... 219
Model Development ............................................................................................... 221
Final Model .............................................................................................................. 222
Model Fit .................................................................................................................. 222
Direct Effects ......................................................................................................... 222
Indirect Effects ....................................................................................................... 223
Discussion ............................................................................................................... 225
Findings .................................................................................................................... 225
Appendices

Appendix A: Schematic Used to Geocode Participants of the Community Awareness Resources and Education Project 3 Study ................................................................. 237

Appendix B: A summary of the Variables Considered in any Modeling Investigating Area-level Effects ................................................................. 242

Appendix C: Generic Notational Representation of the Multilevel Logistic Regression Equation ................................................................. 248

Appendix D: Model Estimates Resulting from a Sensitivity Analysis of the Multilevel and Spatial Investigation of Invasive Cervical Cancer Incidence in Ohio that Restricted to the Appalachian Region ................................................................. 250

Appendix E: Maps Resulting from Neighborhood Definition Sensitivity Analysis of the Multilevel and Spatial Investigation of Invasive Cervical Cancer Incidence in Ohio.... 257

Bibliography .................................................................................................................................................................................. 264
List of Tables

Table 3.1. Latent variables and associated measured variables in a structural equation model of cervical cancer-related factors ................................................................. 114

Table 5.1. Summary results of the principal components analysis of socioeconomic-related census tract-level variables ................................................................. 160

Table 5.2. Effects of demographic and county-level factors on invasive cervical cancer incidence in Ohio, 1996-2009 ........................................................................ 162

Table 5.3. Effects of demographic and school district-level factors on invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................. 168

Table 5.4. Effects of demographic and census tract-level factors on invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................. 170

Table 5.5. Summary of spatial clustering of area-level random intercepts produced from multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................. 183

Table 6.1. Characteristics of participants in the CARE study with complete data on covariates of the hypothetical model ................................................................. 220
List of Figures

Figure 2.1. The U.S. Appalachian region, 2013 ................................................................. 7

Figure 2.2. Percent of uninsured Ohio Adults, ages 18 to 64 years, 2008......................... 22

Figure 2.3. A-C. Self-report health status by Ohio counties, OFHS 2008: A) percent with good, very good, or excellent health; B) percent with less than 14 physically unhealthy days in the previous 30 days; C) percent with less than 20 mentally unhealthy days in the previous 30 days ................................................................. 25

Figure 2.4. A-G Regional trends of Ohio cancer incidence rates per 100,000, 1996-2006: A) female breast; B) cervix; C) colon and rectum; D) lung and bronchus; E) melanoma of the skin; F) oral cavity and pharynx; and G) prostate ........................................ 31

Figure 2.5. Age group-specific invasive cervical cancer incidence rates, U.S., 2003-2007 ........................................................................................................................................ 41

Figure 2.6. Typical temporal pattern of human papillomavirus persistence, clearance and neoplastic progression .................................................................................................................. 44

Figure 2.7. Age group-specific human papillomavirus prevalence among women with normal cervical cytology .................................................................................................................. 45

Figure 2.8. Human papillomavirus types 16 and 18 attributable fractions of various cervical abnormalities and invasive cervical cancer ........................................................................ 47

Figure 3.1. Regions, counties and clinics of the Community Awareness Resources and Education Project 3 study ........................................................................................................... 82

Figure 3.2. A graphical representation of the Bethesda cervical abnormalities nomenclature, other nomenclatures, and risk of cervical cancer incidence .......................................... 86

Figure 3.3. Ohio average annual, age- and race-specific invasive cervical cancer incidence rates, 1996-2009 ........................................................................................................................... 99

Figure 3.4. Hypothesized model with measured variables, latent variables and directed relationships between measured and latent variables ................................................................. 116
Figure 4.1. Age-adjusted cervical cancer incidence rates by Kentucky, Ohio and West Virginia counties, 1996-2008 ................................................................. 129

Figure 4.2. Estimated distribution of socioeconomic deprivation within Kentucky, Ohio and West Virginia counties ................................................................. 129

Figure 4.3. Estimated distribution of healthcare access within Kentucky, Ohio and West Virginia counties ................................................................. 130

Figure 4.4. A detailed representation of the relationships involved in an Ecosocial explanation of disparities of invasive cervical cancer incidence in Central Appalachian ................................................................. 132

Figure 5.1. Spatially-varying probability of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................................. 174

Figure 5.2. Spatially-varying probabilities of Pearson-type residuals from a null, single-level logistic regression model of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................................. 175

Figure 5.3. A & B. Spatially-varying probability of Pearson-type residuals from county null and final multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................................. 177

Figure 5.4. A & B. Spatially-varying probability of Pearson-type residuals from school district null and final multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................................. 178

Figure 5.5. A & B. Spatially-varying probability of Pearson-type residuals from census tract null and final multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................................. 179

Figure 5.6. Results of Ripley’s K-function of invasive cervical cancer incidence case-control data in Ohio, 1996-2009 ................................................................................. 181

Figure 5.7. Perspective density plot of the observed probability of invasive cervical cancer incidence in Ohio, 1996-2009 ................................................................................. 182

Figure 5.8. County-level random intercepts of the invasive cervical cancer incidence null model ................................................................................. 184

Figure 5.9. County-level random intercepts of the invasive cervical cancer incidence final model ................................................................................. 185
Figure 5.10. School district-level random intercepts of the invasive cervical cancer incidence null model ........................................................................................................... 187

Figure 5.11. School district-level random intercepts of the invasive cervical cancer incidence final model ........................................................................................................... 189

Figure 5.12. Census tract-level random intercepts of the invasive cervical cancer incidence null model ........................................................................................................... 190

Figure 5.13. Census tract-level random intercepts of the invasive cervical cancer incidence final model ........................................................................................................... 191

Figure 6.1. Final structural equation model with standardized coefficients of directed paths estimated between measured variables and latent variables among 800 participants in the CARE study ........................................................................................................... 223
Cervical cancer is a significant global, regional, and local public health problem. It is the third most commonly diagnosed cancer and fourth leading cause of cancer death among females worldwide. The burden of cervical cancer is more pronounced in economically underdeveloped regions with age-adjusted incidence and mortality rates of 17.8 and 9.8 per 100,000 females, ranking second among all cancer sites. In the U.S. it is estimated that 12,170 new invasive cervical cancer cases will be diagnosed and 4,220 females will die of cervical cancer in 2012. The 2003-2007 age-adjusted, U.S. cervical cancer incidence and mortality rates were 8.1 (ranking 13th) and 2.4 (ranking 14th) per 100,000 females.

Incidence and mortality rates of cervical cancer in the U.S. have decreased tremendously since the introduction and uptake of the Papanicolaou (Pap) test in the mid twentieth century. Universal distribution of the highly efficacious human papillomavirus (HPV) vaccine allows the possibility for cervical cancer rates to be reduced by another 70%. However, the overall gains in cervical cancer prevention should not obscure the presence of demographic, geographic, historical, and social disparities. African Americans compared to whites, Hispanics compared to non-hispanics, and those of lower socioeconomic position (SEP) have elevated risk of cervical cancer. Certain regions of the U.S. – Central Appalachia, along the Mississippi River in the South, Southern Texas, and the Eastern mid-Atlantic – have reported
persistently elevated cervical cancer rates for decades (only mortality rates were reported prior to the 1990s)\textsuperscript{11,12}.

Central Appalachia has been the subject of much social, political, and environmental attention over the past century. The region was known to have lagging economic and social indicators since the late nineteenth and early twentieth centuries\textsuperscript{13-17}. Parts of Franklin D. Roosevelt’s New Deal (c. 1933) initiatives were directed towards Appalachia\textsuperscript{17(p. 278, 312, 316)}. Media attention reintroduced non-Appalachian American society to a caricature of actual Appalachian culture and deprivation in between the policies of the New Deal and Lyndon B. Johnson’s 1965 “War on Poverty”\textsuperscript{13,14 (p. 29-30)}. The War on Poverty had political roots in John F. Kennedy’s 1960 pursuit of the democratic presidential nominee\textsuperscript{17(p. 339)}. Large-scale environmental catastrophes have been frequent occurrences to Central Appalachia\textsuperscript{18-21}. This attention from varied sectors of society has prompted and been paralleled by a multitude of Appalachia-focused academic work \textsuperscript{14-18,22-25}. A theme unifying much of this literature is that Appalachian poverty – and the social, political and ecologic processes contributing to poverty – are fundamental to understanding the regions past and present conditions.

The rich literature investigating Central Appalachia’s historical poverty along with the extensive biomedical research conducted on cervical cancer etiology and natural history creates unique opportunities for cervical cancer disparity research. Conceptualizing invasive cervical cancer incidence (ICCI) as a biologic consequence of the social, economic, and political milieu acting throughout history within Central Appalachia may provide insight into the conditions causing this persistent cancer disparity. In response to a lack of previous research on ICCI disparities employing
theories of disease distribution incorporating multiple levels across spatial scales and various societal contexts, the following research aims and hypotheses were formulated.

**Research aim #1:** To apply the *Ecosocial* theory and integrate current understanding of ICCI risks and pathways with extant theories of Central Appalachian underdevelopment and poverty so as to inform the social, political, economic and cultural processes contributing to persistent ICCI disparities.

**Hypothesis #1:** Relationships between commonly identified and less-considered factors and processes will be realized, resulting in deeper understanding of the causes of persistent ICCI disparities in Central Appalachia. Applying *Ecosocial* theory will provide explanation for *why* and *how* these relationships exist and *who* and *what* is responsible for the causes of persistent disparities of ICCI.

**Research aim #2:** To investigate and explain the variation of ICCI in Ohio by partitioning it into individual- and various area-level sources (i.e., county, school district, and census tract) according to the theoretical framework of aim one.

**Hypothesis #2:** Several well-established factors (e.g., age, race, area-level measures of SEP, healthcare access, etc) as well as less well-established factors (e.g., built environment and sociopolitical factors) will be associated with ICCI at various area-levels.

**Research aim #3:** Using the results of aim two, describe and summarize the spatial distribution of ICCI within the individual-, county-, school district-, and census tract-levels.
Hypothesis #3: The spatial distribution of ICCI in Ohio will change after adjusting for various factors. Adjustment of socioeconomic factors will reduce geographic disparities of ICCI.

Research aim #4: To further test the theoretical framework developed in aim one and tested in aim two by investigating the relationships between cervical abnormalities (a necessary precursor of ICCI), biologic and behavioral cervical cancer risk factors, and socioeconomic and sociopolitical risk factors of both cervical cancer and cervical cancer-related factors among a sample of women residing in Appalachia Ohio.

Hypothesis #4: Decreased individual-level SEP (i.e., income and education) and area-level SEP (e.g., percent impoverished, median household income) will be related to increased severity of cervical abnormalities, but only through tobacco use. Increased tobacco use, risky sexual behavior and HPV infection status will be related to increased severity of cervical abnormalities.

These research aims will be explored using information from various sources. The first research aim will be addressed using information from historical accounts, texts, and published scientific literature relevant to cervical cancer, cervical cancer-related factors, and to the political ecology of Central Appalachian poverty. The second and third research aims will be addressed using several data sources, including: the Ohio Cancer Incidence and Surveillance System; a synthetic dataset representing the point-level demographic characteristics of women in Ohio; the Census of Population and Housing (Summary File 4), Small-Area Income and Poverty Estimates Program, and the TIGER program of the U.S. Census Bureau; the Rural-Urban Commuting Codes, Rural-Urban Continuum Codes, and County Typology Codes of the U.S. Department of
Agriculture; the County Health Rankings and Roadmaps Program of the University of Wisconsin’s Population Health Institute and the Robert Wood Johnson Foundation33; property tax data of the Ohio Department of Taxation34,35; and Local Report Card data of the Ohio Department of Education36. The fourth research aim will be addressed using data from the Community, Awareness, Resources and Education (CARE) Project 3 along with the several of the State and Federal datasets mentioned above.

This research is intended to expand the current knowledge and produce a more integrated and broad description of ICCI disparities in Central Appalachia. Applying Ecosocial theory to motivate, frame, and guide the investigation may provide better information for the generation of hypotheses explaining ICCI disparities at multiple levels and spatial scales. Empirically testing theoretically defined pathways may further clarify the complex relationships affecting disparities of ICCI and cervical cancer-related factors, highlight relationships between factors and processes previously unrelated to ICCI, and add support for the applicability of using an integrative theory to explain ICCI disparities in Central Appalachia.
CHAPTER 2: BACKGROUND

Appalachia

The Appalachian region of the United States possesses distinct geographic, economic, historical, and societal characteristics. The Appalachian Regional Development Act of 1965 formally defined the U.S. Appalachian region while also establishing the Appalachian Regional Commission (ARC)\textsuperscript{13}. Today this sociodemographically diverse area includes 24.8 million people across 205,000 square miles of 13 states and 420 counties (Figure 2.1\textsuperscript{37}).

At a large spatial scale, the physical geography of Appalachia is homogenous compared to the rest of the United States. The area’s administrative units (i.e., 13 states, 420 counties, etc) are spatially contiguous stretching as far southwest as Mississippi and northeast to New York. The majority of the U.S. Appalachian mountain range – the region’s namesake – is contained within the politically defined region. The geophysical properties of Appalachia have given rise to one of the country’s richest coal deposits, primarily concentrated in West Virginia, Kentucky, and Virginia\textsuperscript{38}.

The presence of coal made many local Appalachian economies reliant on the mining industry, with the fortune of many of these economies paralleling the fortunes of the volatile coal industry\textsuperscript{38,39}. The overwhelming preponderance of coal-fired electricity production within the ARC (approximately 70\%) also ties several local Appalachian
economies – including those not confined to the coal-producing regions – to the electricity industry \(^{40,41}\). These local economic dependencies on powerful industries have been cited as ultimately affecting more than the incomes of local residents as environmental injustices have been documented in the area\(^{41}\). Appalachian residents’ realization that their economic livelihood is inextricably linked to the nearby companies, which may be simultaneously harming their health, may be creating unjust power imbalances (see ‘Disparities in Ohio Appalachia’, p. 16-17 for one such example)\(^{15,18,41,42}\).

![Figure 2.1. The U.S. Appalachian region as of 2013\(^{37}\)](image)

Historically, the U.S. Appalachian region has improved greatly over the past several decades. The region has experienced increases in per capita income, educational attainment, job growth, standards of living (i.e., plumbing, telephone service, electrical / natural gas heating), home ownership, as well as investments in economy-boosting
infrastructure such as the Appalachian Development Highway System\textsuperscript{39,43-46}. These improvements have led some authors to conclude that Appalachia has lost many of the economic characteristics which prompted its congressional designation in 1965. Still admitting to the region’s need of Federal attention, the economic plight of U.S. Appalachia has been framed not as a large-scale regional issue, but one that is localized with pockets of economic well-being contrasted with economic deprivation\textsuperscript{39,43,45,47,48}.

The ARC was charged with addressing the noticeable social and economic deprivations of the region. Since its inception, the ARC has provided both monetary and personnel support to various projects and programs including: transportation and public utilities infrastructure; social and economic development; local capacity and leadership development; as well as research and data reports motivating investments in these projects\textsuperscript{49}. Despite undeniable progress made in these target areas, various subregions of Appalachia can still be characterized as suffering disproportionately when comparing certain indicators to the Non-Appalachian United States.

**Disparities of Appalachia**

**Economic, Social, and Demographic**

**Economic:** The ARC broadly classifies a county’s economic strength into one of five categories using historically set definitions based on poverty, unemployment, and per capita market income (PCMI) relative to the U.S. average\textsuperscript{49}. A distressed county is a classification given to the most economically suppressed counties and satisfies the following criteria: 1) poverty rate at least 150% that of the U.S. average, 2) PCMI no more than \(2/3\) of the U.S. average, and 3) unemployment rate at least 150% that of the U.S. average. A county will also receive a distressed classification if the poverty rate is
at least 200% that of the U.S. average but only meets one of the other two criteria. Over half (54.4%) of Appalachian counties were listed as distressed in 1960. That percentage has steadily decreased to the latest published estimate of 21.7% of all Appalachia counties. Despite this reduction in relative economic deprivation, the spatio-temporal pattern of distressed counties is highly ordered and has become increasingly focused in the Southeastern Ohio, West Virginia, and Kentucky regions with 76% of distressed counties in 2000 also classified as distressed in four of the five previous decennial censuses.

Historically common and influential industries in Appalachia included mining, agriculture and forestry. The predominance of these ‘extractive’ industries provides indication of the richness of Appalachia’s natural resources. The development of these industries of 19th and 20th century Appalachia have been described as “colonial” because natural and human resources were exploited by outside capital. The Appalachian agriculture economy of the mid to late 19th century was highly socially stratified. At the top of the hierarchy were a minority of wealthy individuals who owned large tracts of farmland but also generated extra income through land speculation. These individuals ascended to local political power and frequently married among others of high sociopolitical status who also owned and managed land. Smaller farmers, tenants, and slaves comprised the lower levels of this agriculturally-based social hierarchy. Late 19th century industrialization gave way to the mining and timber industries. Large portions of land were sold off to absentee corporations centralizing operations in larger urban centers (e.g., Pittsburgh, Cleveland, etc), who then imported timber and coal resources from adjacent Appalachian regions.
The U.S. coal industry has ebbed and flowed throughout the last half of the 20th century with a boom in the 1970’s as a result of the U.S. oil embargo (causing an increase in demand and price of coal)\(^4^2\), followed by relatively steady production decreases at the close of the century partly due to increased international competition within coal-dependent industries\(^5^1(p.44^1)\). Another low-wage, labor-intensive industry adversely affected by the increasingly globalized world economy is manufacturing\(^5^1,5^3\); an industry that has been historically overrepresented and disproportionately affected within Appalachia by trade agreements such as the North American Free Trade Agreement (NAFTA)\(^4^8,5^1,5^3,5^4\). Despite evidence of a transition from manufacturing to service-providing industries, 19.2% of Appalachian workers were employed in manufacturing compared to 13.7% of non-Appalachian U.S. employees\(^4^8\). A 2005 economic projection by Bernard, et al. shows that increases in foreign imports of goods from low-wage, labor-intensive industries (e.g., apparel, leather, furniture, etc) will increase the likelihood of manufacturing plant closures, decrease employment, and decrease economic growth in these industries\(^5^3\). The manufacturing industry in general, and a greater number of specific plants in Appalachia, are thought to be especially susceptible to this increased, international trade liberalization\(^5^3,5^4\).

**Demographic:** The human geography and demography of Appalachia is heterogeneous, but not as diverse as the rest of the nation. According to the 2000 census, 27% of Appalachian counties and 58% of Appalachian people resided in metropolitan areas. Approximately 31% of U.S. residents are member to a racial or ethnic minority group (i.e., other than non-Hispanic white), while this percentage is only 12% in Appalachia\(^4^8\). Moreover, these 12% are not uniformly distributed across the region as two
million of the total 2.8 million minorities are concentrated in Southern Appalachia (i.e., Mississippi, Alabama, Georgia, South and North Carolina, and Virginia). Residents of Appalachia tend to be of older age compared to non-Appalachia; median age of the area was 37.3 years in 2000, compared to 35.1 years for the rest of the country. Appalachians 45 years and older comprise 38.0% of the region’s population, compared to only 34.2% among their non-Appalachian counterparts.

**Social:** Despite recent reductions in education disparities between Appalachia and non-Appalachia, a smaller percentage of Appalachian’s completed high school (76.8%) and college (17.7%) compared to non-Appalachians (80.7% with high school diplomas and 25.0% with bachelor’s degrees) in 2000. These percentages are not equally distributed across the region as Appalachia Ohio had the 2nd lowest percent of residents obtaining a bachelor’s degree at 12.3% (Appalachia Kentucky was lowest at 10.5%). In comparison, approximately 21% of Appalachian Georgians, New Yorkers, and South Carolinians had bachelor’s degrees in 2000. A relationship between education and poverty is evident in the region; however, this relationship is dependent on proximity to metropolitan areas. The expected gradient between percentage of Appalachian individuals impoverished by educational attainment (i.e., less than a high school education, high school diploma only, some college, and a bachelor’s degree or more) is similar to that of non-Appalachian residents. A poverty disparity exists within Appalachia as percent impoverished is at least 1.5 times higher when comparing residents of the most urban areas (metro county with a population of 1,000,000 or greater) to those that are most rural (fewer than 2,500 and non-adjacent to any metropolitan areas) within each of the four educational attainment categories. The educational lag of Appalachia has been
exacerbated by a "brain drain" of individuals obtaining college degrees who then migrate out of the region. It has been noted that Appalachian migration patterns during the last five years of the 20th century resulted in a loss of 25,000 – 22% of the total in 2000–college-educated persons who left the region39.

Health

Behaviors: Data on common prognostic indicators of health such as cigarette smoking, obesity, and physical inactivity are sparse. One ARC report summarizes these indicators as the percentage of the population within a county or Labor Market Area (LMA) with the indicator49. For the years 1993-1997, the distributions of smoking percentages among Appalachian white males and females by LMAs appear to be shifted towards higher ranges when compared to non-Appalachians. Geographic concentrations of Appalachian LMAs with high percentages of smokers exist in the Kentucky, West Virginia, and Southeastern sections of Ohio, especially among White males. The county obesity prevalence distributions for Appalachian White males and females appear to be only slightly shifted to the right of non-Appalachians. Again, county obesity estimates tend to geographically cluster in Appalachian Kentucky, West Virginia, Ohio, and Pennsylvania among White males and females. Prevalence of physical inactivity distributions by LMAs comparing Appalachian to non-Appalachian White males and females seem to show large disparities. Geographic concentrations among Appalachian LMAs registering high percentages of physically inactive residents are similar to those of cigarette smoking and obesity with clusters in Kentucky, West Virginia, and Ohio49.

Healthcare Accessibility: The rural, low population density characteristics of Appalachia contribute to a scarcity of medical care resources within the region. Health
Professional Shortage Areas (HPSAs) are federal designations given to areas with a primary care physician-to-population ratio less than 1:4000\(^49\). Of 406 counties analyzed in a report investigating health care disparities in Appalachia, only 109 did not have some portion designated as a HPSA\(^49\). The HPSA’s tended to concentrate in West Virginia, Kentucky, Mississippi, and Alabama. A total of 81 counties did not have a hospital (here hospital was defined as having “at least six inpatient beds, cribs, or pediatric bassinets which shall be continually available for the care of patients”)\(^49\). A minority of Appalachian counties had more than 4 hospitals, and those that did were geographically distributed in a sporadic fashion. When health care accessibility is measured by ratio of total population to physicians active in patient care, counties of Southeastern Ohio, Southern and Central West Virginia, Eastern Kentucky, and North-Central Mississippi have unfavorable ratios\(^49\).

**Cancer Disparities:** As Krieger points out, the term ‘health disparities’ has been variously defined depending on the author and context of the definition\(^55\). The National Cancer Institute’s (NCI) Division of Cancer and Population Sciences definition provides one of the most specific descriptions of what constitutes and who is often affected by a cancer-related health disparity, ‘Health disparities are differences in the incidence, prevalence, mortality, and burden of cancer and related adverse health conditions that exist among specific population groups in the United States. These population groups may be characterized by gender, age, ethnicity, education, income, social class, disability, geographic location, or sexual orientation’\(^56\). It is this working definition of a disparity that will be used to highlight the cancer burden of Appalachians relative to non-Appalachians.
The all-sites combined, age-adjusted, invasive cancer incidence rate (per 100,000) and 95% confidence interval (CI) among Appalachian males, 2001-2003 was 568.7 (95% CI: 565.9-571.5)\textsuperscript{57}. The identical measure among non-Appalachians was 539.4 (95% CI: 538.6 – 540.2). When dividing cancers into major histological primary sites, Appalachian males had incidence rates (per 100,000) that were higher than non-Appalachians in 14 out of 19 cancers (oral cavity and pharynx, esophagus, colon and rectum, pancreas, larynx, lung and bronchus, melanoma of the skin, testis, urinary bladder, kidney and renal pelvis, brain and other nervous system, Hodgkin lymphoma, non-Hodgkin lymphoma, and leukemias). Similarly, the Appalachian and non-Appalachian female all sites, age-adjusted, invasive cancer incidence rates (per 100,000) were 415.1 (95% CI: 413.0 – 417.3) and 398.6 (95% CI: 398.0 – 399.2), respectively. Of 22 major histological primary site cancers, Appalachian females had higher rates compared to non-Appalachian females across 14 sites (colon and rectum, larynx, lung and bronchus, melanoma of the skin, cervix, uterine corpus, ovary, urinary bladder, kidney and renal pelvis, brain and other nervous system, thyroid, Hodgkin lymphoma, non-Hodgkin lymphoma, and leukemias)\textsuperscript{57}.

Cancer mortality rates are also disproportionately distributed between Appalachia and non-Appalachia and within specific Appalachian regions. Smoothed cancer mortality rates among White females during 1990-1997 demonstrate disparate, high-risk areas within Appalachia compared to the non-Appalachian United States\textsuperscript{49}. Within Appalachia, White females ages 35 to 64 years experience high cancer death rates in Southern Ohio, Eastern Kentucky, and Western West Virginia. White females ages 65 years and older experience similar trends with additional high risk areas in Southwest Pennsylvania and Southern New York\textsuperscript{49}. White men of Appalachia, ages 35 to 64 years, and 65 years and
older have nearly identical cancer mortality geographic distributions to one another with high-risk areas in Southeast Ohio, Eastern Kentucky, Western West Virginia, and Northeastern Tennessee.

When comparing temporal trends in mortality over a 13-year period between 1985 and 1997, White males and females ages 65 years and older saw moderate increases (1.0-4.9%) in cancer mortality rates in approximately half of all Appalachian counties. These counties did not exhibit any clear spatial trends. Fortunately, this temporal trend was reversed for White males and females, ages 35 to 64 years as the majority of Appalachian counties experienced moderate declines (1.0% - 4.9%) in cancer death rates.

Ohio Appalachia

Many of the geographic, societal, economic, and historical characteristics of the U.S. Appalachian region are shared by Ohio Appalachia. The Appalachian region of Ohio includes 2,013,203 people across 16,011 square miles in 32 counties. Most of these counties contain a portion of the Appalachian mountain range or are within its foothills. Only 3.7% of Ohio Appalachia land is considered urban, compared to 9.2% of the land in the entire state. Despite the rural nature of Ohio Appalachia, much of the region lies in close proximity to large metropolitan areas such as Cleveland, Akron, Columbus, Cincinnati, and Pittsburgh. Ohio’s 8th largest city by population, Youngstown, is within the Appalachian county of Mahoning.

Disparities of Ohio Appalachia

Demographics: Appalachia Ohio is less demographically diverse compared to the entire State. White residents accounted for 93.3% of the Appalachian population in
2000, compared to only 84.9% in the State (it is important to note that comparisons are Appalachia Ohio to all of Ohio – including Appalachia). The age disparity observed when comparing U.S. Appalachia to the rest of the country is similar when comparing Ohio Appalachia to the State; 38.0% of Appalachians were 45 years and older compared to 35.9% for the entire State in 2000\textsuperscript{58}.

**Socioeconomic:** Using the same county economic indicator as before, five counties were listed as ‘distressed’ in 2000. All five of these counties were also distressed in 1990 and three of these five have been distressed in four of five previous decennial censuses\textsuperscript{50}. Of the ten Appalachian counties that have ever been distressed at any of the decennial censuses from 1960-2000, five have been given this category three or more times (i.e., Adams, Pike, Jackson, Vinton, and Meigs). This deep-rooted historical trend of county-level economic deprivation also appears to cluster spatially as these five contiguous counties are situated in an east-west band across Southern, Appalachia Ohio.

Per capita personal income of Appalachian residents was $6259 (17.4%) less than all Ohio residents (2008). These disparate incomes are also reflected in the percent of individuals with an income (1999) to poverty ratio less than 200%; 33.2% of Appalachians meet this criteria compared to only 26.4% of all Ohioans. Unemployment rates in Appalachia and Ohio in general reflected similar annual patterns from 2005-2009, with slight dips in 2006 followed by sharp increases. However, in all five years Appalachia’s unemployment rate was at least 1 percentage point higher than the State’s, with the most recent reporting year recording an Appalachian unemployment rate of 11.8% compared to 10.2% for the State\textsuperscript{58}.  

16
An interesting comparative evaluation of employment and average weekly wages (2008) shows that Ohio Appalachians earned less than all Ohioans in every major industry sector (i.e., private, federal, state, local)\textsuperscript{58}. Average weekly wage disparities are smallest within State employment (3.4% less among 18,058 Appalachian employees) and largest within the private sector (19.3% less among 549,118 Appalachian employees). The largest private industry sub-sector disparities between average weekly wages of Appalachians and all Ohioans is among those in the financial services and professional and business services; Appalachians earned $294 (28.8\%) and $342 (35.5\%) less than all Ohioans, respectively\textsuperscript{58}. Appalachians earned less within all private industry subsectors (i.e., natural resources and mining, construction; manufacturing; trade, transportation, and utilities; information; financial; professional and business; education and health; leisure and hospitality; and other), with the exception of natural resources and mining where Appalachians’ weekly wages are 12.5\% higher than all Ohioans’. An even clearer picture of Ohio Appalachians’ disparate economic dependence on the mining and manufacturing industries is provided when noted that average weekly wages of Appalachians working in the natural resources and mining industry and manufacturing industry are the first ($901) and second ($885) highest paying of the ten private industry subsectors, but sixth ($801) and third ($998) highest among all Ohioans\textsuperscript{58}.

The importance of the manufacturing industry to U.S. Appalachia is also true for specific counties in Ohio Appalachia. In the same U.S. Appalachia-wide study by Bernard, \textit{et al.}, five of Ohio’s Appalachian counties (i.e., Highland, Pike, Holmes, Coshocton, and Monroe) were projected to be at greatest risk from manufactured foreign imports of low-wage, labor-intensive industries\textsuperscript{53,54}. These risks include an increased
probability of plant closure, decreased employment, and decreased growth, which stems from these counties’ disproportionate, economic dependence on the manufacturing industry. Despite having a lower percentage of displaced workers (all industries) than most other Appalachian states, Ohio’s displaced workers have the highest unemployment rate among all other Appalachian state’s displaced workers; indicating that Ohio Appalachians have a more difficult time finding employment after displacement compared to others in Appalachia\textsuperscript{54}.

An analysis partitioning various sources of variability in wage inequality within the U.S. manufacturing industry found that wage inequality has steadily increased over time and within each source of variation (i.e., within non-ARC counties, between non-ARC counties, within ARC counties, between ARC and non-ARC counties, and between ARC counties) since the mid-1970’s\textsuperscript{59}. Ohio manufacturing wage inequality is highest within non-ARC counties (~77%), followed by between non-ARC counties (~12%), within ARC counties (6%), between ARC and non-ARC counties (~3%), and between ARC counties (~3%). There is evidence that the share of variability of wage inequality between ARC and non-ARC counties has steadily grown. Moreover, the proportions of wage inequality sources of variability and temporal trends are replicated within each ARC state. The authors interpret these findings as evidence that wage inequality at the geographic levels (county, region, state) and temporal scales (five year intervals) studied here may be controlled by national phenomenon related to the manufacturing industry\textsuperscript{59}.

As noted above, the dependency of individuals’ incomes and local economies on a single industry or establishment has created power imbalances which lead to environmental and economic injustices in Ohio Appalachia\textsuperscript{18,41,42}. A series of such
injustices culminated during the Summer of 2001 and involved the village of Cheshire in Gallia County, Ohio Appalachia. American Electric Power (AEP) – listed as one of the 12 largest employers in Ohio Appalachia\(^{58}\) – operates the largest electric plant in the State which neighbors Cheshire\(^{60}\). Upon installing pollution-reducing equipment in Spring 2001, an unexpected chemical reaction resulted in plumes of sulfuric-acid dispersing across local Cheshire residences. Acute health effects were reported including “eye irritation, respiratory problems, and sore throats”\(^{41}\). Air quality readings taken by AEP and the Environmental Protection Agency (EPA) recorded high levels of sulfuric acid in the town of Cheshire. The Agency for Toxic Substances and Disease Registry (ATSDR) concluded the emissions from AEP to be harmful and that steps should be taken to reduce emissions. In 2002, AEP decided to offer a $20 million buyout ($13.5 million to residents) of Cheshire in exchange for the company’s legal immunity from residents suing for any future health problems\(^{41,61}\). Many residents accepted the offer which was up to 3 ½ times their actual property value\(^{61}\). Experts agreed that the move by AEP was cost-effective for the company as the possibility of future individual lawsuits was essentially eliminated\(^{18}\). Buckley, et al. describes the historically lopsided interactions of Cheshire and the powerful utility industry in a 2005 review while also noting, ‘[Cheshire’s demise] causes us to acknowledge the distance, both physical and psychological, that separates our landscapes of power production from our landscapes of power consumption’\(^{18}\).

The educational environment in Ohio Appalachia parallels that of employment and income. Those with a high school diploma or less comprise 64.1% of the Ohio Appalachian population, compared to only 53.1% for Ohio. One in five Ohioans hold a
bachelor’s degree or higher, while only one in eight Ohio Appalachians have attained the same education level\textsuperscript{58}.

An example of educational disparities being promoted by a region’s policies is found in the funding system of Ohio’s primary and secondary public schools; a system deemed unconstitutional by the Ohio Supreme court in 1997\textsuperscript{62}. Previous to the court’s ruling, public school districts were funded based on a function of local property values. Ohio House Bill 920 passed in 1976 limited the funding schools could receive from property taxes, resulting in the introduction of the school levy system which still is in effect today\textsuperscript{63}. Questioning of the constitutionality of this educational funding system originated in Perry County of Ohio Appalachia\textsuperscript{62}. The Ohio high court concluded in 1997 that the state educational system, “fails to provide for a thorough and efficient system of common schools”, and ordered that school funding based on the current property tax system be suspended \textsuperscript{62}. According to the Ohio Office of Policy, Research and Strategic Planning, the median value of owner-occupied housing units was $82,445 in Ohio Appalachia, but $103,700 in all of Ohio (2000)\textsuperscript{58}. These disparate property values along with the unconstitutional public school funding system may be responsible for the observed disparity of expenditures per student between Appalachia and Ohio children. Average expenditure per child attending public schools within Ohio Appalachia is $8,890 compared to $9,939 among all Ohio children (2000)\textsuperscript{58}. Statistical analyses provide evidence that Ohio Appalachian students are more likely to be impoverished, reside at homes with lower assessed property values, and therefore have lower revenues available for school funding per a pupil when compared to non-Appalachia Ohioans\textsuperscript{41}. 
Health: The health of Ohio Appalachians – as measured by various self-reported healthcare access; self-report behavioral indicators; self-reported health status; self-reported cancer screening; and registry-based cancer incidence and mortality rates – is generally worse than Ohioans residing outside of Appalachia. Health care access as measured by health insurance status shows that 18.7% Ohio Appalachians between the ages of 18 and 64 years were without health insurance in 2002. This is compared to only 13.4% of all Ohioans between the ages of 18 and 64 years (Ohio Behavioral Risk Factors and Surveillance System). A map utilizing 2008 Ohio Family Health Survey (OFHS) data showing county percent uninsured, among adults, ages 18-64 years is displayed in figure 2.2. All but three Ohio counties with a percent uninsured, adult population that is at least 25% greater than the State average of 17.0% are located in Appalachia. Two spatial concentrations of contiguous counties in this disparate, uninsured category are located in Southern (Adams, Highland, Pike, Scioto, and Lawrence) and Eastern Appalachia (Guernsey, Monroe, Belmont, Harrison, Carroll, and Columbiana).

Disparities of health care access, as measured by number of health care resources per capita, between Appalachia and Ohio vary depending on the resource analyzed. Ohio Appalachians have an unfavorable physician (both medical and osteopathic physicians)-to-population ratio (1:717) compared to all of Ohio (1:388). This disparity is further underscored by noting that if Ohio Appalachia was to match the physician-to-population ratio for the entire State, the current number of physicians practicing in the region would need to double. When medical care availability is measured by number of registered hospitals, Ohio Appalachia has a more favorable ratio with one hospital for every 51,620 people compared to all of Ohio with one hospital for every 65,213 people. However,
insight into this seemingly favorable relationship is gained when noted that the number of hospital beds-to-population is lower in Appalachia with one hospital bed for every 326 people compared to one bed for every 261 people in Ohio\textsuperscript{58}. This suggests that despite having more hospitals per capita in Appalachia compared to the State, these hospitals have fewer beds and may have a decreased capacity for inpatient care. A disparity of licensed residential care between Appalachia and Ohio is also evident with one residential care license for every 21,647 people in Appalachia compared to one for every 19,464 in the entire State.
A report utilizing 2004-2007 Ohio Behavioral Risk Factors and Surveillance System (OBRFSS) demonstrated a prevalence ratio (PR) of current cigarette smoking among males and females of Ohio Appalachia (32 counties) of 1.27 and 1.29 compared to non-Appalachian Ohio males and females, respectively. A similar analysis using older OBRFSS data from 1999-2003 demonstrates that the absolute prevalence of smoking has decreased over time, but that the disparity between Ohio Appalachia (29 counties) and non-Appalachia may have widened slightly. The PR of current cigarette smoking during 1999-2003 comparing Appalachians to non-Appalachians was 1.21. Similarly, a disparity between Appalachian and non-Appalachian Ohioans involving prevalence of no physical activity in the past month appears to have widened when comparing OBRFSS results of the two time periods; from 1999-2003 prevalence of no physical activity among Ohio Appalachians was only 1.03 times that of non-Appalachians, while males and females of Appalachia reported no physical activity at a prevalence that was 1.35 and 1.22 times that of non-Appalachians during 2004-2007. Prevalence of obesity and fruit and vegetable consumption comparing Appalachians to non-Appalachians across the two OBRFSS time periods exhibit similar trends but to a much smaller degree.

A recent, in-depth analysis of health care access and associated outcomes (including various self-reported health status indicators) utilized data from the 2008 and 2010 OFHS. Risk of self-reporting a health status as ‘fair’ or ‘poor’ among Appalachians (all 32 counties except Mahoning) is 1.30 (95% CI: 1.09 – 1.53) times that of suburbanites. Appalachians have a risk of reporting 14 or more physically unhealthy days and 20 or more mentally unhealthy days in the previous 30 days that is 1.33 (95% CI: 1.07 – 1.64) and 1.47 (95% CI: 1.04 – 2.06) times that of suburbanites, respectively.
When measuring psychological distress level using the Kessler-6 scale for non-specific recent distress, Appalachians have a risk of “very high risk for distress” that is 1.41 (95% CI: 1.02 – 1.94) times that of suburbanites. These measures of health status are not uniformly distributed across Ohio as the maps in figures 2.3 A-C demonstrate. The county map – classed into quintiles – displaying percentage of survey respondents reporting good, very good, or excellent health shows that only one county, Washington, of Appalachia, ranks within the top 40% of all counties (figure 2.3-A). Moreover, a large spatial cluster of counties occupying the lowest 20% of self-report health status lie in Southern Appalachia (Adams, Pike, Scioto, Jackson, Lawrence, Gallia, Meigs, Vinton, Hocking, and Perry). Similar spatial patterns exist for the measures involving frequency of recent physically and mentally unhealthy days (figures 2.3-B and C). The Ohio counties of Adams, Gallia, Pike, Scioto, Hocking, Jackson, and Lawrence were deemed as having the most unfavorable health care access outcomes on eight measures investigated in the analysis (i.e., medical care utilization, dental care utilization, foregone medical care, foregone dental care, foregone prescriptions, self-reported health status, physically unhealthy days, and mentally unhealthy days). It is important to note that these seven, contiguous (except for Hocking) counties occupy Southern Appalachia.

A U.S. Appalachian-wide study utilizing 2002 thru 2005 data from the National Survey on Drug Use and Health (NSDUH) reports prevalence estimates of various drug use as well as serious psychological distress in the past year for 44 NSDUH sub-state county aggregates. The four county aggregates across Ohio Appalachia report decreased use of most drugs (except tobacco use across Appalachia and alcohol use in the subregion containing Ohio University). However, the entire region is among the highest 40% for
2.3.A. Percent self-reporting a health status of good, very good, or excellent

Figure 2.3. Self-report health status by Ohio counties, OFHS 2008: A) percent with good, very good, or excellent health; B) percent with less than 14 physically unhealthy days in the previous 30 days; C) percent with less than 20 mentally unhealthy days in the previous 30 days. Adapted from Hull, et al. 2011[67].
2.3.B. Percent self-reporting less than 14 physically unhealthy days in the previous 30 days
2.3.C. Percent self-reporting less than 20 mentally unhealthy days in the previous 30 days

serious psychological distress in the past year, with the Southern Ohio counties of Adams, Highland, Ross, Pike, Scioto, Lawrence, Jackson, Gallia, and Meigs occupying the highest 20\%^{69}. Focus groups involving key leaders (e.g., law enforcement, medical/mental health practitioners, school personnel, social service providers, etc) from communities within Kentucky, West Virginia, and Virginia cited several factors for why they believed substance abuse was a problem in their communities, including: depression,
anxiety, psychological trauma; geographic isolation; economic stressors such as loss of community resources and gainful employment; intergenerational substance abuse; boredom; and limited recreational opportunities for youth\textsuperscript{69}.

Disparities comparing Ohio Appalachians to non-Appalachians based on prevalence of cancer screening reported in the 1999-2003 or 2004-2007 OBRFSS exist for the following procedures: mammogram in past two years in females (PR 1999-2003 = 0.89; PR 2004-2007 = 0.92;), colonoscopy or sigmoidoscopy in past five years (PR 1999-2003 = 0.83; PR 2004-2007 = 0.82;), digital rectal exam (DRE) in past year in males (PR 1999-2003 = 1.06; PR 2004-2007 = 0.86;), and Pap smear in past three years in females (PR 1999-2003 = 0.91; PR 2004-2007 = 0.95;\textsuperscript{65,66}). As these PRs indicate, Ohio Appalachians have reported lower prevalence of screening for breast (i.e., mammogram) and colon cancer (i.e., colonoscopy/sigmoidoscopy) consistently between the two BRFSS surveys. While Ohio Appalachians actually reported higher screening prevalence for prostate cancer (i.e., DRE) from 1999-2003 than non-Appalachians, this prevalence dropped disproportionately during 2004-2007 such that Appalachian males reported a lower prevalence compared to non-Appalachians. Despite a narrowing of the disparity between Appalachian and non-Appalachian females’ reporting screening for cervical cancer (i.e., Pap smear) between the two BRFSS surveys, Appalachian females still report a slightly lower prevalence compared to non-Appalachian females\textsuperscript{65,66}.

When investigating disparities in cancer incidence among the four most frequent primary cancer sites (i.e., female breast, colon and rectum, lung and bronchus, prostate) and in addition, three screenable primary cancer sites (i.e., cervix, melanoma of the skin, and oral cavity and pharynx), Appalachians have higher rates compared to non-
Appalachia in all but female breast and prostate\textsuperscript{65,66}. The largest of these disparities during 1996-2003 was age-adjusted, invasive cervical cancer incidence (ICCI) rates with an incidence ratio (IR) of 1.38 comparing Appalachian to non-Appalachian females\textsuperscript{66}. Much smaller increases comparing Appalachians to non-Appalachians were observed for colon and rectum (IR=1.13), lung and bronchus (IR=1.11), melanoma of the skin (IR=1.03), and oral cavity and pharynx (IR=1.01). Of these five primary cancer sites recording Appalachian disparities from 1996-2003, all but oral cavity and pharynx (IR=1.12) demonstrated a narrowing or plateauing of disparities comparing Appalachian to non-Appalachians in 2002-2005 (cervix IR=1.08, colon and rectum IR=1.12, lung and bronchus IR=1.09, melanoma of the skin IR=0.98)\textsuperscript{65}. Nearly identical trends, by specific site, and reporting year-span were observed for age-adjusted, cancer mortality rate ratios comparing Appalachians to non-Appalachians: ICCI disparities decreased sharply from 1996-2003 (IRR=1.44) to 2002-2005 (IR=1.21); colon and rectum (IR 1999-2003 = 1.14, IR 2002-2005= 1.08), lung and bronchus (IR 1999-2003 = 1.08, IR 2002-2005= 1.08), and melanoma of the skin (IR 1999-2003 = 1.23, IR 2002-2005= 1.16) disparities decreased or plateaued to a smaller degree; while female breast, oral cavity and pharynx, and prostate rates are consistently lower in the Ohio Appalachian population\textsuperscript{65,66}.

Reporting cancer incidence as an average-annual rate for such large regions as the Appalachian and non-Appalachian portions of Ohio eliminates the ability to discern any inter-regional patterns at a smaller spatial scale. More meaningful patterns may be gleaned from analysis of smaller areal units, such as counties. Figures 2.4 A – G display the 1996-2006 average annual, age-adjusted, cancer incidence rates per 100,000 within Ohio for the four major primary sites and three screenable sites mentioned above. A
moving-window average of county rates was utilized to highlight any regional patterns within the State. Figure 2.4-A demonstrates how 1996-2006, average annual, age-adjusted, invasive female breast cancer incidence rates (hereafter but only in this paragraph, all rates are invasive, age-adjusted, and average annual during 1996-2006) tend to concentrate in Ohio’s major metropolitan areas (e.g., Cleveland, Columbus, Cincinnati), leaving a cluster of low-rate counties in Southeast Appalachia. ICCI incidence rates exhibit a strong Appalachian tendency, but are further concentrated in the Southern (Lawrence, Jackson, Gallia, Vinton, Meigs, and Hocking) and Eastern (Muskingum and Harrison) Appalachian counties (Figure 2.4-B). In contrast low ICCI rates in the Northern Appalachian counties of Ashtabula, Trumbull, and Mahoning seem to cluster with one another and adjacent non-Appalachian counties. Cancer incidence rates of the colon and rectum also demonstrate a strong predilection for the Appalachian region of Ohio (Figure 2.4-C). As with ICCI, smaller clusters exist in Southern (Brown, Adams, Scioto, and Lawrence) and Eastern (Perry, Morgan, Noble, Guernsey, Belmont, Harrison, Jefferson) Appalachia. Lung and bronchus cancer incidence rates also demonstrate strong Appalachian tendencies while clustering with the Southern (Clermont, Brown, Adams, Scioto, Lawrence, Jackson, Vinton, Hocking, Athens, and Meigs) counties (Figure 2.4-D). Melanoma of the skin cancer incidence rates are relatively sporadic across the State with notable concentrations spanning Appalachia and non-Appalachia regions of Southwest-Central Ohio and a low-rate area clustering in Southeastern (Pike, Vinton, Hocking, Perry, Athens, Meigs, and Washington) Appalachia (Figure 2.4-E). Cancer incidence rates of the oral cavity and pharynx exhibit some of the
2.4.A. Female Breast

Figure 2.4. Regional trends of Ohio cancer incidence rates per 100,000, 1996-2006\(^\text{70}\): A) female breast; B) cervix; C) colon and rectum; D) lung and bronchus; E) melanoma of the skin; F) oral cavity and pharynx; and G) prostate

cont’d
2.4.B. Cervix

Figure 2.4 (cont’d)

Rate per 100,000 Females


cont’d
2.4.C. Colon and Rectum

Figure 2.4 (cont’d)

2.4.D. Lung and Bronchus


cont’d
2.4.E. Melanoma of the Skin


Rate per 100,000

- 7.93 - 12.87
- 12.88 - 14.01
- 14.02 - 14.75
- 14.76 - 15.28
- 15.29 - 15.85
- 15.86 - 16.37
- 16.38 - 17.04
- 17.05 - 17.81
- 17.82 - 18.96
- 18.97 - 22.43

cont’d
Figure 2.4 (cont’d)

2.4.F. Oral Cavity and Pharynx

2.4.G. Prostate


The U.S. rate, 171.2, exceeds the upper boundary of the Ohio County
most sporadic spatial patterns of all seven sites preventing identification of any meaningful regional patterns (Figure 2.4-F). Similar to female breast cancer, prostate cancer incidence rates appear to cluster in Ohio’s metropolitan areas, with a notable collection of high-rate contiguous counties in Northeast Appalachia (Ashtabula, Trumbull, Mahoning, Columbiana, and Jefferson) (Figure 2.4-G). In contrast, Southeastern (Adams, Pike, Ross, Scioto, Jackson, and Lawrence) Appalachia represents a regional cluster of low-rate counties. A better understanding of ICCI disparities requires more detailed information on the various factors affecting ICCI exposure, susceptibility and resistance.

**Cervical Cancer**

Cervical cancer is a significant social, global, and regional public health problem. It is the third most commonly diagnosed cancer and fourth leading cause of cancer death among females worldwide\(^1\). In more developed areas the age-adjusted cervical cancer incidence and mortality rates of 9.0 and 3.2 per 100,000 females both rank seventh among all cancer sites (24 major sites). The burden of cervical cancer is more pronounced in less developed areas with age-adjusted incidence and mortality rates of 17.8 and 9.8 per 100,000 females, ranking second among all cancer sites\(^1\). In the U.S. it is estimated that 12,170 new invasive cervical cancer cases will be diagnosed and 4,220 females will die of cervical cancer in 2012\(^2\). The 2003-2007 average annual, age-adjusted, U.S. cervical cancer incidence and mortality rates were 8.1 (ranking 13\(^{th}\)) and 2.4 (ranking 14\(^{th}\)) per 100,000 females\(^3\).

Similarly, the 2003-2007 average annual, age-adjusted, Ohio cervical cancer incidence and mortality rates were 8.0 (ranking 13\(^{th}\)) and 2.4 (ranking 9\(^{th}\)) per 100,000
Regional variation of rates within Ohio is also evident when comparing Ohio-Appalachia to non-Appalachia. The 2002-2005, average annual, age-adjusted incidence and mortality rates for Ohio Appalachia and non-Appalachia were 8.7 and 2.9 per 100,000 and 7.8 and 2.4 per 100,000 females, respectively. The distribution of ICCI rates within Ohio are visually evident in the map within Figure 2.2-B. As noted above, ICCI rates from 1996-2006 seem to exhibit strong spatial patterns not only between the Ohio Appalachia and non-Appalachia regions, but within Appalachia.

Both incidence and mortality rates within the U.S. have been decreasing over the several past decades, mainly attributable to the effective screening strategies associated with the Papanicolaou (Pap) test. The 1999-2006, 5-year U.S. relative survival of cervical cancer was 70.2%, compared to 66.0% for all cancer sites combined.

The median age of diagnosis and death from cervical cancer in the U.S. from 2003-2007 was 48 and 57 years old, compared to 65 and 73 years for all cancer sites combined. Because of this younger median age at diagnosis and death, cervical cancer had the third highest average years of life lost per person dying of cancer with 26.0 years of life lost in the U.S. (year 2007). The 2003-2007, age group-specific incidence rates by race in the U.S. are shown in Figure 2.2.5. ICCI rates in the U.S. may exhibit a bimodal distribution with a peak at 40-44 years old and another at 65-69 years old. ICCI rates in African Americans exhibit a minor peak at 40-44 years old and a much larger one at 65-69 years old. In contrast, rates among whites exhibit a slightly larger peak at 40-44 years old followed by a slightly smaller peak at 65-69 years old. The age group-specific ICCI rate pattern for all races combined naturally follows that of Caucasians since the majority of ICCI diagnoses are among Caucasians (35,170 of 45,904 (76.6%) diagnosed from
This pattern has been observed in other countries and by other researchers, however the distribution has yet to be explained.

As Figure 2.5 depicts, ICCI disparities exist across various racial and ethnic groups within the U.S. The 1998-2003, average annual, age-adjusted ICCI rates in the U.S. for Whites, African Americans, and Asian and Pacific Islanders were 8.4, 12.6, and 8.3 per 100,000, respectively. Hispanics experienced a markedly higher ICCI rate compared to non-Hispanics with 14.2 compared to 8.4 cases per 100,000. African Americans and Asian and Pacific Islanders diagnosed with ICCI had a larger proportion of late-stage (regional and distant) diagnoses compared to whites (African American=47.4%, Asian and Pacific Islanders=46.8%, White= 39.0%). Similarly, 43.6% of invasive cervical cancer cases in Hispanics were late-stage compared to 39.5% of those in non-Hispanic women. These diagnoses at later stages may provide partial explanation for the higher age-adjusted, cervical cancer mortality rates observed in African Americans (5.2 per 100,000), and Asian and Pacific Islanders (2.6 per 100,000) compared to Whites (2.4 per 100,000) as well as Hispanics (3.4 per 100,000) compared to non-Hispanics (2.7 per 100,000).

Factors Related to Cervical Cancer Incidence

Human Papillomavirus

The importance of the Human Papillomavirus (HPV) in the etiology of cervical cancer incidence cannot be understated. The seminal work of zur Hausen initiated the progression of HPV’s role in cervical cancer incidence etiology from speculation in the mid-1970’s to a reified necessary cause today. This distinction of ‘necessary cause’ has important biologic and epidemiologic implications as any other ICCI risk factor is
then considered to interact with HPV in either a synergistic or antagonistic way, dependently increasing or decreasing the risk of ICCI\textsuperscript{76,77}.

**Molecular Biology:** Over 100 different HPV types have been identified, of which 40 infect the genital tract\textsuperscript{78,79}. These are categorized as ‘low risk’ or ‘high risk’ depending on their carcinogenic potential. HPV types 16 and 18 together account for approximately 70% and 83% of all squamous cell carcinoma and adenocarcinomas, respectively. Other high risk types accounting for much smaller fractions of ICCI include 45, 31, 33, and 59. Low risk types such as 6 and 11 are commonly implicated as causative agents in genital warts (i.e., condylomas)\textsuperscript{72}. It is estimated that 99.7% of all cervical cancers contain HPV DNA\textsuperscript{80}.
HPV contains double-stranded, circular DNA enclosed in an icosahedral capsid. The viral genome varies by type but is approximately 8000 base pairs, encoding only eight genes: E1, E2, E4, E5, E6, E7, L1, L2\textsuperscript{80}. Much of the variability in each HPV type’s oncogenic potential is determined by the phylogenic variability in the E6 and E7 genes, that is, high risk HPV types have evolutionary conserved polymorphisms in their E6 and E7 genes which provide the ability to cause cancer in their hosts\textsuperscript{81-84}. E6 has been shown to inhibit p53 – the host cell’s main protein controlling cellular apoptosis or cell death. E7 inhibits pRB or retinoblastoma tumor protein, which normally regulates the host’s cell cycle and replication\textsuperscript{85}. Moreover, the high risk HPV types possess two main strategies for subverting the host immune system: 1) avoidance during the early stages of infection, and 2) directly inhibiting portions of the host’s viral and carcinogenesis defense systems\textsuperscript{86}. The fact that high risk HPV infection – and therefore ICCI – is predicated on host immune system suppression is an important concept that will be detailed later.

**Natural History:** Four general, biologic steps have been identified in the causal pathway of HPV and invasive cervical cancer development: 1) infection with a high-risk HPV type, 2) establishment of viral persistence, 3) pre-invasive neoplastic progression, and 4) invasive cervical cancer\textsuperscript{85,87}. Genital HPV is most commonly transmitted through sexual intercourse, but evidence shows that the virus can be transmitted through non-penetrative, skin-to-skin sexual contact\textsuperscript{88-90}. Virus particles must enter the epithelial basal layer of the cervical tissue and gain access to the dividing basal cells, presumably via microabrasions of the cervical mucosa\textsuperscript{91,92}. Once infection is successful, the virus must establish persistence. However, 80%-90% of all HPV infections are cleared before persistence is established\textsuperscript{86}. 
Measurements of persistence have been criticized as the distinction between a persistent infection and transient one is not clear\textsuperscript{85,87}. Prospective cohort studies suggest that the highest risk of detectable cervical abnormalities occur between 1-2 years of first HPV detection (not necessarily first infection) and decline to baseline abnormality levels after 4 years\textsuperscript{93,94}. It is important to note that cervical abnormalities classified as low grade squamous intraepithelial lesions (LSIL) (i.e., Bethesda system\textsuperscript{95}) or cervical intraepithelial neoplasia grade 1 (CIN1) (i.e., World Health Organization\textsuperscript{96}) are better proxies of HPV infection/persistence than neoplastic progression. It has been suggested that high risk infections lasting longer than 2 years may serve as a cut point distinguishing cervical abnormalities which are at an increased risk of progressing to precancer from those which may naturally resolve and clear on their own\textsuperscript{97,98}. Figure 2.6 is thought to represent a typical pattern of HPV persistence, clearance, and neoplastic progression within a population of females\textsuperscript{85}. Of the 10-20\% of females who become persistently infected, approximately 80\% will have cleared the infection at 30 months. Shortly thereafter, the slope of the clearance and persistence curves appear to decrease drastically to nearly horizontal (thus, the idea that infections/persistence lasting longer than 2 years signifying an increased risk of neoplastic progression). As duration of infection continues past three years, the risk of neoplastic progression increases while the risk of clearance decreases.

A meta-analysis published in 2007 estimated the prevalence of HPV among women with normal cytology across the globe, stratified by region. A plot of age group-specific HPV prevalence by region is shown in Figure 2.7\textsuperscript{99}. The plots of the Americas and Europe all exhibit noticeable inflection points between 35-54 years old where the
steep decreases in HPV prevalence of younger years is met with an equally steep increase in older years. This inflection is comparatively less pronounced in Africa and seemingly negligible in Asia. However, it is important to note the differences in variability between each region as the age-specific prevalence estimates in Africa and Asia are far less variable than the other regions, making relative comparisons difficult. Three main hypothesis have been put forth to explain these trends: 1) an impaired immune response experienced as the result of hormonal changes in post-menopausal women causing reactivation of latent infections, 2) changes in sexual behavior of women and/or their partners at this age, or 3) a period or cohort effect.\textsuperscript{85,99}

Integration of viral genome into the host cell chromosome is viewed as an important molecular event in the progression from viral persistence to neoplastic
Figure 2.7. Age group-specific human papillomavirus prevalence among women with normal cervical cytology. Shaded regions are 95% confidence bounds.

progression\textsuperscript{100,101}. It is thought that this integration results in the deregulation of the critical E6 and E7 genes, leading to their overexpression and further inhibition of the host’s cancer defenses\textsuperscript{80}. The presence of molecular factors preceding or coinciding with integration, such as glucocorticoids, estrogen, and progesterone, has been shown to affect viral oncogene expression\textsuperscript{102-110}. Progression to CIN3 is considered a marker for and synonym of “precancer” while CIN2 has mixed and unpredictable tumorigenic potential\textsuperscript{85}. Further representing the complexity in each of the four main stages of the HPV-ICCI casual chain, CIN1 has been shown to be no better of a predictor of CIN3 than absence of any cervical abnormalities\textsuperscript{111}. However, the induction period between initial HPV infection to CIN3 formation usually occurs five years from HPV infection\textsuperscript{85}, with some studies providing evidence of precancer development within two years of age at coitarche (age at first sexual intercourse)\textsuperscript{89,90,112}. The majority of CIN3 cases is diagnosed between 25-35 years of age and is hypothesized to vary with age at coitarche\textsuperscript{85}.
Identification of independent risk factors attributed to viral persistence from those of neoplastic progression is difficult because of the associated complications in analyzing the transition between the two processes. It has been noted that those with a persistent infection of HPV 16 for 3-5 years have a 40% (absolute) risk of developing CIN3 \textsuperscript{98,113,114}.

The induction period from HPV infection to precancerous lesion (CIN3) development seems to be much shorter than the induction period from precancer to invasive cervical cancer. Obvious ethical considerations prevent current studies from determining factors contributing to the progression of precancerous lesions to invasive cancer. However, estimates of early studies suggest a 20%–30% absolute risk of invasive cancer from precancerous lesions over a 5-10 year span \textsuperscript{115,116}. Figure 2.8 represents the estimated HPV 16 and 18 attributable fractions of different stages of cervical abnormalities and invasive cancer \textsuperscript{117}. Fortunately, the role of HPV in the development of invasive cervical cancer has prompted the development of vaccines against certain HPV types.

**Human Papillomavirus Vaccine:** Two HPV vaccines have been developed which protect against cervical cancer. Gardasil (Merck and Co, Bluebell, PA, USA) provides protection against HPV types 16, 18, 6, and 11. As previously mentioned, HPV types 16 and 18 account for 70%-80% of all ICCI. HPV types 6 and 11 are responsible for 90% of all cases of genital warts \textsuperscript{85}. Cervarix (GlaxoSmithKline, Rixensart, Belgium) is focused solely on ICCI prevention and only affords protection to HPV types 16 and 18. Cervarix also contains agents to boost the immune response caused by the vaccines. Gardasil is currently recommended for use in females and males 9 through 26 years of age. Similarly, Cervarix is recommended for females 10 through 25 years of age.
Randomized controlled trials (RCT) provided strong evidence of both vaccines’ high efficacy, preventing nearly 100% of the infections of their respective HPV types after several years of follow up \(^7,^{118-121}\). Development of this efficacious primary preventive measure seems to make any future etiologic investigation of ICCI moot. However, such is not the case and justification for such analyses can be made on the following grounds: 1) current vaccines only provide protection against HPV types 16 and 18 (non-vaccine HPV types account for as much as 30% of ICCI), 2) barriers to vaccination – including finances, religious beliefs, and cultural norms may prevent universal vaccine uptake, and 3) even if the preceding two points were adequately addressed, a more precise and explanatory model of ICCI (and its antecedent causes) could provide insight for models of other diseases or conditions.

**Risks of Human Papillomavirus Infection:** Infection with HPV is the leading sexually transmitted infection (STI) in the U.S. \(^{122,123}\) HPV can infect all forms of anogenital tissue – not only the cervical tissue of females \(^{124}\). The same high-risk HPV types implicated in ICCI are also responsible for some cancers of the penis and anus in males \(^{125}\). The realization that males are also affected by HPV has increased focus on the
role that males have in heterosexual female’s ICCI risk. Risks for HPV infection are identical to those of most any STI: condom use, total lifetime sexual partners, recent sexual partners, and partner’s sexual behaviors.

Regular male condom use has been shown to provide protection for both males and females against HPV infection. A cross-sectional investigation among men estimated a significantly reduced prevalence of HPV among “always” condom users compared to “not always” users (OR=0.50, 95% CI: 0.30 – 0.83), adjusted for age, date of HPV analysis, and concurrent detection of genital warts. A prospective study aiming to evaluate the effectiveness of condom use in male-to-female HPV transmission found that females engaging in sexual intercourse only when their male partners wore a condom had a risk of incident HPV infection that was 0.3 (95% CI: 0.1 – 0.6) times that of females engaging in sexual intercourse when their male partners wore condoms only 5% of the time, adjusted for number of new partners and number of previous partners of the male partner. An ongoing longitudinal study investigating factors influencing HPV transmission in new (< 6 months) heterosexual couples report that males always using condoms have a significantly decreased prevalence of HPV compared to men never using condoms (prevalence ratio=0.62, 95% CI: 0.40 – 0.94). Interestingly, this study failed to demonstrate the identical association among females.

An individual’s risk of HPV infection is strongly related to number of sexual partners – personal number of both lifetime and recent partners, and number of partner’s recent and lifetime partners. The longitudinal study mentioned above reported HPV prevalence ratios (PRs) among females comparing 10 or more, 5-9, and 2-4 to 1 or less lifetime partners of 6.67 (95% CI: 2.91 – 15.2), 5.39 (95% CI: 2.34 – 12.4),
and 3.46 (95% CI: 1.48 – 8.10), respectively. Prevalence estimates among males were similar and both sets of sex-specific estimates yielded significant tests for trend, suggesting a dose-response relationship between lifetime partners and HPV prevalence\textsuperscript{127}. A prospective study among female college students estimated an HPV infection hazard ratio (HR) of 10.10 (95% CI: 3.24 – 31.50) for every 1 partner increase per month, adjusted for history for HSV2 and vulvar warts, and current oral contraceptive (OC) use\textsuperscript{93}.

The male partner’s role in increasing a female’s risk is well demonstrated in a pooled analysis of studies investigating HPV prevalence across various countries: females married to men who engaged in extramarital affairs had an odds of HPV positivity that was 1.45 (95% CI: 1.24-1.70) times that of females married to monogamous men, adjusted for age and lifetime number of female partners\textsuperscript{130}. Although HPV is a necessary cause, several other factors may be related to risk of ICCI dependently with HPV.

**Human Papillomavirus-Cofactors**

**Viral Load:** Earlier studies suggested HPV viral load as a possible predictor in formation of cervical abnormalities\textsuperscript{133-137}. However, more recent longitudinal studies had failed to show consistent associations with viral load and cervical abnormalities\textsuperscript{138-140}. Further analysis has provided evidence of other cofactors which are associated with viral load to affect disease severity such as HPV type, neoplastic progression, and cervical lesion characteristics\textsuperscript{139,141,142}. Despite this, it has been found that HPV type 16 viral load does increase with increasing disease severity with appreciable reliability\textsuperscript{137,143,144}.
**Tobacco Smoking:** Tobacco smoking was naturally implicated early on in the etiologic research of ICCI, albeit with inadequate control for confounding\textsuperscript{145,146}. Subsequent research on the association with better confounding control has found consistently increased risks of neoplastic progression and cervical cancer incidence among women who were tobacco smokers and were HPV positive\textsuperscript{147-153}. A 2006 collaborative study combining data from multiple sources estimated a relative risk (RR) and 95% confidence interval (CI) of 1.95 (95% CI: 1.43 – 2.65) comparing HPV positive women who were current smokers to never smokers for the risk of squamous cell carcinoma, with null findings for adenocarcinoma\textsuperscript{150}. A nested case-control study of HPV positive women investigating the risk of CIN3 estimated that ever smokers had 2.2 times (95% CI: 1.4-3.3) the odds of developing CIN3 compared to never smokers. Moreover, it appears as if a dose response was present as measured by years of smoking and average number of cigarettes per day. This association was observed after adjustment for age at coitarche, total number of sexual partners, year since start of last regular relationship, and spontaneous abortion\textsuperscript{148}. A prospective study investigating the risk of CIN3 among HPV positive women reported univariate RR’s (95% CI) of 2.1 (1.1-3.9), 2.2 (1.2-4.2), and 2.9 (1.5-5.6) comparing former smokers, women who smoked less than one pack of cigarettes per day, and women who smoked one or more packs of cigarettes per day to never smokers, respectively. These significant associations remained in multivariate settings after adjustment for oral contraceptive (OC) use and parity. Moreover, women in this cohort study were matched on cytologic interpretation of baseline Pap smears, age, and screening behavior\textsuperscript{154}. 

50
Exact mechanisms of the relationship between tobacco smoke and cervical cancer incidence have yet to be determined. Previously identified carcinogens of tobacco smoke have been found in elevated concentrations within the cervical mucosa of cigarette smokers\textsuperscript{155}. A more recent study demonstrated that in vitro exposure of HPV infected cervical cells to benzo[a]pyrene (B[a]P), a known carcinogen found in cigarette smoke, increased the viral titers of various HPV types, including types 16 and 18\textsuperscript{156}. The metabolic activation of B[a]P has also been shown to be modified by HPV infection such that cells immortally infected (a biologic step along the causal pathway to invasive cancer) with HPV type 16 are more susceptible to the DNA damage caused by B[a]P\textsuperscript{157}. Another hypothesis holds that tobacco smoke may indirectly increase risk of neoplastic progression and ICCI via its speculative immunosuppressive effects\textsuperscript{158}. Reductions in frequencies of important immune system cells have been observed in the cervical epithelium of smokers\textsuperscript{159}.

**Parity:** Several studies have estimated significant positive associations between parity and cervical cancer incidence\textsuperscript{149,160-166}. However, those studies which did not properly stratify by HPV status will not be reviewed as these ignore the necessary presence of HPV in ICCI\textsuperscript{162,165,167}. A prospective study utilizing a subcohort (N=1812) of a larger study (N=20810) found no association between parity and CIN3 or ICCI among HPV positive women after a follow up as long at 122 months\textsuperscript{113}.

A pooled analysis of retrospective, case-control studies investigated the associated risks of various reproductive factors with cervical cancer incidence among HPV positive women\textsuperscript{163}. Number of full-term pregnancies (FTP) was associated with squamous-cell carcinoma of the cervix (a histological subtype of cervical carcinoma) in a dose-response
fashion as those reporting 1-2, 3-4, 5-6, and at least 7 pregnancies yielded odds that were 1.81 (95% CI: 1.31 – 2.52), 2.55 (95% CI: 1.95 – 3.34), 2.83 (95% CI: 2.02 – 3.96), and 3.82 (95% CI: 2.66 – 5.48) times the odds of squamous-cell carcinoma of the cervix in nulliparous women, adjusting for study center, age, education, smoking status, age at coitarche, number of sexual partners, OC use, and history of Pap smears. Also of interest were the adjusted associations between age at first FTP and squamous-cell carcinoma of the cervix (at least 25 years old, 20-24, 17-19, less than 17 years old yielding odds that were 2.16 (95% CI: 1.44 – 3.24), 2.21 (95% CI: 1.72 – 2.84), 2.49 (95% CI: 1.81 – 3.44), and 4.38 (95% CI: 2.43 – 7.91) times the odds of squamous-cell carcinoma of the cervix in nulliparous women); vaginal vs. caesarean delivery 2.18 (95% CI: 1.17 – 4.05); and number of abortions (1 and at least 2 yielding odds that were 0.87 (95% CI: 0.65-1.16), and 0.58 (0.40-0.84) times the odds of squamous-cell carcinoma of the cervix in women reporting zero abortions). It is plausible that the women of the reference group in the abortion associations are more parous than the women reporting 1 or at least 2 pregnancies, providing explanation for the inverse associations with squamous-cell carcinoma of the cervix. The authors discuss how the associations with aborted pregnancies may also provide insight into the period of exposure related to the pregnancy-cervical cancer incidence relationship: the inverse association may indicate that it is not simply the act of pregnancy which increases risk of cervical cancer incidence, but pregnancy lasting into the later stages. It is possible that cervical trauma associated with live birth may explain the associations between squamous-cell carcinoma of the cervix and FTP, abortions, live birth vs caesarean delivery, and parity.
The pooled analysis also reported possible effect modification of the FTP-squamous-cell carcinoma of the cervix relationship by age at cervical cancer diagnosis. Diagnoses made before age 35 years old resulted in much higher odds ratios compared to those made at older ages. For example, stratifying the original study population and analyzing the FTP-squamous-cell carcinoma of the cervix association by those less than 35 years old at diagnosis resulted in ORs and 95% CIs of 1.63 (95% CI: 0.88-3.03), 5.56 (95% CI: 2.39-12.98), and 8.96 (95% CI: 1.43-56.13) for those experiencing 1-2, 3-4, and at least 5 FTPs compared to those that were nulliparous, respectively. Conversely, among those at least 55 years old, the ORs and 95% CIs were 0.87 (95% CI: 0.38-2.02), 2.24 (95% CI: 1.10-4.56), 1.55 (95% CI: 0.88-2.74), and 1.77 (95% CI: 1.09-2.58) for those experiencing 1-2, 3-4, 5-6, and at least 7 FTPs compared to those that were nulliparous. Again, it is possible that cervical trauma associated with live birth may explain the associations between squamous-cell carcinoma of the cervix and FTP and parity. The synergistic association with age may be due several reasons: the possibility that the births of those younger than 35 years were closer together compared to the births of those older than 55 years; or, cohort effects (i.e., those younger than 35 years had children at older ages compared to those older than 55 years. As HPV infection and persistence is related to age, this could indicate that HPV positivity during birth may increase ICCI risk).

**Hormonal Contraceptive Use:** Similar to the numerous studies reporting on the associations between parity and cervical cancer incidence, those analyzing the effects of hormonal contraceptives on cervical cancer acknowledged the necessary presence of HPV in the development of cervical cancer to varying degrees. In light of studies with more current analytic strategies, results of the earlier
studies which did not stratify on HPV status will not be discussed. Two studies provide
the most compelling evidence of hormonal contraceptive’s modifying effects on the
HPV-cervical cancer incidence relationship: an International Agency for Research on
Cancer (IARC) collaborative study by Moreno, et al\textsuperscript{191} and a meta-analysis conducted by
Smith, et al\textsuperscript{192}(which includes Moreno, et al’s study).

The IARC study drew upon data from eight previously conducted case-control
studies investigating independent risks and cofactors of cervical cancer incidence among
various countries\textsuperscript{161,162,171,172,180,183,193,194}. All effect measures were adjusted for study center,
age, education, lifetime number of Pap smears, number of births, lifetime number of
sexual partners, age at coitarche, and interactions with center and age, center and
education, and center and number of Pap smears. Self-reported ‘ever’ use of oral
contraceptives (OC) was associated with an insignificantly increased risk of both ICCI
and cervical carcinoma in situ (OR: 1.29, 95% CI: 0.88-1.91; OR: 1.42, 95% CI: 0.99-2.04, respectively). However, stratification on duration of OC use produces a more
refined view of the association as those using OC for at least 10 years, 5-9 years, 2-4
years, and 1 year had an odds of cervical cancer incidence that was 4.03 (95% CI: 2.09-
7.79), 2.82 (95% CI: 1.46-5.42), 0.80 (95% CI: 0.51-1.24), and 0.67 (95% CI: 0.41-1.08)
times that of never users, respectively.

The meta-analysis by Smith, et al which includes the IARC collaborative study
stratified on OC duration of use. The meta-analytic effect measure of cervical cancer
when women used OC for less than 5 years, 5-9 years, and at least 10 years was 0.9 (95%
CI: 0.7 – 1.2), 1.3 (95% CI: 1.0 – 1.9), and 2.5 (95% CI: 1.6 – 3.9) times that of never
users of OC, respectively. Level of confounding control was not explicitly stated. he tests
for heterogeneity between the study ORs used in calculation of the pooled OR were insignificant for all three OC duration groups. The dose-response relationships between duration of OC use and ICCI found in the IARC study and meta-analysis provides evidence that OC use may increase risk of ICCI.

Possible mechanisms explaining the OC-cervical cancer incidence relationship are still being determined. Serum estrogen levels have been shown to be elevated in women using OC whom were also at an increased risk of cervical neoplasia. Various estrogen by-products were found to increase risk of breast cancer in mouse models. More recently, a study of high risk HPV infected mice found that estrogen administration increased the risk of cervical cancer onset and progression. A study investigating estrogen levels in women found that women with CIN1/2 had significantly decreased estrogen levels compared to women with CIN1/2 and high-risk HPV infection. More precisely, the important E6 and E7 genes of high-risk HPV types demonstrate increased expression when treated with estrogen and estrogen by-products, which are contained in OC’s.

**Age at Coitarche:** Few etiologic investigations of cervical cancer incidence have explicitly analyzed age at coitarche, or age at first intercourse, as a possible cofactor with HPV infection. A nested case-control study designed to contrast the risks of acquiring HPV from the risk of neoplastic progression found that among HPV positive females, odds of CIN3 in women 16 years and younger and 17-20 years at coitarche was 3.22 (95% CI: 1.33 – 7.69) and 1.75 (95% CI: 1.06 – 2.86) that of women aged 21 years and older at coitarche, respectively, adjusted for total number of partners, years since last regular relationship, tobacco smoking, and spontaneous abortion. A cross-sectional
study designed to investigate the determinants of CIN3 estimated similar unadjusted effects for age at coitarche\textsuperscript{190}. Among HPV positive women, odds of CIN3 in women 16 years and younger, 17 years, 18 years, and 19-20 years at coitarche was 1.35 (95% CI: 0.92 – 2.00), 1.96 (95% CI: 1.19 – 3.23), 1.92 (95% CI: 1.06 – 3.45), 1.85 (95% CI: 0.88 – 3.85), respectively. However, adjustments by race, income, lifetime number of sexual partners, parity, tobacco smoking, OC use, and body mass index (BMI) rendered the age at coitarche – CIN3 relationship insignificant\textsuperscript{190}.

Absent from the literature are tested hypotheses of the biologic mechanisms involved in earlier age at coitarche leading to an increased risk of cervical cancer incidence. Authors of the previously discussed nested case-control study speculate that the observed associations between age at coitarche and CIN3 risk are presumed to serve as proxy to duration of HPV exposure; that is, individuals engaging in sexual contact at earlier ages will be exposed to HPV at earlier ages and will also experience neoplastic progression more quickly than those who were not exposed until later years\textsuperscript{148}. This hypothesized mechanism would then place age at coitarche as a risk for HPV infection or persistence as opposed to cervical cancer incidence. An alternative mechanism holds that the adolescent cervix is especially susceptible to HPV infection\textsuperscript{202}. This suggests age at coitarche and age are dependently related to HPV infection (i.e., interact with one another). Indirect evidence supporting a similar mechanistic hypothesis provides support that age at coitarche could indeed be a ICCI cofactor with HPV infection, not simply an antecedent. The adolescent cervix has been determined to be biologically and structurally different than the adult cervix and sexual intercourse among adolescents may initiate this
developmental transition\textsuperscript{203-206}. This transitional period from adolescent to adult cervix may provide a more favorable environment for the HPV virus to establish persistence\textsuperscript{206}.

**Coinfection with other sexually transmitted pathogens:** The roles of *Chlamydia trachomatis* and Herpes Simplex virus – 2 (HSV2) as cofactors with HPV in the development of cervical cancer is conflicted in the literature. Older studies will be omitted from discussion as they either failed to limit analyses to HPV positive females \textsuperscript{174,207-210}, or used inadequate laboratory assays to identify the presences of HPV and HSV2\textsuperscript{211}. A recently published nested case-control study drew upon participants of the Costa Rica HPV Natural History Study to investigate the effect of *Chlamydia trachomatis* on the HPV – CIN2 + relationship\textsuperscript{212,213}. Among females determined to be HPV positive at study enrollment, the unadjusted odds of incident CIN2+ among females positive for *Chlamydia trachomatis* DNA was 0.95 (95% CI: 0.44 – 2.04) that of females negative for *Chlamydia trachomatis* DNA. Interestingly, among HPV negative females, the unadjusted odds of incident CIN2+ among females positive for *Chlamydia trachomatis* DNA was 2.19 (95% CI: 1.03 – 4.66) that of females negative for *Chlamydia trachomatis* DNA. Moreover, females positive for *Chlamydia trachomatis* DNA at enrollment were at an increased risk of incident HPV infections during follow-up (RR=2.6, 95% CI: 1.6 – 4.2). The authors conclude that *Chlamydia trachomatis* infection may increase risk of establishing an HPV infection, but may not interact with HPV to increase the risk of neoplastic progression or cervical cancer incidence.

A pooled analysis of the IARC multicenter case-control study investigated the *Chlamydia trachomatis* – ICCI relationship among HPV positive females\textsuperscript{214}. Females testing positive for *Chlamydia trachomatis* DNA at enrollment had an odds of squamous-
cell ICCI that was 1.80 (95% CI: 1.22 – 2.66) that of females testing negative for
*Chlamydia trachomatis* DNA, adjusted for study center, age, history or Pap, OC use,
parity, and HSV2 seropositivity. *Chlamydia trachomatis* titer levels also exhibited a
dose-response relationship with ICCI among HPV positive women (Chi-square for trend
= 13.59, p-value < 0.001), adjusted for the same factors noted before. The positive
results of this study relating *Chlamydia trachomatis* coinfection to ICCI risk among HPV
positive women seem to conflict with those of the nested case-control above. However,
the inability of the IARC study to establish temporal coincidence between HPV
*Chlamydia trachomatis* may indicate that both studies converge on an identical
conclusion: *Chlamydia trachomatis* infection may be an antecedent to HPV infection and
increase host susceptibility to establishment of HPV infection, or *Chlamydia trachomatis*
coinfection with HPV may increase host susceptibility to HPV persistence.

Several competing hypotheses exist attempting to explain the mechanism behind
the role of *Chlamydia trachomatis* in cervical cancer occurrence. Inflammation
associated with *Chlamydia trachomatis* infections has been hypothesized to damage
DNA of HPV infected cells, increasing the probability of neoplasic progression. Another hypothesis holds that *Chlamydia trachomatis* infection increases the
susceptibility of HPV persistence by disrupting the normal cervical epithelial layer
allowing easier access to the dividing cells at the basal layer. Despite a lack of
consensus regarding a specific mechanism of action, most agree that *Chlamydia
trachomatis* is a cofactor with HPV in risk of ICCI.

A single case-control study summarizes the relationship between HSV2 and
cervical cancer incidence among HPV positive women. Results from the IARC,
pooled, multicenter analysis indicate that females positive for HSV2 DNA have 2.19 (95% CI: 1.41 – 3.40) the odds of squamous-cell ICCI compared to females negative for HSV2 DNA, adjusting for age, study center, HPV DNA type, history of Pap, OC use, parity, and *Chlamydia trachomatis* seropositivity. Further exploratory analyses indicate that the HSV2 association is not appreciably confounded by sexual behavior factors (i.e., lifetime number of sexual partners and age at coitarche), nor do effect measures vary when stratified on lifetime number of sexual partners. The authors note that these findings provide evidence that HSV2 infection is not serving proxy to sexual behavior factors.

The mechanistic hypotheses put forth in attempts to explain the HSV2 – cervical cancer incidence relationship are similar to those explaining the *Chlamydia trachomatis* role in cervical cancer occurrence: 1) the inflammatory response to HSV2 infection may suppress the host’s immune system or damage the DNA of HPV infected cells, and 2) ulcers created by establishment of HSV2 infection may provide opportunity for HPV to access the basal layer of the cervical epithelium. Another alternative hypothesis involves the ability of HSV2 to increase the rate of HPV replication or DNA integration in infected cervical cells.

**Immune Suppression:** Immunosuppressed individuals – renal transplant recipients and human immunodeficiency virus (HIV) positive women – have increased risks of HPV infection as well as cervical abnormalities supporting the evidence that the immune system plays a critical role in cervical cancer development. A cross-sectional study of female sex workers in Kenya estimated a high-risk HPV prevalence within HIV-positive females that was 3.6 (95% CI: 2.6-5.1) times that of HIV-negative
females, adjusting for age, number of children, and condom use\textsuperscript{222}. A prospective study investigating the natural history of HPV in relation to HIV infection found that the risk of incident HPV infection among HIV positive females with a relatively strong immune system (i.e., greater than 500 CD4\textsuperscript{+} T-cells per mm\textsuperscript{3}, less than 4000 copies of HIV RNA per ml) was 1.7 (95\% CI: 1.35 – 2.15) times that of HIV-negative women. Moreover, risk of incident HPV infection increased linearly as HIV titer levels increased\textsuperscript{220}.

A prospective cohort study investigating the incidence of SIL among women presenting to urban health clinics found that HIV positive women had 3.2 (95\% CI: 1.7-6.1) times the risk of SIL compared to HIV negative women, adjusting for transient and persistent HPV infections, cigarette smoking, and age. Despite the authors decision to adjust for HPV as opposed to stratifying, it is important to note the significant presence of an interaction between baseline HPV and HIV infections \textit{(HR=2.7, 95\% CI: 1.5-4.8)}\textsuperscript{224}. The HPV natural history in relation to HIV infection study mentioned above found that irrespective of HIV titer levels and immune status, HIV positive women have statistically significant increased risks of SIL compared to HIV negative women. Similar to the associations with HPV infection, dose response relationships between immune status and HIV titer levels with risk of incident SIL were evident\textsuperscript{220}.

Proposed mechanisms explaining the synergistic effects of HIV infection or renal transplant on HPV infection and progression to ICCI include the immunosuppressive effects\textsuperscript{221,225,226} and possible HPV promoting abilities of HIV\textsuperscript{227,228}. HIV studies have documented the lack of benefit the immune system-boosting drug regimen, combined antiretroviral therapy (ART), has conferred on cervical cancer progression\textsuperscript{229,230}. This observation is placed in context by noting that ICCI is not at epidemic proportions within
HIV-positive females, and other opportunistic infections common among HIV-positive individuals do successfully resolve with ART administration. It has been hypothesized that the lack of co-beneficial effect by ART on ICCI serves to highlight the idea that HPV is well-equipped to subvert the immune system during the establishment of persistence regardless of HIV’s immunosuppressive effects. This would then suggest either confounding by a common antecedent (e.g., sexual activity), HIV’s positive effect on HPV infection, or selection bias by HAART’s life-prolonging effect as explanations of the HIV-HPV relationship in regards to ICCI.

Several factors have been demonstrated to increase the risk of ICCI. Infection with HPV, tobacco smoking, long-term hormonal OC use, higher HPV viral load, increased parity, younger age at coitarche, coinfection with *Chlamydia trachomatis* or Herpes Simplex virus-2, and immune suppression have been shown to increase risk of ICCI. Identification of HPV – a STI – as a necessary ICCI cause implies that any other factor dependently acts with HPV to affect ICCI risk. Factors that increase the risk of infection with HPV include condom use, number of recent and lifetime sexual partners, and number of partner’s sexual partners. The HPV vaccine is one of two efficacious measures commonly used to prevent ICCI. The other preventive measure is the Pap smear.

**Pap Smear Test**

The Pap smear test is a common cervical cancer screening tool that is attributed with large reductions in cervical cancer incidence and mortality among populations since the middle of the 20th century. Lack of screening is also thought to be a predictor in an
individual’s risk of ICCI\textsuperscript{231,232}. Recently released American College of Obstetricians and Gynecologists (ACOG) and United States Preventive Services Task Force (USPTF) provides guidelines for cervical cancer screening that recommend females begin screening at age 21\textsuperscript{233}. Between the ages of 21 and 30 years, females should be screened using Pap smear tests every three years. Between the ages of 30 and 65 years, HPV testing should accompany Pap smear tests and take place every five years (or Pap smears alone every three years). Pap smear tests are not recommended for any women older than 65 years with no prior history of CIN 2 or higher. Disparities of Pap smear utilization by Appalachian residence have been demonstrated.

According to the 2010 Behavioral Risk Factor Surveillance System (BRFSS) of the U.S., 81.0\% of all women aged 18 and older with an intact cervix reported having had a Pap smear within the past 3 years\textsuperscript{234}. This same prevalence measure is essentially identical among all women in Ohio (81.7\%), but varies depending on Appalachian residence. The 2004-2007 estimated prevalence of obtaining a Pap smear within the past 3 years is 80.4\% among Ohio Appalachians, compared to 84.6\% among Ohio non-Appalachians\textsuperscript{71}. A recent study among Ohio Appalachian females found that higher socioeconomic position (SEP) (e.g., educational attainment, household income, etc), fewer major life events, lower perceived discrimination, and older age at coitarche were all associated with being within risk-appropriate cervical cancer screening guidelines (Using ACOG cervical cancer screening guidelines current through 2005)\textsuperscript{235}. Other Appalachian studies have found that a lack of medical facilities\textsuperscript{236,237}, increased distance to medical facilities\textsuperscript{237}, and a lack of mass transit\textsuperscript{238,239} may decrease the risk of having a recent Pap smear. Studies not focusing on Appalachia have demonstrated associations
between increased rurality and decreased Pap smear utilization\textsuperscript{240,241}. Pap smear utilization has also been shown to vary by demographic and socioeconomic factors.

Pap smear utilization has been shown to increase with older age\textsuperscript{242,243}. Hispanic women may be less likely than non-Hispanic women and White women may be less likely than Black women to have had a recent Pap smear\textsuperscript{241,242,244}. Those without health insurance\textsuperscript{242,244,245}, of lower educational attainment\textsuperscript{242-244}, lower household income\textsuperscript{242,246}, or residing in neighborhoods of lower income\textsuperscript{245} or higher poverty\textsuperscript{243} report decreased recent use of Pap smear. It has also been shown that single, widowed, separated, or divorced women have reduced odds of reporting recent Pap smear use compared to married women\textsuperscript{242,243,245,246}.

Research has identified several biologic and behavioral factors or measures that increase or decrease the risk of ICCI. Widespread implementation of the Pap smear has led to the steady decline of cervical cancer rates since the mid 20\textsuperscript{th} Century. The HPV vaccine may further reduce ICCI rates. Despite what is known about ICCI, geographic disparities have persisted in regions of the U.S that have been shown to also suffer from persistent socioeconomic deprivation. Details of the effects that the social environment has on health may be needed to understand the persistence of ICCI disparities.

**Socioenvironmental Influences of Health**

The health of an individual is thought to be partially dependent on factors exogenous to that person, such that socioenvironmental phenomena an individual may be member to or engage, may yield differential health outcomes\textsuperscript{247-257}. Constructs of these phenomena can be divided into two main categories: individual-level and group-level\textsuperscript{258,259}. Examples of individual-level measures may include educational attainment,
income, health insurance status, and relative SEP, to name a few. Group-level social measures are frequently tied to a certain ecologic or administratively defined level (e.g., household, census tract, county, etc.) and include median household income, economic deprivation, income inequality, social disorganization, social cohesion, rurality, sunlight, pollution, and many others\(^\text{258}\). Important differences within both individual-level and group-level socioenvironmental constructs have prompted further classification. Individual-level measures may be of absolute (also termed ‘objective’ and including educational attainment, health insurance status)\(^\text{260}\) or relative nature (also termed ‘subjective’ and include relative SEP)\(^\text{260,261}\). Group-level measurements of exposure have been further classified as aggregate\(^\text{262,263}\) (also termed ‘contextual’\(^\text{263}\), ‘analytical’\(^\text{264}\), or ‘derived’\(^\text{259}\)), environmental\(^\text{262}\), structural\(^\text{264}\), or integral\(^\text{263}\) (also termed ‘global’\(^\text{262,264}\)). These numerous classifications have been created based on theoretical and pragmatic grounds indicating that measurements of each respective class may relate to the same phenomenon in a different way\(^\text{258,259,262,263,265-267}\). The relationships between these classes of socioenvironmental variables, heath, and ICCI have been demonstrated to varying degrees and capacities (see Kogevinas, et al, 1997 for a review on social inequalities and cancer\(^\text{268}\) and Wilkinson and Marmot, 2003\(^\text{269}\) or Marmot and Wilkinson, 2006\(^\text{250}\) for reviews of social influences with other health outcomes). Moreover, several researchers note the importance of these constructs across the life-course such that the longitudinal relationships between socioenvironmental measures and health determines an individual’s risk of adverse health consequences at any given point in time\(^\text{249,250,270-272}\).
Individual-level, Absolute Socioenvironmental Measures

Individual-level, absolute measures of socioenvironmental constructs have frequently been investigated as factors affecting health (measured in various ways) and ICCI. Two seminal reports of health disparities along social strata were results of the Whitehall study (1978)\textsuperscript{273} and the Black report (1980)\textsuperscript{274} – both conducted in England. These early studies utilized occupational grade as proxy to social status and found dose-response relationships between various morbidity and mortality disease measures and occupation. Countless investigations have since been conducted using different individual-level, absolute socioenvironmental measures, but report mostly consistent, proportional associations: health generally increases with increasing socioenvironmental status.

Only two studies have reported on the associations of ICCI risk and any absolute, individual-level socioenvironmental measures\textsuperscript{275,276}. An international meta-analysis investigated the effect of individual-level income and education on ICCI. Significant, unadjusted associations were found comparing odds of ICCI among females of middle (OR=1.41, 95% CI: 1.25 – 1.59) and low (OR=1.97, 95% CI: 1.80 – 2.15) education level to those of a high education level. Similar associations were estimated between income (middle OR: 1.32, 95% CI: 1.12 – 1.55; low OR: 2.68, 95% CI: 2.30 – 3.14; reference = high) and ICCI. However, tests of heterogeneity between studies suggest that the studies are significantly different (p-value for chi-square < 0.05 for all measures reported above). Restricting analysis to all North American studies yielded significant, inverse associations when comparing levels of education to ICCI (middle OR: 1.39, 95% CI: 1.15 – 1.67; low OR: 2.30, 95% CI: 2.01 – 2.63; reference = high), and chi-square tests of
heterogeneity between studies resulted in high p-values, suggesting that effect measures of these studies did not significantly differ. The unadjusted estimates of these associations preclude discussion involving mechanistic explanations between education and ICCI due to the high probability of confounding by other established causes.

The 2005 study by Khan, et al produces a more refined picture of the association between education and cervical cancer development. A multinomial logistic regression framework was utilized to compare odds of CIN 2, or CIN 3 or greater, to less than CIN 2. Among HPV positive females, odds of CIN 3 or greater among those with less than a high school education was 2.4 (95% CI: 1.5 – 3.7) times that of females completing college, adjusted for race, source of medical care, age, study center, referral diagnosis, smoking history, parity, and number of recent Pap smears. This identical measure was insignificant when comparing the equivocal measure of CIN 2 to less than CIN 2, suggesting that education (or a proxy of) may play a role in instances where HPV infection and persistence progress to pre-cancer or invasive disease, but not the more mild forms of cervical dysplasia.

Specific biologic mechanisms linking absolute, individual-level socioenvironmental constructs to health are dependent on the specific measure (e.g., education, health insurance status, income, etc) and outcome (e.g., diabetes mellitus II, ICCI, etc). Some researchers have made general statements about the associations between this class of constructs and health by noting that they may serve proxy to, or predict more well-established associations with health such as disease prevention efforts (e.g., cancer screening, vaccines, etc), behaviors (i.e., diet, exercise, drug use), or medical care access (i.e., physical access, affordability, and medical care availability).
Others conducting research within the biomedical sciences have investigated possible, direct biologic mechanisms between absolute, individual-level socioenvironmental measures and disease, or established disease precursors (cancer-related pathways; other health pathways). The mechanistic depth and complexity of the numerous relationships involved in these neurobiological pathways are beyond the scope of this review. It should suffice to note that all pathways connecting absolute, individual-level socioenvironmental measures to any health condition, which do not serve proxy to other well-established health-affecting factors, operate as a temporal antecedent to stress (conceptually defined as, ‘… a transactional process arising from real or perceived environmental demands that can be appraised as threatening or benign, depending on the availability of adaptive coping resources to an individual.’).

Individuals of lower, absolute socioenvironmental strata have increased biologic stress levels (measured in various ways) which have been shown to have adverse effects on multiple pathways involving numerous body systems, including: endocrine, nervous, digestive, excretory, and immune.

**Individual-level, Relative Socioenvironmental Measures**

Relative, or subjective, socioenvironmental constructs measure an individual’s perceived status or rank relative to others in society. A measure commonly employed is the MacArthur Scale of Subjective Social Status which asks individuals to self-rank their social status by choosing a ladder rung among a series of rungs which reflect the various social strata among all of society. This construct is theorized to be functionally different than objective constructs such that it may more precisely measure: an individual’s appraisal of material resources longitudinally, standards of living, and social
The degree of correlation between objective and subjective measures may be dependent on demographic and other socioeconomic factors.\textsuperscript{261}

An investigation of healthy, white women conducted in 2000 estimated the effects of subjective social status after controlling for objective measures of social status (income and education) on various physiologic and psychological measures. Subjective social status was not significantly related to self-reported physical health after adjustment for education and income, but was significantly, inversely correlated with heart rate, chronic stress, and “poor” sleep.\textsuperscript{287} Similarly, subjective socioeconomic status (SES) was demonstrated to increase the risk of rhinovirus- and influenza-induced ‘common colds’,\textsuperscript{289} as well as metabolic syndrome,\textsuperscript{290} after adjustment for objective SES (both studies using measures of education and income).

Biologic mechanisms explaining results of the observational studies above are similar to those put forth attempting to explain the absolute SES – health relationships. A cross-sectional neuroimaging study found an inverse association between subjective SES and volume of a portion of the brain involved in experience, regulation, and response to psychological stress, after adjustment for individual-level, absolute SES, group-level SES and a host of other potential confounding psychological factors.\textsuperscript{291} Limitations of the study not only included the inability to establish temporality due to study design, but other research suggests the possibility of reverse-causation as individuals with smaller volumes of this section of the brain may be at an increased risk of perceiving their SES as lower compared to those with larger brain volumes.\textsuperscript{249}
Group-level Socioenvironmental Measures

Group-level measures of socioenvironmental constructs of exposure have a storied history of use and acceptance as valid measures and predictors of health (see Pearce, 2000 for a brief overview)\textsuperscript{292}. These measures share similarities in that they represent characteristics of a group – usually defined at an ecologic or administrative level. Despite this unifying similarity, group-level measures of socioenvironmental constructs differ on their measurement, use, and existence of an individual-level analog\textsuperscript{258,265}.

Group-level measurements that are formed from attribute aggregations (e.g., income, percent smoking, income inequality, etc) of individuals are commonly referred to as aggregate\textsuperscript{262} or derived\textsuperscript{263,265}. Despite these measures often having an individual-level analog, it has been demonstrated in practice\textsuperscript{293,294} and theory\textsuperscript{295,296} that rarely can one reliably use aggregate measures as proxy to the individual-level. Moreover, research suggests that aggregate measures may be assessing a different construct than the individual-level analog\textsuperscript{263,265,297}.

Environmental group-level constructs measure some aspect of the physical environment (e.g., rurality, sunlight, elevation, ambient pollution level, etc) and are often referenced to a geographic area\textsuperscript{262}. Some of these measures may indeed have an individual-level analog, but for financial and logistical feasibility, are oftentimes measured at a defined geographic level. Since many of these physical processes can be reliably modeled as a continuous distribution across space, geostatistical methods can be applied to yield exposure estimates at a smaller spatial scale than that of the group-level data\textsuperscript{298}. 
Global\textsuperscript{262} (also known as integral\textsuperscript{263,265}) constructs measure attributes of a group and do not have an individual-level analog. These measures are thought to affect all individuals within the group in a similar fashion and include social cohesion, social disorganization, and legislation. Structural constructs such as social networks, which measure relationships between individuals of a group, are oftentimes considered global\textsuperscript{258,264}.

Studies incorporating group-level socioenvironmental measurements while excluding any individual-level socioenvironmental measurement are subject to the well known ecologic fallacy, especially if the group-level measurement is aggregate\textsuperscript{262,265}. Including both individual-level and group-level constructs into a single framework, oftentimes called a multilevel or hierarchical framework is a common approach to investigating the unique contribution each construct may have on a given outcome\textsuperscript{265,266,296,299}. Despite the attention given to the ecologic fallacy, it is equally important to be mindful of the atomistic, psychologistic, and sociologistic fallacies when the outcome under study is thought to operate in a multilevel fashion \textsuperscript{259}. These fallacies can arise when, like the ecologic fallacy, the unit of measurement does not match the unit of inference (i.e., atomistic), or when both units of measurement and inference match but relevant information about one of the levels is omitted (i.e., psychologistic or sociologistic). Numerous population-based studies investigating the relationships between group-level socioenvironmental constructs and various health outcomes and prognostic health indicators have been conducted in ways that avoid these fallacies\textsuperscript{250,266,268,269,272}. 
Studies estimating the effects of group-level socioenvironmental constructs on ICCI are few and use either aggregate or environmental group-level measures. A study using registry data and including 80% of the U.S. population found that the relationship between county rurality and ICCI is dependent on age, adjusted for race, county poverty, and diagnosis year (younger than 35 years, rural versus metropolitan incidence ratio (IR) = 1.72, 95% CI: 1.44-2.05; 35 thru 44 years, rural versus metropolitan IR = 1.31, 95% CI: 1.11-1.55; 45 thru 54 years, 55 thru 64 years, 65 thru 74 years, and 75 thru 84 years IRs are all insignificant; at least 85 years, rural versus metropolitan IR = 0.73, 95% CI: 0.55-0.99).

A similar study by the same author employed a multilevel model to determine the effect that county aggregate measures of education (i.e., less than 75 percent greater than a high school education versus at least 85 percent), income (i.e., median household income less than $35,000 versus at least $50,000), poverty (i.e., at least 20 percent below 100% federal poverty level versus less than 10 percent), Pap smear usage (i.e., by 5 percent increase of ever reporting a Pap test), and tobacco use (i.e., by 5 percent increase of current tobacco smokers), as well as the environmental, county measure of rurality (i.e., rural, suburban, metropolitan), have on ICCI, adjusting for individual-level age and race. County education, income, poverty, and tobacco smoking were all significantly related to ICCI (education IR = 1.41, 95% CI: 1.35 – 1.47; income IR = 1.15, 95% CI: 1.11 – 1.20; poverty IR = 1.20, 95% CI: 1.15 – 1.26; tobacco use IR = 1.02, 95% CI: 1.01 – 1.03). A significant cross-level interaction between race and rurality was observed as Asian and Pacific Islanders (API), African Americans (AA), and Caucasians all had significant associations with ICCI when comparing those residing in rural versus
metropolitan counties (API IR = 1.73, 95% CI: 1.24 – 2.39; AA IR = 1.15, 95% CI: 1.05 – 1.26; Caucasian IR = 0.96, 95% CI: 0.93 – 0.99), but insignificant associations when comparing suburban versus metropolitan area of residence\textsuperscript{302}.

One study investigated ICCI temporal trends by county poverty level, as well as cross-sectional incidence rates by various census tract aggregate measures and race using U.S. registry data\textsuperscript{300}. Results of the temporal analysis demonstrate that age-adjusted ICCI rates have steadily decreased from 1975 – 1999 across all county-poverty strata. Despite evidence of a reduction in disparities between rates of counties by percent impoverished, those counties with at least 20 percent of their population in poverty had an ICCI rate that was at least 33 percent higher than counties with less than 10 percent of an impoverished population, maintained over all 25 years. Gradients exist of ICCI rates between census tract aggregate measures of poverty and education strata by race. Incidence rates of census tracts occupying the lowest quintile of education (percent with at least a high school diploma) were double those of the highest quintile among all races. Race-specific, incidence rates comparing census tracts of the lowest quintile of education to the highest were four times higher among American Indians, slightly less than double among Hispanics and non-Hispanic Caucasians, 1.5 times among African Americans, and only 1.2 times higher among Asian and Pacific islanders\textsuperscript{300}.

The aforementioned mechanisms attempting to link individual-level socioenvironmental constructs to health via stress apply to group-level measurements as well. One study demonstrated a relationship between a major neurotransmitter, serotonin, and an aggregate group-level measure, such that those living in less-advantaged communities had blunted serotonin responses, after adjustment for individual-level SEP,
age, sex, diet, cognitive ability, and trait conscientiousness. This cross-sectional study was unable to establish temporality between the associations and is subject to ecologic fallacy as a multilevel model was not utilized.

Socioenvironmental constructs that may affect health have been identified and classified according to operating level (i.e., individual- or group-level). Further differentiation of individual- and group-level constructs can be made depending on various criteria. Several of these specific socioenvironmental construct types have been demonstrated to affect health via various physiologic mechanisms and pathways. However, these physiologic mechanisms do not provide explanation for how these socioenvironmental factors arise or relate to population-level patterns of health and health disparities. The development of health-related socioenvironmental factors and their relationships to health outcomes is the subject of several social epidemiologic theories and models.

**Social Epidemiologic Theoretical Frameworks**

The numerous epidemiologic theories which explicitly frame the health of populations in terms of social and political processes can be traced back to the mid-twentieth century. Commonality between these frameworks include: 1) the belief that societal phenomena (e.g., economic, social, political, etc) cause population level distributions of health, and 2) fluctuations in these phenomena will cause fluctuations in population health. All socially-oriented theories of disease distribution also reject the main implications of biomedical reductionism and methodologic individualism, which together posit that the study of disease is best conducted within the realm of the biologic, chemical and physical sciences, and that all characteristics of disease are reducible to the
individual as only the individual should be the target of preventive and therapeutic efforts. Just as these shared notions and philosophies help define social theories of disease distribution, several divergent characteristics add unique perspectives to various theoretical frameworks. The frameworks most relevant to the issue of disparities of ICCI are detailed below.

**Social Determinants of Health**

Many ideas of Marmot and Wilkinson’s *social determinants of health* (SDH) framework were borne of the previously established theories of *political economy of health* and *social production of disease*. These two preceding theories of SDH emphasized the role that politics, economics, and social phenomena played – while highlighting processes of consumption, production, reproduction, and exchange associated with these phenomena – in not only the societal patterning of health distributions, but the persistence and spread of health disparities. The idea that social determinants govern societal patterns of disease is shared by SDH and the parent theories; however, SDH diverges in several important ways: 1) de-emphasizes the role of production and associated processes which ultimately give rise to the social determinants of health, 2) increases attention to the biologic pathways linking social determinants and their adverse presentation as disease morbidity and mortality within individuals, and 3) lifecourse approaches highlight the necessity to acknowledge the cumulative effects of unfavorable exposures longitudinally. That SDH eschews explicit analysis of sociopolitical processes that may give rise to social determinants of health is an important point which helps to define a boundary of the theory. Although not assumed to develop *ex nihilo*, social determinants of health are situated ‘upstream’ in the causal chain.
connecting health-related outcomes, which “Good social policy…”\textsuperscript{5}(p. 5) should intervene upon to improve social conditions; omitting analysis of inequitable policies which may analogously be considered ‘bad’.

The Centers for Population Health and Health Disparities Multilevel Model of Population Health and Health Disparities

The Centers for Population Health and Health Disparities (CPHHD) is a National Institutes of Health (NIH)-funded program that aims to, ‘address disparities and inequities in the prevalence and outcomes of several diseases, particularly cancer and heart disease’\textsuperscript{4}. Although not intended to act as a stand-alone theory providing foundation for constructs and the relationships between observed phenomena, the model developed by Warnecke, \textit{et al} encompasses a hybrid of social theories and frameworks\textsuperscript{250,305-311}, serving as the conceptual model used by the CPHHD\textsuperscript{312}. Broad categories of determinants, also referred to as levels, are labeled distal, intermediate, and proximal. Distal determinants include constructs of social conditions and policies (e.g., poverty, public policy, culture, norms, discrimination, etc) as well as institutional context (health care system, economic system, legal system, political system, family, organized religion, etc) and comprise ‘fundamental’ causes of population health inequities\textsuperscript{306}, for ‘… their influence is solely reflected at the population level in the variation of disease or poor health’\textsuperscript{312}(p. 1610). Intermediate determinants include such constructs as social context (e.g., collective efficacy, poverty, SES gradient, etc), social relationships (e.g., social networks, social isolation, religious participation, employment, etc), and physical context (e.g., building quality, pollution, open space, neighborhood stability, etc) which serve as ‘links through which the environment affects … proximal determinants’\textsuperscript{312} (p. 1611). Lastly,
proximal determinants are those constructs measured at an individual-level (though may represent context, as in race) such as sociodemographics (e.g., age, SES, race, health status, etc), risk behaviors (e.g., tobacco use, alcohol use, diet, sexual behavior, etc), biologic responses (e.g., obesity, depression, stress, alcoholism, etc), and biologic and genetic pathways (e.g., allostatic load, genetic ancestry, genetic mechanisms, etc).

Additional characteristics of the CPHHD model are the possibility of constructs to interact across all three broad categories of determinants, or the possibility for constructs to affect population disparities independently and directly. This conceptual model seems to be an improvement over the SDH model by nature of the hybridization of SDH and other theories resulting in a greater emphasis on sociopolitical and economic processes giving rise to other possible social determinants of health\(^{250,305-311}\). However, operationalizing the relationships between levels, constructs, interactions, and population health is somewhat difficult because the CPPHD model is not intended to serve as a theory. Models motivated by other theories may better guide these relationships, while expounding at length on others of importance, resulting in more clarified frameworks and consistent applicability. Nancy Krieger’s Ecosocial theory of disease distribution – first described in a 1994 paper calling for epidemiologists to frame studies with more robust theories – is one such theory\(^{255}\).

Ecosocial Theory

_Ecosocial_ theory serves as basis for determining, “who and what drives current and changing patterns of social inequalities in health”\(^{313}(p. 672)\). _Ecosocial_ theory incorporates elements of the socially-oriented theories _social production of disease_\(^{256,314}\), _political economy of health_\(^{256,315}\), and _social determinants of health_\(^{250,269}\), while
additionally emphasizing the importance of historical context and ecology. Nearly identical to the theory of political ecology invoked by ecologists and human geographers (see King), *Ecosocial* theory posits that population patterns of health are the consequences of historically interacting social, economic and political processes. Emphasizing *processes* and *relationships* as opposed to static independent factors, interactions between phenomena should be dynamically considered across spatiotemporal scales.

Four key constructs theoretically ground the design, conduct, analysis and conclusions drawn from investigations employing *Ecosocial* theory: 1) embodiment; 2) pathways of embodiment; 3) cumulative interplay of exposure, susceptibility and resistance; and 4) accountability and agency. Embodiment is the biologic manifestation of an individual’s interactions with the biophysical and social world. That which is embodied is a health condition of interest (e.g., ICCI, diabetes, stress, etc). Embodiment stipulates that contributors of health disparities originate mostly from sources external to the individual, and ‘…cannot be reduced to allegedly ‘innate’ characteristics…’ (p. 215).

Moreover, embodiment implies that static disease mechanisms or fundamental causes cannot account for cause-specific population patterns of disease as phenomena are context-dependent. The framing of and possible contributors to the specific health condition are directed by how societal phenomena become embodied.

Pathways of embodiment are the interacting relationships between given factors and processes that contribute to the embodied consequence. Specific health disparities may have multiple causative pathways and these pathways should be jointly detailed. The various pathways producing a specific health disparity can be generalized and
categorized as involving: ’ 1) economic and social deprivation, 2) toxic substances, pathogens, and hazardous conditions; 3) discrimination and other forms of socially inflicted trauma; 4) targeted marketing of harmful commodities; 5) inadequate or degrading health care; and 6) degradation of ecosystems’.

The third Ecosocial construct – cumulative interplay between exposure, susceptibility, and resistance – highlights the importance of timing, accumulation of and responses to embodied exposures. Krieger admits to the complexity of Ecosocial theory and the impossibility of incorporating all possible pathways along with their context-contingent relationships between exposure, susceptibility, and resistance into a single investigation. This limitation of causal theorizing accords with philosophical debates contending that complete and perfect explanation of causal relationships may be impossible pragmatically, if not theoretically. The purpose of Ecosocial theory’s first three constructs is to more accurately and thoroughly characterize causes of health disparities.

The last construct involving accountability and agency explicitly identifies by who and what, and to what extent disparities are created and propagated in a population, as well as why the disparity is studied and how it is explained. Agency refers to the ability to act, while accountability describes the degree of responsibility for actions taken and avoided. Coinciding with the social production of disease and political economy of health, this fourth construct requires a reflexivity of epidemiologic theories and theorists. Reflexivity, as described in social theory, is referring to the recursive relationships between researchers and that which is researched.
This review demonstrates that several theoretical frameworks have been
developed in order to explain how socioenvironmental factors and processes may relate
to health disparities. Consistent links have been found between various
socioenvironmental measures and several health outcomes. These models can be useful to
explain health disparities in regions where socioeconomic deprivation has persisted; such
as the Central Appalachian region of the U.S. Although much is known about the natural
history, risk factors, and preventive measures of ICCI, little is known about the factors or
processes that may contribute to disparities of ICCI. Accordingly, research that aims to
integrate and empirically test relationships between the causes of socioeconomic
disparities with the causes of ICCI disparities within Central Appalachia may be needed.
CHAPTER 3: METHODS

Despite steady reductions overall, disparities of invasive cervical cancer incidence (ICCI) exist across socioeconomic and geographic gradients. ICCI within the Central Appalachian region of the United States has remained higher than adjacent non-Appalachian regions for decades. Several social, economic, political, and ecologic factors are believed to have contributed to the lower socioeconomic conditions of Central Appalachia. The goals of this research were to theoretically integrate current knowledge of the causes of socioeconomic disparities with the causes of ICCI disparities within Central Appalachia, to describe the multilevel and spatial characteristic of ICCI in Ohio, and to investigate relationships between sociopolitical, economic, and biobehavioral factors together with the prevalence of cervical abnormalities. This chapter describes the methods used to conduct this research.

Datasets

Several datasets were used to explore the goals of this research. These datasets are described in detail in the following sections.

Community Awareness Resources and Education Study

The Community Awareness Resources and Education (CARE) Project is a Center for Population Health and Health Disparities, funded by the National Cancer Institute (CA105632, 2003-present). The purpose of CARE I (2003-2007) was to utilize
transdisciplinary approaches to address disparities of cervical cancer incidence among females of Appalachia Ohio. Three separate studies within CARE I were designed, each addressing different issues related to cervical cancer incidence. The specific goal of CARE I project 3 (hereafter referred to as ‘study’ or ‘project’) was to determine the social, behavioral, and biologic variables which may contribute to an increased risk of abnormal cervical cytology. The theoretical frameworks serving as basis for this project’s design, data collection, analysis, and interpretation were the *Social Determinants of Health* and the CPHHD multilevel model of population health and health disparities.

**Study Design and Data Collection**

A two-stage sampling scheme was utilized to achieve a sample of females that would allow estimation of the effects of area-level socioenvironmental and individual-level biobehavioral and socioeconomic factors on abnormal cervical cytology. At the time of this study’s conception, Ohio Appalachia consisted of 29 counties (all present day counties but Ashtabula, Trumbull, and Mahoning) with females being recruited only from these 29 counties. Participants were sampled using a stratified sampling plan. Each of the 29 counties was placed into one of four contiguous regional groupings (regions 1, 2, 3, and 4 contained 6, 7, 9, and 7 counties, respectively). Counties within regions were classified as urban or rural according to their membership to a U.S. Census Bureau designated metropolitan statistical area (MSA). All urban counties were included while two rural counties were randomly chosen within each of the regions to enable a heterogeneous sample of females with respect to urban/rural residence (Figure 3.1).
A census of clinics within these counties was generated (N=171 counties). A total of 22 clinics met the following inclusion criteria: performed Pap smear tests, reported at least 200 Pap screenings per month, and served a demographically diverse population. These clinics were approached and asked to participate in the study of which 14 agreed (63.6%). An additional three clinics that were approached following enrollment of the initial 14 also participated in the study (Figure 3.1).
Patient listings were gathered from each of the 17 enrolled clinics and females were randomly selected on a monthly basis. Study personnel reviewed medical records of 6028 randomly selected females to deem eligibility. Females were eligible if at least 18 years of age, a resident of Appalachia Ohio, not pregnant, no history of hysterectomy or invasive cervical cancer, and seen within a participating clinic within the last two years. Possibly eligible participants were approached and asked to participate in person by a study nurse on the day of the female’s routinely scheduled Pap smear examination. Of the 4043 total eligible females, 2394 agreed to participate (59.2%). Participants were informed of the study, signed a written informed consent, completed a self-administered questionnaire, and provided blood and saliva samples before undergoing the Pap examination.

The self-administered questionnaire included information related to a participant’s demographics; socioeconomics; perceived stress; cervical screening history; sexual history and behavior; and alcohol and tobacco use. Relevant demographic information included age and race. Basic socioeconomic data included: detailed work status (e.g., employment status, unemployment length, type of work, etc), annual household income, educational attainment, and marital status. Cervical screening information included date of last Pap smear, age at first Pap smear, number of Pap smears within the last six years, and ever had an abnormal Pap smear. Several questions related to sexual-health and included: age at menarche / menarche cessation, number of pregnancies, age at first pregnancy, number of miscarriages / stillbirths / abortions, age at first intercourse, age of first sexual partner, sex of sexual partners, number of sexual partners (and various patterns and characteristics of these partners), total sexual partners in past year, number
of new sexual partners in past year, sexual partner’s sexually transmitted infection (STI) status, doctor diagnosis of genital warts / condylomas / human papillomavirus (HPV), prostitution history, contraceptive use / type (including condom use), condom use prior to and after 18 years of age, and current condom use. Lastly, information of alcohol and drug use was gathered including: number of days alcohol was consumed in past month, number of drinks consumed on days when alcohol was consumed, number of days of binge drinking (at least 5 drinks in one sitting), greater than 100 cigarettes ever smoked, smoking status, and cumulative smoking years. The data generated from this questionnaire included nearly all established individual-level factors associated with cervical cancer incidence. After completion of the self-administered survey, participants proceeded to undergo the routine pelvic examination and Pap smear.

Two additional cervical tissue samples (ThinPrep: Hologic, Inc., Bedford, MA and STM: QIAGEN Inc., Valencia, CA) were collected either for storage within the CARE biological specimens repository (ThinPrep) or for HPV typing (STM). Pap smear results of each participant were reported using the Bethesda system. The Bethesda system was developed with the primary intent of facilitating communication between laboratory personnel interpreting Pap smears and clinicians who were responsible for providing proper medical care based on the results. Pap smear results were broadly classified into three categories: 1) negative for intraepithelial lesion or malignancy (NILM), 2) epithelial cell abnormalities, and 3) other. Results deemed as representing epithelial cell abnormalities were further interpreted and classified into cell types (i.e., squamous or glandular). Squamous cell abnormalities, the most common type, were classified into the following: 1) atypical squamous cells (ASC), which are often further
classified and presented as, of undetermined significance (ASC-US), or cannot exclude high-grade squamous intraepithelial lesion (ASC-H); 2) low-grade squamous intraepithelial lesion (LSIL); 3) high-grade squamous intraepithelial lesion (HSIL); and 4) squamous cell carcinoma. Figure 3.2 is a graphical representation of the Bethesda system cervical abnormalities nomenclature\(^\text{95}\), other nomenclatures, and risk of cervical cancer incidence\(^\text{85}\). The additional cervical sample was assayed for HPV typing (AMPLICOR: Roche, Indianapolis, IN) at the Centers for Disease Control and Prevention (CDC).

A total of 284 females presented with epithelial cell abnormalities leaving 2081 with normal Pap smear tests (29 had an unsatisfactory or missing sample). Females with normal Pap smear tests were matched to those with abnormal Pap smear tests based on clinic site and date of Pap smear (within a three month window) and randomly selected to achieve a three to one, normal (n=847) to abnormal Pap smear test ratio. HPV typing was conducted for these 1131 females and classified into: 1) HPV negative, 2) high risk (HR)-HPV positive (types 16, 18, 31, 33, 39, 45, 51, 52, 56, 58, 59, 68, 73, or 82), and 3) low risk (LR)-HPV positive (any other type).

**Geocoding Participants**

Geocoding of participants was necessary for any analysis involving geographically or administratively defined constructs (e.g., county of residence, etc). An ad hoc process was developed to translate each female’s residential address into a set of georeferenced coordinates (i.e., latitude and longitude). ArcGIS (version 9.3, ESRI, Redlands, CA) was the primary platform used to accomplish this task. However, many addresses which did not successfully geocode were processed utilizing
the U.S. Postal Service (USPS), manual queries utilizing a year 2000 road file originating from the Ohio Department of Development, county auditor websites, and Google maps.

A general schema of the geocoding process can be found in Appendix A. In brief, addresses were visually scanned for errors or non-geocodable formats (i.e., no street address provided). Those that appeared to have viable addresses were geocoded using ArcGIS software. An alphanumeric street and zip zone address-locator was built using year 2000 street data from the Ohio Department of Development. This address locator served as the reference data to which participant’s addresses were matched. ArcGIS uses
probabilistic matching, the details of which are beyond the scope of these analyses. However, probabilistic matching allows for users to define ‘confidence thresholds’ for various parameters in the matching process\textsuperscript{320}. Spelling sensitivity and minimum matching score were both set to 80 while minimum candidate score was left at the default value of 10. Therefore, those addresses that did not achieve a matching score of 80 were considered unmatched. These addresses were further scrutinized and analyzed using the various means mentioned above (see schema in Appendix A). Of all cases providing some form of a mailing address, eight (2.8\%) provided a post office box (P.O. box), and four (1.4\%) addresses could not be assigned a latitude or longitude. Fifty-seven (6.7\%) controls only provided P.O. boxes while only one (0.1\%) control address could not be assigned a latitude or longitude.

The geographic coordinates of each participant will be used to determine residence in various administrative units as well as distance to clinic performing the pelvic exam. Clinics will be geocoded following the same procedures and schema described above. Euclidian network distance between female’s residence and clinic will be calculated using the latitudes and longitudes previously generated along with the road file used in constructing the address locator. Residence within various administrative units will be ascertained using tools within ArcGIS’s ArcToolBox (i.e., Spatial Join).

**Ohio Cancer Incidence and Surveillance System**

The Ohio Cancer Incidence Surveillance System (OCISS) at the Ohio Department of Health (ODH) is the population-based central cancer registry for the state of Ohio. The passage of Ohio House Bill 213 in 1991 made cancer a reportable disease in Ohio and established the OCISS\textsuperscript{321}. All primary malignancies, with the exception of basal and
squamous cell carcinoma of the skin and carcinoma in situ of the cervix, diagnosed on or after Jan. 1, 1992, are required to be reported to the OCISS at ODH within six months of diagnosis (Ohio Revised Code Section 3701.26)\textsuperscript{26}. The mission of the OCISS is to "provide high-quality cancer incidence and mortality data, and analyses; to monitor the occurrence of cancer, and to identify high-risk populations for the various types of cancer; to support qualified research efforts and the design and evaluation of interventions concerned with the prevention, early detection and control of cancer in Ohio"\textsuperscript{26}. Goals of the OCISS include: 1) gather high quality cancer incidence data that are accurate, complete and timely; 2) promote and provide data for studies designed to identify factors relating to cancer etiology, prevention, and control; 3) respond to requests from individuals and organizations in Ohio and across the US for cancer data and analyses; and, 4) provide data and expertise for cancer research and education\textsuperscript{26}.

Reporting sources of OCISS data include Ohio hospitals, ambulatory surgery and radiation therapy centers, free-standing pathology laboratories, nursing homes and physician offices. Ohio residents diagnosed with cancer in neighboring states as well as Florida are reported to the OCISS by central cancer registries in those states\textsuperscript{26}. Cancer cases are also identified via Ohio death certificate files from the Vital Statistics Program at ODH\textsuperscript{322}. Follow-back surveillance is conducted for each of these cases in which a type of cancer is listed as the underlying cause of death to determine if the case has been previously reported. OCISS requires each reporting source to submit a minimum of 39 variables on all incident cancer cases within six months of diagnosis and/or the first date of contact. These required variables include demographic factors (e.g., age, race, sex, etc.) and additional information required to compute incidence rates (e.g., county of
residence, residential address, etc). County of residence is assigned based on the residential mailing address provided by the reporting source.

Cancer cases in the OCISS are coded according to the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), codes C00.0-C80.9. Cancer cases are grouped by ICD-O code into 24 major cancer site/type groupings (including “Other Sites/Types”) based on those created by the Surveillance, Epidemiology and End Results (SEER) Program of the NCI.

Completeness is calculated with a formula that uses SEER incidence to mortality rate ratios and Ohio mortality rates to estimate Ohio’s expected incidence rate. Then, Ohio’s “reported” rate is calculated based on the data collected and compared to the expected rate to determine how complete data submission is for any given diagnosis year at 24 months after the end of that year. The national standard for complete cancer case reporting is 95 percent. Using this formula, the OCISS exceeds national standards for every diagnosis year. The integrity of OCISS data is examined and maintained through the procedures of the OCISS Data Quality Assurance Plan, which is based on the NAACCR Standards for Cancer Registries, Volumes I, II and III. Quality assurance and data cleaning are conducted using the CDC CRS Plus software (Centers for Disease Control and Prevention, Atlanta, GA), with additional and final edit checks and analyses performed via SAS (version 9.2, SAS Institute Inc., Cary, NC) and SEER*Stat software (Surveillance, Epidemiology, and End Results Program, Bethesda, MD).
Federal and State Databases

A Synthesized U.S. Geospatial Database

A database of the U.S. population representative of the age, race, and sex geospatial distribution was created by researchers of RTI International. Details of the methodology used to create this database can be found elsewhere. Briefly, the database was synthesized from various data sources: Public Use Microdata Area (2005-2009), U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing Data (TIGER) Block Group Boundaries (2010), American Community Survey (2005-2009), and LandScan USA (2008). The resulting synthetic population estimated the 2005-2009, individual-level and household demographic characteristics reported by the U.S. Census.

Area-level Covariates

Area-level covariate data were gathered from the 2000 Census of Population and Housing (Summary File 4) and 1995 Small-Area Income and Poverty Estimates Program of the U.S. Census Bureau; the 2000 Rural-Urban Commuting Codes, 2003 Rural-Urban Continuum Codes, and 2004 County Typology Codes of the U.S. Department of Agriculture; the 2010 and 2012 County Health Rankings and Roadmaps Program of the University of Wisconsin’s Population Health Institute and the Robert Wood Johnson Foundation; 1990 and 2004 property tax data of the Ohio Department of Taxation; and 1994 and 2004 Local Report Card data of the Ohio Department of Education (see Appendix B for years of measurement, dataset, data source, description, classification, and geographic level of support for area-level variables).
County-level variables considered in modeling were: tobacco smoking (percent of adults identifying as current smokers and reporting at least 100 lifetime cigarettes, 2002-2008); Chlamydia rate (Chlamydia rate per 100,000 population, 2007); teen birth rate (teen birth rate per 1,000 female population, 15-19 years, 2000-2006); percent without health insurance (percent of the population less than 65 years without health insurance, 2005); primary care provider rate (primary care provider rate per 100,000 population, 2006); income (composite of income inequality at the household level, 2007, and percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 2005-2007); high school graduation rate (percent of 9th grade cohort that graduates in 4 years, 2005-2006); percent with a bachelor’s degree (percent of population of at least 25 years with 4-year college degrees or higher, 2005-2007); family and social support (composite of percent of single-parent households, 2005-2007, and percent of adults without social or emotional support, 2005-2008); childhood poverty (percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 2007); physical inactivity (percent of adults reporting no leisure time physical activity, 2009); recreational facility rate (recreational facilities per 100,000 population, 2009); urban (Rural-Urban Continuum Code, 2003); high mining employment (income earnings from mining industry is at least 15% of all income earnings, 2004); and high manufacturing employment (income earnings from manufacturing industry is at least 25% of all income earnings, 2004).

School district-level variables considered in modeling were: childhood poverty (percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 1995); pupil support expenditures per pupil (pupil support services,
guidance services, health services, psychological services, speech pathology, audiology services, social activities, work activities, and student instructional media services ÷ average daily membership, 1994 and 2004); instructional expenditures per pupil (salaries of teachers, teacher aides, or paraprofessionals, as well as materials, computers, books and other consumables used with students within the classroom ÷ average daily membership, 1994 and 2004); total expenditures per pupil (administration, building operations support, staff support, pupil support, and instructional support ÷ average daily membership, 1994 and 2004); and taxable land value per pupil (value of all property subject to taxation and after application of assessment rates ÷ average daily membership, 1990 and 2004).

Census tract-level variables considered in modeling were: high mobility (Rural-Urban Commuting Codes, 2000); commute time (mean travel time to work among workers of at least 16 years, 2000); percent employed in manufacturing (percent of total civilian population of at least 16 years employed within the manufacturing industry, 2000); median household income (median of household incomes, 2000); percent unemployed (percent of total civilian population of at least 16 years in the labor force whom are not employed, 2000); percent impoverished (percent of total population with incomes less than 100% the federal poverty level, 2000); percent without a high school diploma (percent of population of at least 25 years without a high school diploma or equivalent, 2000); and percent without a bachelor’s degree (percent of population of at least 25 years with 4-year college degrees or higher, 2000).

Ohio’s county subdivision (year 2000), unified school district (year 2012), and census tract (year 2000) spatial boundary files were gathered from the U.S. Census
Bureau’s TIGER program. All shapefiles were projected in *equidistant conic* prior to any spatial operations. A series of one to one spatial join operations enabled the linkage of area- to individual-level data. Spatial join match type ‘closest’ was chosen due to small topological mismatches between the individual- and area-level data (e.g., home address coordinates lying 10 meters beyond the Ohio border). This join match type ensured that every individual point was assigned to a given area.

**Data Use and Institutional Review Board Approval**

CARE protocol was approved through the institutional review boards (IRB) at The Ohio State University (OSU), University of Michigan, and the CDC. Identifying information was needed for the analysis of CARE and OCISS data (residential address for geocoding and neighborhood residence). Accordingly, data use agreements were completed and IRB approval of this study protocol was granted from IRBs of both OSU (Protocol number 2012C0014) and ODH (Protocol number 2012-23).

**Data Analyses**

**Research Aim 1: An Ecosocial Perspective of Disparities of Invasive Cervical Cancer Incidence in Central Appalachia**

**Research aim 1:** To apply the *Ecosocial* theory and integrate current understanding of ICCI risks and pathways with extant theories of Central Appalachian underdevelopment and poverty so as to inform the social, political, economic and cultural processes contributing to persistent ICCI disparities.

This research aim did not explicitly use quantitative data or statistical analyses. Literature searches for any aspects related to the historical, social, economic, political, ecologic, environmental, demographic, and health-related characteristics of Central
Appalachia was conducted. The databases, directories, and websites searched included: PubMed\textsuperscript{332}, Google Scholar\textsuperscript{333}, Web of Knowledge\textsuperscript{334}, The Ohio LINK Library Catalog\textsuperscript{335}, and The Appalachian Regional Commission webpage\textsuperscript{336}. Material types returned from the literature search and subsequently used, included: peer-reviewed scientific articles, non-peer-reviewed reports, conference proceedings, laws and statues, and secondary historical texts. It is important to note that primary historical documents were not used.

Research Aims 2 & 3: Multilevel and Spatial Characteristics of Invasive Cervical Cancer Incidence in Ohio

**Research aim 2:** To investigate and explain the variation of ICCI in Ohio by partitioning it into individual- and various area-level sources according to the theoretical framework of aim one.

**Research aim 3:** To use the results of aim two and describe and summarize the spatial distribution of ICCI within each level.

**Study Sample**

**Study Design:** Invasive cervical cancer cases were from Ohio’s population-based state cancer registry, the Ohio Cancer Incidence Surveillance System (OCISS)\textsuperscript{26}. Females diagnosed with invasive cervical cancer (International Classification of Disease – Oncology codes C530-C539\textsuperscript{323}) from 1996 to 2009 were considered for inclusion (N=7124). Race, age at diagnosis and geographic coordinates of home address at diagnosis were also gathered from OCISS. Nine cases were excluded due to nonsensical or missing geographic coordinates resulting in 7115 cases. A control sample was generated from an extant database simulating the point location and demographic characteristics of Ohio’s entire female population. Details of the methodology used to
create this database can be found elsewhere. Briefly, the database was simulated from a synthesis of various data sources: Public Use Microdata Area (2005-2009), U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing Data (TIGER) Block Group Boundaries (2010), American Community Survey (2005-2009), and LandScan USA (2008). The resulting synthetic population estimated the 2005-2009, individual-level and household demographic characteristics reported by the U.S. Census. A total of 14,230 controls were randomly selected using a random number generator to achieve a 1 to 2 case to control ratio (21,345 cases and controls total). Each control’s estimated age, race and geographic coordinates of home address were also gathered from the synthetic dataset. The final individual-level dataset consisted of Ohio invasive cervical cancer cases and synthetically generated controls all containing complete information on age, race and geographic coordinates of home address.

**Measures:** Area-level covariate data were gathered from the 2000 Census of Population and Housing (Summary File 4) and 1995 Small-Area Income and Poverty Estimates Program of the U.S. Census Bureau, the 2000 Rural-Urban Commuting Codes, 2003 Rural-Urban Continuum Codes, and 2004 County Typology Codes of the U.S. Department of Agriculture; the 2010 and 2012 County Health Rankings and Roadmaps Program of the University of Wisconsin’s Population Health Institute and the Robert Wood Johnson Foundation; 1990 and 2004 property tax data of the Ohio Department of Taxation; and 1994 and 2004 Local Report Card data of the Ohio Department of Education (see Appendix B for years of measurement, dataset, data source, description, classification, and geographic level of support for area-level variables).
County-level variables considered in modeling were: tobacco smoking (percent of adults identifying as current smokers and reporting at least 100 lifetime cigarettes, 2002-2008); Chlamydia rate (Chlamydia rate per 100,000 population, 2007); teen birth rate (teen birth rate per 1,000 female population, 15-19 years, 2000-2006); percent without health insurance (percent of the population less than 65 years without health insurance, 2005); primary care provider rate (primary care provider rate per 100,000 population, 2006); income (composite of income inequality at the household level, 2007, and percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 2005-2007); high school graduation rate (percent of 9th grade cohort that graduates in 4 years, 2005-2006); percent with a bachelor’s degree (percent of population of at least 25 years with 4-year college degrees or higher, 2005-2007); family and social support (composite of percent of single-parent households, 2005-2007, and percent of adults without social or emotional support, 2005-2008); childhood poverty (percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 2007); physical inactivity (percent of adults reporting no leisure time physical activity, 2009); recreational facility rate (recreational facilities per 100,000 population, 2009); urban (Rural-Urban Continuum Code, 2003); high mining employment (income earnings from mining industry is at least 15% of all income earnings, 2004); and high manufacturing employment (income earnings from manufacturing industry is at least 25% of all income earnings, 2004).

School district-level variables considered in modeling were: childhood poverty (percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 1995); pupil support expenditures per pupil (pupil support services,
guidance services, health services, psychological services, speech pathology, audiology services, social activities, work activities, and student instructional media services ÷ average daily membership, 1994 and 2004); instructional expenditures per pupil (salaries of teachers, teacher aides, or paraprofessionals, as well as materials, computers, books and other consumables used with students within the classroom ÷ average daily membership, 1994 and 2004); total expenditures per pupil (administration, building operations support, staff support, pupil support, and instructional support ÷ average daily membership, 1994 and 2004); and taxable land value per pupil (value of all property subject to taxation and after application of assessment rates ÷ average daily membership, 1990 and 2004).

Census tract-level variables considered in modeling were: high mobility (Rural-Urban Commuting Codes, 2000); commute time (mean travel time to work among workers of at least 16 years, 2000); percent employed in manufacturing (percent of total civilian population of at least 16 years employed within the manufacturing industry, 2000); median household income (median of household incomes, 2000); percent unemployed (percent of total civilian population of at least 16 years in the labor force whom are not employed, 2000); percent impoverished (percent of total population with incomes less than 100% the federal poverty level, 2000); percent without a high school diploma (percent of population of at least 25 years without a high school diploma or equivalent, 2000); and percent without a bachelor’s degree (percent of population of at least 25 years without a 4-year college degree, 2000).

Ohio’s county subdivision (year 2000), unified school district (year 2012), and census tract (year 2000) spatial boundary files were gathered from the U.S. Census
Bureau’s TIGER program\textsuperscript{330}. The individual- and three area-level shapefiles were projected in \textit{equidistant conic} prior to any spatial operations. A series of one to one spatial join operations enabled the linkage of area- to individual-level data. Spatial join match type ‘closest’ was chosen due to small topological mismatches between the individual- and area-level data (e.g., home address coordinates lying 10 meters beyond the Ohio border). This join match type ensured that every case and control was assigned to a census tract, school district and county. All of Ohio’s 88 counties, 611 of 615 school districts, and 2861 of 2940 census tracts were represented in the spatially joined dataset (i.e., the difference represent areas registering zero cases or controls).

Transformations, reclassifications or composites of both individual- and area-level variables were created prior to analyses. Individual-level race was collapsed into ‘White’(N=18,184), ‘African American’ (N=2,506), and ‘Other’(N=655). Appropriate age categorizations were investigated in response to previous research suggesting that age and race interact to affect ICCI risk\textsuperscript{3,8,72}. Age- and race-specific (white or African American only) crude ICCI rates calculated using 19 age groups suggested age categories of: less than 39 years, between 39 years and 53 years, and greater than 53 years (Figure 3.3). All subsequent statistical model building included age, race and an age-race multiplicative interaction term.

Except for urban, high mining employment and high manufacturing employment all county-level variables were standardized and modeled as continuous Z-scores. County-level Rural-Urban Continuum codes were dichotomized into code=1 (part of a metropolitan statistical area with at least 1,000,000 population) versus codes 2 through 9 (all counties not part of a metropolitan statistical area with at least 1,000,000 population).
to create the county variable ‘urban’. County-level high percent of mining employment and high percent of manufacturing employment were dichotomous by definition (see above or Appendix B). All school district-level variables were standardized and modeled as continuous Z-scores. Except for Rural-Urban Commuting Area codes, all census tract-level variables were standardized and modeled as continuous Z-scores. Rural-Urban Commuting Area codes were dichotomized into codes less than or equal to 5 (all metropolitan, micropolitan core, and micropolitan non-core/high commuting) and code=8 (small town high commuting) versus codes 5.1-7.4 and 8.1-10.6 create the census tract variable ‘high commuting’. Principal Component Analysis (PCA) of five variables commonly used to indicate area-based socioeconomic constructs – median household income, percent unemployed, percent impoverished, percent without a high school diploma, and percent without a bachelor’s degree – was utilized to reduce data
dimensionality and possibility of collinearity between census tract-level variables. PCA analysis proceeded using the data correlation matrix and only components with Eigenvalues greater than 1.0 were considered for inclusion in models. The standardized components resulting from the PCA were used in subsequent modeling.

**Statistical Modeling**

Multilevel logistic regression models were developed to estimate associations between ICCI and covariates, explain variation in ICCI and to partition the odds of ICCI into individual- and area-level sources of variation. This partitioning, through the use of area-specific random intercepts, enabled the geographic investigation of individual- and area-level patterns in the odds of ICCI through the creation of maps. Cases and controls were assumed to follow a binomial distribution while the area-level random intercepts were assumed to follow a normal distribution. The probability of ICCI was related to the explanatory variables through the logit link function (see Appendix C for model notation). A compound symmetry random effects covariance structure was used in all models.

Model building was conducted separately for each of the three geographic levels. This approach allowed for comparisons of area-level variance across geographic levels. Explanatory models were built in a backwards elimination fashion to yield a parsimonious model that explained the most variation possible in ICCI. The individual-level, fixed-effects covariates age, race, and an age-race multiplicative interaction term were included in all model building. Type-III statistics were used to judge the order of potential fixed-effects covariate removal from the model; at each step of the model building process, that covariate with the associated lowest model fit (i.e., highest p-value
of the type-III test of fixed-effects) was removed from the model. This removal continued until all covariate p-values associated with type-III tests of fixed-effects were below the \textit{a priori} defined Type-I error level, 0.05. At this point, those covariates that were previously removed from the model were reevaluated for inclusion into the model; covariates initially removed were entered into the model, in the order they were removed. Any reentered fixed-effects covariates associated with statistically significant type-III tests were included to produce a final ‘main-effects’ model (note: ‘main effects’ in this application includes the age-race interaction term). Two-way cross-level interactions were judged for model inclusion by entering multiplicative interaction terms between any fixed-effects individual- and area-level main effect into the model. Those fixed-effect interactions associated with statistically significant type-III tests were included in the final model.

Sensitivity analyses were conducted using each of the three final models described above. Models calibrated using the entire Ohio dataset were stratified by Appalachian residence. This enabled the comparison of relationships between factors and ICCI throughout the State (the full dataset) to those in Appalachia, Ohio. Differences between estimates of the two models may suggest a possible interaction between that factor, Appalachian residence, and ICCI. Results of models restricted to Appalachian residence are tabulated in Appendix D.

Exponentiation of any fixed-effect model coefficient yielded an odds ratio (OR). Odds ratios and 95\% confidence intervals (CI) were calculated for all fixed-effect covariates of the final model. A statistical test of whether the log-OR = 0 was conducted and assessed using a t-test. Stratum-specific ORs and 95\% CIs were reported for those
covariates involved in an interaction. Odds ratios of covariates modeled as Z-scores were back calculated and reported in terms of the data and not the Z-score. Main-effect ICCI ORs estimated from Z-score variables were reported as a one standard deviation (SD) change in those variables. The original data of these Z-score variables were analyzed to ensure that a one SD change was not beyond the range of the data.

Changes in the level-specific distributions of ICCI were investigated through analyses of individual-level residuals and random intercepts estimated from multilevel null and final models. Multilevel null models are those only including an overall intercept and area-specific random intercepts (i.e., a model excluding explanatory variables).\textsuperscript{342-344} The partitioning of variation between individual- and area-level sources of a null model depends only on the case-control distribution within individuals, within areas, and between individuals and areas. The residual variation of the null model serves as a baseline to which the residual variation of a final model can be compared. Any changes in the residual spatial distribution between null and final models of the same geographic-level can be attributed to the covariates within the final model.\textsuperscript{344,345}

Various items of information were reported from each of the two types (i.e., null and final) by three geographic-level explicit models (six sets of results). Model fit was estimated from the quantity two times the negative log-likelihood, conditional on the random intercepts (-2LLR|RI). Model fit could be assessed by comparing -2LLR|RI values of null and final models of the same geographic level; -2LLR|RI\textsubscript{Null} – (-2LLR|RI\textsubscript{Final}) is approximately distributed Chi square with degrees of freedom (df) equal to df\textsubscript{Final} – df\textsubscript{Null}.\textsuperscript{341} A random intercept variance parameter, \(\tau^2\), was estimated and used in calculations of the intraclass correlation coefficient (ICC).\textsuperscript{342,343,346} Random intercept
variance and ICCs can be used to quantify the variation attributable to both individual- and area-level processes. The latent response method was applied in calculating the ICCs; the individual-level variance of the models was assumed to equal the constant, $\pi^2/3$. Therefore, model ICC was equal to, $\tau^2 / (\tau^2 + \pi^2/3)$. A statistical test of whether $\tau^2=0$ was conducted and assessed using a Chi-square mixture distribution. Tests of $\tau^2=0$ yielding a sufficiently high Chi-square value provide evidence that a model including random intercepts is better than one without.

Laplace approximation estimation techniques were used to estimate fixed-effects, individual-level residuals, and model fit. Fixed-effect coefficients calculated from models calibrated using Laplace approximation estimation techniques represent maximum likelihood estimators. Individual-level residuals were reported as Pearson-type, on the scale of the data, and using area-level predictors in computing the statistic. Pearson-type residuals were used in anticipation of unequal residual variance across space. Area-level predictors were used in residual calculations so that the spatial distributions of individual-level residuals between null and final models could be compared (i.e., individual-level residuals of final models would be nearly identical across all three geographic-specific models if area-level predictors were not included in their calculations).

Residual Subject-specific Pseudo Likelihood (RSPL) estimation techniques were used to determine the model random effects (i.e., area-level variance parameter, ICC and random intercept estimates). Use of RSPL estimation techniques to derive random effect components of the model ensured that the values were the estimated best linear unbiased predictors (EBLUP). Estimates of EBLUP variability were based on the
prediction standard errors. A statistical test of whether a given area’s associated random intercept=0 was conducted for all areas of each model and assessed using a t-test (using the EBLUP and prediction standard errors). Those tests yielding sufficiently high t-values provide evidence that a given area’s ICCI odds were significantly different from the overall odds of the entire State.

Spatial Analysis

Basic exploratory spatial data analyses (ESDA) techniques were used to characterize the spatial pattern of ICCI in Ohio. A nonparametric estimate of the spatially-varying relative risk was calculated to produce a fine-grain, yet stable surface of ICCI probabilities. These probabilities were visualized in a map and represent the degree to which ICCI risk varies across the State. Details of the methods used to produce the surface are described elsewhere. Briefly, local probabilities of ICCI are calculated using a kernel smoother. The magnitude of estimated variability across the region is mainly determined by the radius defining the local neighborhood in which ICCI probabilities are calculated, or bandwidth. The optimal bandwidth was chosen from a series of bandwidth values as that which minimized the likelihood function of a cross-validation procedure. The resulting probability surface may be superior to those that aggregate cases and controls into discrete geographic areas for the purpose of displaying rates; the choice of geographic level at which to display rates is precluded as risk estimates are not restricted to discrete geographic units in probability surface calculations. However, the resulting probability surface should not be used to locate potential ‘clusters’, but rather as an exploratory tool that allows visualization of the spatial variation of ICCI across the state in a manner not restricted by geographic areas.
A global summary of the inter-point spatial dependence, or global clustering, of ICCI cases and controls was estimated using Ripley’s K-function\textsuperscript{352-354}. Details of the methods used in calculating the K-function are described elsewhere\textsuperscript{352-354}. In this application, the K-function is providing an estimate of the distances at which the density of cervical cancer cases may be statistically different than the density of controls. Evidence of statistically significant deviations between case and control density was assessed visually using significance bands. Significance bands represent the K-function null values at which to compare the K-function values estimated from the observed data, \((\text{Kidot-K})_{\text{obs}}(r)\)\textsuperscript{352-354}. The distribution of null values was generated under a random labeling null hypothesis, using 99 replications. Those distances at which \((\text{Kidot-K})_{\text{obs}}(r)\) are higher than the highest values of the null distribution, \((\text{Kidot-K})_{\text{hi}}(r)\), or lower than the lowest values of the null distribution, \((\text{Kidot-K})_{\text{lo}}(r)\), provide evidence of spatial clustering or spatial regularity, respectively, at those distances\textsuperscript{352-354}. Results of the K-function are oftentimes presented graphically with distance on the X-axis, differences in case and control density on the Y-axis, and plots of \((\text{Kidot-K})_{\text{obs}}(r)\), \((\text{Kidot-K})_{\text{hi}}(r)\) and \((\text{Kidot-K})_{\text{lo}}(r)\). In addition to characterizing any spatial clustering of ICCI in Ohio, the distance at which cervical cancer cases significantly clustered was used to inform subsequent definitions of a ‘neighborhood’ in later analyses.

Spatial patterns of individual-level residuals (i.e., Pearson-type residuals of the six multilevel models described above) were investigated the same way as the original case-control data. Probability surfaces were calculated from the residuals and results were visualized in maps. The kernel bandwidth of the probability surface calculations and the resulting scale of the legends representing those probabilities were held constant across
individual-level maps. The use of identical bandwidths and scales eased comparisons of spatial distributions across maps.

Spatial patterns of area-level residuals (i.e., random intercepts from the six multilevel models described above) were investigated through map visualization and calculation of the global and local version of Moran’s I\(^{355,356}\). The global Moran’s I provides a single summary measure of the spatial clustering of an area-level variable (e.g., area-level random intercepts) throughout the entire study region (e.g., Ohio). A positive Moran’s I index indicates global clustering of area-level values, a negative index indicate global regularity or ‘repulsion’, and an index=0 indicates no global clustering\(^{356}\). A statistical test of whether the index is significantly different from ‘0’ is also used. The local Moran’s I, or local indicator of spatial autocorrelation (LISA), belongs to a class of methods applied to discrete, area-based spatial data for the purpose of measuring the local “…similarity between each region’s associated value…and those of nearby regions”\(^{345}(p.237)\). As the description indicates, calculation of LISA statistics necessitates defining nearby regions or ‘neighborhoods’. The choice of neighborhood definition may have substantial influence on LISA results\(^{345}\). As a sensitivity analysis, two different neighborhood definitions were used in each of the six sets of random intercepts: 1) edge and node contiguity (i.e., queen), and 2) zone of indifference (ZINDF). Queen contiguity defines a given region’s neighbors as those regions which share any border portion with that region’s border. ZINDF defines neighbors using a combination fixed distance and inverse distance weighting approach: neighbors within a region’s distance threshold contribute equal weight to the region, beyond this zone neighbors contribute a weight that is inversely proportional to the distance specified. Because the K-function quantifies the
distances at which individual-level ICCI cases cluster, t K-function results were used to determine the distance threshold value for ZINDF neighborhood definitions. The degree to which individual-level based K-function results will accurately measure a given geographic level’s series of neighborhoods may be related to the average unit area within each geographic level; as the average area of units within a geographic level increase beyond the distance indicated by the K-function, the degree to which the neighbors will be accurately characterized within that geographic level will decrease. LISA results are presented in tables and visualized in maps (ZINDF LISA results will be presented as main findings while analyses using queen contiguity neighborhood definitions can be found in Appendix E). It is important to note that evidence of spatial dependence within the random intercepts indicates that standard errors of area-level model estimates may be underestimated, resulting in inflated Type-I error. Conclusions regarding area-level covariates estimated from models exhibiting statistically significant spatial autocorrelation of area-level residuals should be interpreted cautiously.

Software

Data management and multilevel modeling were conducted in SAS using PROC GLIMMIX (SAS Institute Inc., version 9.3, Cary, North Carolina). SPSS was utilized for PCA analysis (IBM Corp., IBM SPSS, version 20.0.0, Armonk, NY). Spatial data management, visualization and LISA calculations took place in ArcGIS (Environmental Systems Research Institute, Inc. ArcMap version 10.0, Redlands, California). R statistical software (R Development Core Team, version 2.14.0, Vienna, Austria) with the Spatstat package (version 1.30-0) were utilized for probability density and Ripley’s K-function calculations.
Research Aim 4: Pathways from Socioeconomic and Biobehavioral Factors to Cervical Abnormalities Among Women in Appalachia Ohio

**Research aim 4:** To further test theoretical framework developed in aim one and tested in aim two by investigating the relationships between cervical abnormalities (a necessary ICCI precursor), biologic and behavioral ICCI risk factors, and socioeconomic and sociopolitical risk factors of both ICCI and ICCI-related factors among a sample of women residing in Appalachia Ohio.

**Study Sample**

**Study Design:** The CARE Project is a Center for Population Health and Health Disparities, funded by the National Cancer Institute (CA105632, 2003-present). The purpose of CARE I (2003-2007) was to utilize transdisciplinary approaches to address disparities of cervical cancer incidence among females of Appalachia Ohio. Data used in this analysis are from CARE I project 3, a case-control study investigating the social, behavioral, and biologic variables that may contribute to increased risk of abnormal cervical cytology.

A two-stage sampling scheme was utilized to achieve a sample of females that would allow estimation of the effects of area-level socioenvironmental (e.g., county rurality) and individual-level biobehavioral and socioeconomic (e.g., tobacco use, age at coitarche [age of first sexual intercourse], educational attainment) factors on abnormal cervical cytology. At the time of this study’s conception, Ohio Appalachia consisted of 29 counties (all present day counties but Ashtabula, Trumbull, and Mahoning) with females being recruited only from these 29 counties. Participants were sampled using a stratified sampling plan. Each of the 29 counties was placed into one of four contiguous
regional groupings (regions 1, 2, 3, and 4 contained 6, 7, 9, and 7 counties, respectively). Counties within regions were classified as urban or rural according to their membership to a U.S. Census Bureau designated metropolitan statistical area (MSA). All urban counties were included while two rural counties were randomly chosen within each of the regions to enable a heterogeneous sample of females with respect to urban/rural residence.

A census of clinics within these counties was generated (N=171 counties). A total of 22 clinics met the following inclusion criteria: performed Pap smear tests, reported at least 200 Pap screenings per month, and served a demographically diverse population. These clinics were approached and asked to participate in the study of which 14 agreed (63.6%). An additional three clinics that were approached following enrollment of the initial 14 also participated in the study.

Patient listings were gathered from each of the 17 enrolled clinics and females were randomly selected on a monthly basis. Study personnel reviewed medical records of 6028 randomly selected females to deem eligibility. Females were eligible if at least 18 years of age, a resident of Appalachia Ohio, not pregnant, no history of hysterectomy or invasive cervical cancer, and seen within a participating clinic within the last two years. Possibly eligible participants were approached and asked to participate in person by a study nurse on the day of the female’s routinely scheduled pelvic examination. Of the 4043 total eligible females, 2394 agreed to participate (59.2%). Participants were informed of the study, signed a written informed consent, completed a self-administered questionnaire, and provided blood and saliva samples before undergoing the pelvic examination.
The self-administered questionnaire gathered all demographic, socioeconomic, behavioral, and biologic information considered for inclusion in the current study. After completion of the self-administered survey, participants proceeded to undergo the routine pelvic examination and Pap smear. Two additional cervical tissue samples (ThinPrep: Hologic, Inc., Bedford, MA and STM: QIAGEN Inc., Valencia, CA) were collected either for storage within the CARE biological specimens repository (ThinPrep) or for HPV typing (STM). Pap smear results of each participant were reported using the Bethesda system. Pap smear results were broadly classified into three categories: 1) negative for intraepithelial lesion or malignancy (NILM), 2) epithelial cell abnormalities, and 3) other. Results deemed as representing epithelial cell abnormalities were further interpreted and classified into cell types (i.e., squamous or glandular). Squamous cell abnormalities, the most common type, were classified into the following: 1) atypical squamous cells (ASC), which are often further classified and presented as, of undetermined significance (ASC-US), or cannot exclude high-grade squamous intraepithelial lesion (ASC-H); 2) low-grade squamous intraepithelial lesion (LSIL); 3) high-grade squamous intraepithelial lesion (HSIL); and 4) squamous cell carcinoma. The additional cervical sample was assayed for HPV typing (AMPLICOR: Roche, Indianapolis, IN) at the Centers for Disease Control and Prevention (CDC).

A total of 284 females presented with epithelial cell abnormalities leaving 2081 with normal Pap smear tests (29 had an unsatisfactory or missing sample). Females with normal Pap smear tests were matched to those with abnormal Pap smear tests based on clinic site and date of Pap smear (within a three month window) and randomly selected to achieve a three to one, normal (n=847) to abnormal Pap smear test ratio. HPV typing was
conducted for these 1131 females and classified into: 1) HPV negative, 2) high risk (HR)-HPV positive (types 16, 18, 31, 33, 39, 45, 51, 52, 56, 58, 59, 68, 73, or 82), and 3) low risk (LR)-HPV positive (any other type).

Participants’ address was extracted from medical record and geocoded to the point-level. Of all participants for which a mailing address was provided, 66 (5.8%) were post office box only, and five (0.4%) addresses could not be assigned a latitude or longitude due to mismatches with address location methods.

**Measures:** Relevant demographic data included age, race (“White”, “Black or African American”, “American Indian or Alaskan Native”, “Asian”, “Native Hawaiian/Other Pacific Islander”, “Mixed”, and “Other”), and marital status (“Married”, “Divorced”, “Widowed”, “Separated”, “Never been married”, “Member of unmarried couple”, and “Don’t know”). Age was centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. All race categories besides “White” were collapsed to create a race dichotomy of “White” and “Non-White”. Three marital status categories were created through the following combinations: 1) “Married” and “Member of unmarried couple”, 2) “Divorced”, “Widowed”, and “Separated”, and 3) “Never been married”.

Relevant socioeconomic data included previous year’s total household income (“Less than $10,000”, “$10,001 to $15,000”, “$15,001 to $20,000”, “$20,001 to $25,000”, “$25,001 to $35,000”, “$35,001 to $50,000”, “Over $50,001”, “Don’t know”, and “I prefer not to answer”) and educational attainment (“None”, “Preschool/Kindergarten”, “1st grade”, … , “12th grade”, “GED”, “Some college, technical or trade school, but less than 1 year and no degree”, “1 or more years of college, 111
technical or trade school, but not degree”, “technical or trade degree or certification”, “Associate degree”, “Bachelor’s degree”, “Master’s degree”, “Doctorate degree or other advanced degree”, and “Don’t know”). Income categories were combined to create the following groups: 1) $20,000 or less, 2) $20,001-$50,000, and 3) at least $50,001. Education was centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. Non-responses or responses of “Don’t know” or “I prefer not to answer” for either socioeconomic-related question was treated as missing.

Current cigarette smoking (“Former smoker”, “Current smoker”, “Never smoker”, and “Don’t know”) and various sexual health questions were used to characterize participants’ relevant behaviors that may affect risk of abnormal cervical cytology. Sexual health data included age at coitarche, total number of male partners with whom a female had sexual intercourse within the previous 12 months (recent male partners), ever used condoms (“Yes”, “No”, “Don’t know”, and “I prefer not to answer”), and previous history of an abnormal Pap smear (“Yes”, “No”, and “Don’t know”). Age at coitarche was centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. Due to major normality violations number of recent male partners was grouped to create the following categories: 1) zero partners, 2) one partner, 3) two partners, and 4) more than two partners (maximum of twelve). Non-responses or responses of “Don’t know” or “I prefer not to answer” were treated as missing.

Area-level covariate data were gathered from the 2000 Census of Population and Housing (Summary File 4) of the U.S. Census Bureau (census tract)\textsuperscript{28} and 2004 property tax data of the Ohio Department of Taxation (school-district level)\textsuperscript{34,35}. Census tract-level variables were: median household income (median of household incomes, 1999); percent
unemployed (percent of total civilian population of at least 16 years in the labor force whom are not employed, 2000); percent impoverished (percent of total population with incomes less than 100% the federal poverty level, 1999); percent without a high school diploma (percent of population of at least 25 years without a high school diploma or equivalent, 2000); and percent without a bachelor’s degree (percent of population of at least 25 years without a 4-year college degree, 2000). The school district-level variable was taxable land value per pupil (value of all property subject to taxation and after application of assessment rates ÷ average daily school membership). All six area-level covariates were centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. Centering and standardization was conducted only for those census tracts and school districts represented in the sample (i.e., omitting those regions that did not have a study participant residing within). Due to heavy skewness log transformations were used for percent unemployed, percent impoverished, percent without a high school diploma, and taxable land value per pupil.

Ohio’s census tract (year 2000) and unified school district (year 2012) spatial boundary files were gathered from the U.S. Census Bureau’s TIGER program. A series of one to one spatial join operations enabled the linkage of area- to individual-level data through the point-level geocoded participant’s address. Eighty-seven of 615 school districts, and 201 of Ohio’s 2940 census tracts were represented in the spatially joined dataset.

Latent Variables

Seven latent variables were constructed from one or more measurement variables and used in structural modeling: 1) taxable land value, 2) area-level deprivation, 3)
individual-level SEP, 4) risky sexual behavior, 5) HPV infection and persistence, 6) tobacco use, and 7) age (Table 3.1). Taxable land value per a pupil served as the sole indicator of the taxable land value latent variable. Area-level deprivation was indicated by the five census tract-level covariates listed above. Previous year’s total household income and educational attainment were indicators of individual-level SEP. Choice of risky sexual behavior indicators was partially informed by previous work conducted by

<table>
<thead>
<tr>
<th>Latent Variable</th>
<th>Measured Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxable Land Value (School district)</td>
<td>School district level taxable land value per a pupil (1990)</td>
</tr>
<tr>
<td>Area-level Deprivation (Census tract)</td>
<td>Median household income (median of household incomes, 1999); percent unemployed (percent of civilian population of at least 16 years in the labor force whom are not employed, 2000); percent impoverished (percent of population with incomes less than 100% the federal poverty level, 1999); percent without a high school diploma (percent of population of at least 25 years without a high school diploma or equivalent, 2000); and percent without a bachelor’s degree (percent of population of at least 25 years without a 4-year college degree, 2000)</td>
</tr>
<tr>
<td>Individual-level socioeconomic position</td>
<td>Previous year’s total household income and educational attainment</td>
</tr>
<tr>
<td>Risky Sexual Behavior</td>
<td>Age at coitarche, total number of male partners with whom a study participant had sexual intercourse within the previous 12 months, ever used condoms and marital status</td>
</tr>
<tr>
<td>Human Papillomavirus Infection and Persistence</td>
<td>Previous abnormal Pap smear test and current human papillomavirus status</td>
</tr>
<tr>
<td>Tobacco Use</td>
<td>Current cigarette smoking</td>
</tr>
<tr>
<td>Age</td>
<td>Current age</td>
</tr>
</tbody>
</table>

Table 3.1. Latent variables and associated measured variables in a structural equation model of cervical cancer-related factors
Reiter, et al357 and included: number of recent male sexual partners, condom use, age at coitarche, and marital status. HPV infection and persistence was indicated by previous abnormal Pap smear test and current HPV infection status. Current cigarette smoking was the sole indicator of tobacco use. Current age was the sole indicator of age.

Statistical Modeling

Descriptive statistics (means and standard deviations for continuous variables and totals and percents for categorical variables) were computed for all measured variables. Structural equation modeling (SEM) was used to estimate the relationships of Figure 3.4 and test how well the data fit the hypothesized model. Because SEM techniques allow examination of relationships between multiple independent and dependent variables, both direct and indirect effects were investigated. Relationships were reported as standardized coefficients (i.e., interpreted as standard deviation changes). Relationship statistical significance was assessed through calculation of scaled z-tests (accounting for non-normality). Type-I error was held at 0.05. The total variation of cervical dysplasia explained by the latent variables was estimated and reported.

Similar to linear regression, SEM requires that many assumptions be true, including: multivariate normality and absence of outliers; relationship linearity; absence of multicollinearity and singularity; and residual normality, homoskedasticity and independence358. The categorical natures of several measured variables of the hypothesized model complicated the examination of these assumptions. Multicollinearity and singularity were assessed though examination of the determinant of the model correlation matrix; small determinants indicating assumption violation. It was necessary
that all categorical variables represented ordered categories of an underlying latent continuous distribution (i.e., “No HPV”, “LR-HPV”, and “HR-HPV” are categorized as zero, one, and two from an unobserved distribution of HPV types that continuously vary with respect to risk of cervical abnormalities). All categorical variables were ordered and could be said to have been grouped from a continuous distribution except marriage. It was assumed that marital status would be coded by the following: “Never been married” = zero (highest risk for cervical abnormalities), “Divorced”, “Separated”, or “Widowed” =
one (moderate risk for cervical abnormalities), and "Married" and "Member of an unmarried couple" = two (lowest risk for cervical abnormalities).

Due to the use of categorical variables, model estimation was performed using the correlation matrix. Due to the expected non-normality of the data, scaled Satorra-Bentler maximum likelihood estimation techniques were used to estimate parameter standard errors. Accordingly, model fit was assessed using the Satorra-Bentler scaled chi-squared, comparative fit index (CFI), and root mean-square error of approximation (RMSEA). As in other goodness of fit indices, the chi-square test statistic measures the degree to which the observed correlation matrix differs from that of the models, and thus a nonsignificant chi square test is desired. Because the p-value of the chi-square test statistic is dependent on the sample size, the CFI and RMSEA will be the primary indices used to determine fit. Acceptable fit is indicated by a CFI value greater than 0.92 and RMSEA value less than 0.05.

**Software**

Data management and descriptive statistics were computed using SAS (SAS Institute Inc., version 9.3, Cary, North Carolina). SEM analyses were conducted using EQS (Multivariate Software, Inc., version 6.2, Encino, California).

**Summary**

These methods were developed to directly address the lack of knowledge concerning relationships of persistent socioeconomic and ICCI disparities within Central Appalachia, and will be used to integrate and empirically test relationships between the causes of socioeconomic disparities with the causes of ICCI disparities within Central Appalachia. First, a thorough and multidisciplinary literature review will be used to
theoretically integrate current knowledge of the causes of socioeconomic disparities with the causes of ICCI disparities within Central Appalachia. Second, multilevel logistic regression models, spatial analysis methods, and maps will be used to describe the multilevel and spatial characteristic of ICCI in Ohio. Finally, structural equation modeling will be used to investigate relationships between sociopolitical, economic, and biobehavioral factors together with the prevalence of cervical abnormalities. Together, these methods may yield results that provide insight on the factors or processes that contribute to disparities of ICCI within Central Appalachia.
CHAPTER 4: AN ECOSOCIAL PERSPECTIVE OF DISPARITIES OF INVASIVE CERVICAL CANCER INCIDENCE IN CENTRAL APPALACHIA

Abstract

**Background:** Cervical cancer disparities have persisted in the Central Appalachian region of the United States for decades despite identification of several risk factors that account for and the development of efficacious interventions to reduce these disparities. Empirical evidence links invasive cervical cancer incidence (ICCI) to lower socioeconomic conditions – a historically rooted economic characteristic of Central Appalachia. The goal of this research is to integrate current knowledge of the causes of socioeconomic disparities with the causes of ICCI disparities within Central Appalachia.

**Methods:** *Ecosocial* theory and its associated social, political, economic and cultural processes and factors that explain gradients of population health will be used to gain a more complete understanding of ICCI disparities in Central Appalachia. Literature searches of various databases, directories, and websites was conducted to integrate the current understanding of the processes contributing to both socioeconomic and ICCI disparities within Central Appalachia.

**Results:** Relationships may exist between historically dependent sociopolitical (i.e., regional economic integration, sociopolitical clientelism and factionalism, land-use/land-tax policy, public education funding system), environmental (i.e., natural ecologic limits of a mountainous region, rurality), built environment (i.e., physical activity facilities and
resources), socioeconomic factors (i.e., farm-and-forest economy, community resources, individual-level socioeconomic position [SEP]), and ICCI or ICCI-related factors (i.e., human papillomavirus (HPV) infection and persistence, biologic stress, immune suppression, tobacco use, Pap smear utilization) in Central Appalachia.

**Discussion:** Explanations of factors contributing to the development or persistence of ICCI disparities that do not include explanations of socioeconomic disparities may be incomplete.

**Introduction**

Cervical cancer is a significant global, regional, and local public health problem. It is the third most commonly diagnosed cancer and fourth leading cause of cancer death among females worldwide\(^1\). The burden of cervical cancer is more pronounced in economically underdeveloped global regions with age-adjusted incidence and mortality rates of 17.8 and 9.8 per 100,000 females, respectively, ranking second among all cancer sites in these regions\(^1\). In the U.S. it is estimated that 12,170 new invasive cervical cancer cases will be diagnosed and 4,220 females will die of cervical cancer in 2012\(^2\). The 2003-2007 age-adjusted, U.S. cervical cancer incidence and mortality rates were 8.1 (ranking 13\(^{th}\)) and 2.4 (ranking 14\(^{th}\)) per 100,000 females\(^3\).

Incidence and mortality rates of cervical cancer in the U.S. have decreased tremendously since the introduction and uptake of the Papanicolaou (Pap) test in the mid twentieth century\(^4\)^5. Universal distribution of the highly efficacious human papillomavirus (HPV) vaccine allows the possibility for cervical cancer rates to be reduced by another 70%\(^6\)^7. However, the overall gains in cervical cancer prevention should not obscure the presence of demographic, geographic, historical and social
disparities. African Americans compared to whites\textsuperscript{2,8}, Hispanics compared to non-hispanics\textsuperscript{2,8} and those of lower socioeconomic position\textsuperscript{9,10} have elevated risk of ICCI. Certain regions of the U.S. – Central Appalachia, along the Mississippi River in the South, Southern Texas, and the Eastern mid-Atlantic – have reported persistently elevated ICCI rates for decades (only mortality rates were reported prior to the 1990s)\textsuperscript{11,12}.

The Appalachian Regional Commission, a federally-funded regional economic development agency\textsuperscript{360}, has defined five Appalachian subregions based on shared topography, demographics, and economics\textsuperscript{361}. The North Central and Central regions include the Appalachian counties of Northern Tennessee, Western Virginia, all of Kentucky and West Virginia, and Southern Ohio. This area (hereafter referred to as Central Appalachia) encompasses nearly all of Appalachia that has reported a historically high ICCI burden and will be the region of focus for this report.

Central Appalachia has been the subject of much social, political and environmental attention over the past century. The region was known to have lagging economic and social indicators since the late nineteenth and early twentieth centuries\textsuperscript{13-17}. Parts of Franklin D. Roosevelt’s New Deal initiatives (c. 1933) were directed towards Appalachia\textsuperscript{17}(p. 278, 312, 316). Media attention reintroduced non-Appalachian American society to a caricature of actual Appalachian culture and deprivation in between the policies of the New Deal and Lyndon B. Johnson’s 1965 “War on Poverty”\textsuperscript{13,14} (p. 29-30). The War on Poverty had political roots in John F. Kennedy’s 1960 pursuit of the democratic presidential nominee\textsuperscript{17}(p. 339). Large-scale environmental catastrophes have been frequent occurrences to Central Appalachia (Buffalo Creek, West Virginia, 1972\textsuperscript{19}; Martin County, Kentucky, 2000\textsuperscript{20}; Cheshire, Ohio, 2002\textsuperscript{18}; and Upper Big Branch Mine,
West Virginia, 2010). This attention from varied sectors of society has prompted and been paralleled by a multitude of Appalachia-focused academic work. A theme unifying much of this literature is that Appalachian poverty – and the social, political and ecologic processes contributing to poverty – are fundamental to understanding much of the region’s past and present conditions.

The rich literature investigating Central Appalachia’s historical poverty along with the extensive biomedical research conducted on ICCI etiology and natural history enables the unique possibility to formulate a comprehensive theory of the causes of ICCI disparities in Central Appalachia. Conceptualizing ICCI as a biologic consequence of the social, economic, and political milieu acting throughout history within Central Appalachia provides insight into the conditions initiating and maintaining such a health disparity. The primary purpose of this paper is to integrate current understanding of ICCI risks and pathways with political ecology theories of Central Appalachian underdevelopment and poverty. The resulting synthesis may help inform the social, political, economic and cultural processes contributing to persistent ICCI disparities. The goal is to reposition the view of ICCI disparities from the idea that they are, “…marker(s) for low access to health care in poor communities” as this may be an oversimplification. Indeed, some counties with high rates of Pap test uptake exhibit high ICCI burden while some counties with low Pap test rates have low ICCI rates. A more comprehensive and integrative framework could provide insight on the relationships between commonly identified and less-considered factors and processes, resulting in a deeper understanding of the causes and persistence of this disparity.
Biomedical Approach to Invasive Cervical Cancer Incidence Risk and Disparity Research

Biomedical research related to ICCI has largely focused on the identification of associated biologic, behavioral, and lifestyle factors\textsuperscript{6,85,87,362}. The importance of this biomedical research cannot be understated as highly efficacious preventive and therapeutic measures have resulted: a test to detect pre-malignant lesions of the cervix (i.e., Pap test); a vaccine preventing ICCI; and surgical, chemotherapeutic, and radiotherapeutic treatment interventions. Several risk factors have also been identified: HPV infection (a necessary cause), tobacco use, parity, oral contraceptive use, infection with other sexually transmitted infections (STI), and immune suppression\textsuperscript{6,85,87,362}. This research has greatly contributed to the absolute reduction of ICCI, morbidity and mortality. Epidemiologic research concerned with ICCI disparity reduction follows similarly from that pursuing reduction in the absolute burden of ICCI.

Epidemiologic research regarding ICCI disparities necessarily begins with identification of sociodemographic or geographic subpopulations with elevated risks of disease. Research addressing the existence of a disparity may borrow from work focused on reducing the absolute burden of ICCI by hypothesizing about the subpopulation’s disparate prevalence of factors related to ICCI (i.e., increased prevalence of tobacco use, increased risky sexual behavior, decreased healthcare access, etc). Socioeconomic indicators associated with ICCI disparities (e.g., educational attainment, income, employment status, etc) may be considered initiators of the pathways to ICCI. Interventions may target these subpopulations with the goal of impacting risk-reduction differentially compared to subpopulations of lower-risk. What are intervened upon are the bio-behavioral or sociodemographic factors contributing to ICCI – tobacco use, Pap test
and HPV vaccine uptake, knowledge of these factors, etc – that may mediate the socioeconomic-ICCI relationship.

Explanation of the pathways linking SEP and disease clarifies how these factors may contribute to ICCI disparities. The association between SEP and the myriad of adverse health conditions has been demonstrated by others\textsuperscript{250,268,269}. Measures of SEP (e.g., individual-level education and income, area-level poverty rate, etc) have been specifically linked to cervical cancer incidence\textsuperscript{275,276,300,302} and Pap smear utilization\textsuperscript{235,242-246,363}. Recent research suggests that the SEP – cancer relationship may be mediated by physiologic stress and reduced immune capacity; those of lower SEP may have increased markers of stress\textsuperscript{277-283}. Those of lower socioeconomic levels may have decreased immune functioning compared to those of higher socioeconomic levels\textsuperscript{249,286,287,289}. Physiologic stress is thought to increase risk of ICCI by: increasing cellular susceptibility to viral infection, promoting viral replication, increasing tumor growth or suppressing host immune system defenses\textsuperscript{278,283}. The historically rooted sociopolitical factors and processes contributing to the socioeconomic and bio-behavioral factors are largely ignored in traditional ICCI disparity research.

An Ecosocial Perspective of Population Health Disparities

Krieger’s Ecosocial theory of disease distribution was first described in a 1994 paper calling for epidemiologists to frame studies with more robust theories\textsuperscript{255}. Ecosocial theory serves as basis for determining, “who and what drives current and changing patterns of social inequalities in health”\textsuperscript{313}(p. 672). Ecosocial theory incorporates elements of the socially-oriented theories social production of disease\textsuperscript{256,314}, political economy of health\textsuperscript{256,315}, and social determinants of health\textsuperscript{250,269}, while additionally emphasizing the
importance of historical context and ecology\textsuperscript{303,313}. Nearly identical to the theory of political ecology invoked by ecologists and human geographers (see King\textsuperscript{316}), Ecosocial theory posits that population patterns of health are the consequences of historically interacting social, economic and political processes\textsuperscript{303,313}. Emphasizing processes and relationships as opposed to static independent factors, interactions between phenomena should be dynamically considered across spatiotemporal scales. Processes of production, consumption, reproduction and exchange are central to describing relationships between differential social, economic and political arrangements of power and property.

Four key constructs theoretically ground the design, conduct, analysis and conclusions drawn from investigations employing Ecosocial theory: 1) embodiment; 2) pathways of embodiment; 3) cumulative interplay of exposure, susceptibility and resistance; and 4) accountability and agency. Embodiment is the biologic manifestation of an individual's interactions with the biophysical and social world. That which is embodied is a health condition of interest (e.g., ICCI, diabetes, stress, etc). Embodiment stipulates that contributors of health disparities originate mostly from sources external to the individual, and ‘…cannot be reduced to allegedly ‘innate’ characteristics…’\textsuperscript{303}(p. 215). Static disease mechanisms or fundamental causes cannot account for cause-specific population patterns of disease as phenomena are context-dependent. The framing of and possible contributors to the specific health condition are directed by how societal phenomena become embodied.

Pathways of embodiment are the interacting relationships between given factors and processes that contribute to the embodied consequence. Specific health disparities may have multiple antecedent pathways and these pathways should be jointly detailed.
The various pathways producing a specific health disparity can be generalized and categorized as involving: 1) economic and social deprivation, 2) toxic substances, pathogens, and hazardous conditions; 3) discrimination and other forms of socially inflicted trauma; 4) targeted marketing of harmful commodities; 5) inadequate or degrading health care; and 6) degradation of ecosystems\(^\text{303}(p.223)\).

The third *Ecosocial* construct – cumulative interplay between exposure, susceptibility, and resistance – highlights the importance of timing, accumulation of and responses to embodied exposures. Krieger admits to the complexity of Ecosocial theory and the impossibility of incorporating all possible pathways along with their context-contingent relationships between exposure, susceptibility, and resistance into a single investigation\(^\text{303}\). This limitation of causal theorizing accords with philosophical debates contending that complete and perfect explanation of causal relationships may be impossible pragmatically, if not theoretically\(^76\). The idea of *Ecosocial* theory’s first three constructs is to more accurately and comprehensively characterize contributors of health disparities\(^\text{303}(p.225)\).

The last construct involving accountability and agency explicitly identifies by *who* and *what*, and to *what extent* disparities are created and propagated in a population, as well as why the disparity is studied and how it is explained. Agency refers to the ability to act, while accountability describes the degree of responsibility for actions taken and avoided\(^313\). Coinciding with the *social production of disease* and *political economy of health*, this fourth construct requires a reflexivity of epidemiologic theories and theorists. Reflexivity, as described in social theory, is referring to the recursive relationships between researchers and that which is researched\(^\text{317} (p.255)\).
Central Appalachian Cervical Cancer Disparities from an Ecosocial Perspective


Over half of all Appalachian counties were economically “distressed” (a relative indicator of economic deprivation) in 1960\(^50\). That percentage has decreased to 27.6% in 2010\(^364\). Despite this reduction the spatiotemporal pattern of distressed counties is highly ordered and has become increasingly concentrated in the Central Appalachian regions with 76% of distressed counties in 2000 also classified as distressed in four of the five previous decennial censuses\(^50\).

Evidence suggests that women from Central Appalachia may have increased exposure, increased susceptibility and decreased resistance to conditions related to socioeconomic well-being, HPV infection and ICCI. The percent of the workforce employed within the manufacturing industry is overrepresented in Central Appalachia\(^48\). Research has concluded that overreliance on the manufacturing industry, undiversified employment opportunities and neoliberal trade policies (e.g., North American Free Trade Association) are acting together to increase unemployment susceptibility within several local economies of Central Appalachia; including many of the same areas already reporting elevated unemployment rates \(^51,53,54\). Employment status has been shown to predict health insurance coverage and healthcare utilization – both are decreased in Central Appalachia\(^49,67\). Regions of Central Appalachia have higher poverty rates\(^39,50\), higher unemployment rates\(^43,50\), lower educational attainment\(^43\), and lower incomes\(^43,49\). Some of these lagging economic indicators of present-day have been shown to be related to cervical cancer incidence\(^275,276,300,302\).
The persistently depressed socioeconomic conditions of Central Appalachia may also contribute to increased unhealthy behaviors related to ICCI\textsuperscript{69,365,366}. Independent surveys provide evidence that Central Appalachian residents are at an increased risk of high psychological distress\textsuperscript{69,133}. Exercise may be one of the best ways to increase resistance to stress as well as strengthen one’s immune system\textsuperscript{286}. Females of Central Appalachia report a markedly higher prevalence of recent physical inactivity compared to non-Appalachian women\textsuperscript{65}. Moreover, prevalence of cigarette use in Appalachia is highest in Central Appalachia and a higher prevalence of cigarette use is reported by females of the Appalachian regions of Kentucky and Ohio compared to non-Appalachian females\textsuperscript{49,65,66,69}. Cigarette smoke has demonstrated immunosuppressive properties as well as the ability to decrease the local cervical environment’s resistance to HPV infections\textsuperscript{158,159}.

The highest county ICCI rates of the Central Appalachian states tend to concentrate in Southeastern Kentucky, Southeastern Ohio and Central West Virginia (Figure 4.1). Moreover, socioeconomic deprivation and healthcare access demonstrate similar trends within the region (Figures 4.2 and 4.3)\textsuperscript{12,33}. Despite the strong relationships between county-level socioeconomic deprivation, healthcare access and cervical cancer these maps do not provide explanation for why and how these relationships exist nor who or what is responsible for their production and reproduction. Operationalizing Ecosocial’s four constructs will address these questions and yield a fuller description of the factors responsible for the creation and persistence of ICCI disparities in Central Appalachia.
Figure 4.1. Age-adjusted cervical cancer incidence rates by Kentucky, Ohio, and West Virginia counties, 1996-2008

Figure 4.2. Estimated distribution of socioeconomic deprivation within Kentucky, Ohio, and West Virginia counties (data from University of Wisconsin’s Population Health Institute)\textsuperscript{33}. 

129
Reconceptualization of Invasive Cervical Cancer Incidence: A Biologic Consequence of Socioecologic Disparities

Applying Ecosocial’s first construct of embodiment to ICCI in Central Appalachia involves considering the physiologic presentation of cervical cancer and all biologically relevant ICCI antecedents (e.g., establishment of HPV infection, persistence and neoplastic progression; suppressed immune system; decisions to use tobacco; etc). It is necessary to consider the possible social, political and ecologic contexts in which ICCI may become biologically embodied; as opposed to a conceptualization that eschews external forces and only details ICCI development within an individual. Attention is then placed on determining why females of Central Appalachia biologically incorporate their life experiences in such a way that results in invasive malignancies of the cervix (or HPV
infection, tobacco use, etc) in comparative disproportions to females not residing in this region. Despite being diagnosed within individuals, this perspective shifts the possible causes and responsibility for the disease away from the individual and towards processes that operate largely external to the individual being diagnosed.

Interdependent Pathways Connecting Invasive Cervical Cancer Incidence to Ecologic, Social, Economic and Political Historical Conditions within Central Appalachia

Detailing of the multiple intersecting pathways of embodiment along with the cumulative interplay of exposure, resistance, and susceptibility to ICCI in Central Appalachia focuses on describing the possible causal links between ICCI and social, political, economic and cultural processes acting throughout history. These relationships suggest that contemporary ICCI disparities of Central Appalachia may be the product of individual-and regional socioeconomic deprivation. The relationships between SEP and biologic ICCI-related factors suggest that explanations of ICCI disparities may be incomplete when explanations of socioeconomic disparities are excluded. These socioeconomic disparities are not assumed to develop ex nihilo, but rather as a result of interacting historical processes. Specifically, Central Appalachian socioeconomic disparities may be causally linked to interdependencies between a patriarchal-based farm-and-forest economy and culture, integration into surrounding market-based economies, economic under-regulation, ecologic limits of the mountainous region, land-use policies, political corruption and public quiescence of this corruption (Figure 4.4). These relationships are described in detail below.
Figure 4.4. A detailed representation of the relationships involved in an Ecosocial explanation of disparities of invasive cervical cancer incidence in Central Appalachian

Historically common and influential industries in Appalachia included mining, agriculture and forestry. The predominance of these ‘extractive’ industries provides indication of the richness of Appalachia’s natural resources. The development of these industries of 19th and 20th century Appalachia have been described as “colonial” because natural and human resources were exploited by outside capital. The Appalachian agriculture economy of the mid to late 19th century was highly socially stratified. At the top of the hierarchy were a minority of wealthy individuals who owned large tracts of farmland but also generated extra income through land speculation. These individuals ascended to local political power and frequently married among others of
high sociopolitical status who also owned and managed land. Smaller farmers, tenants, and slaves comprised the lower levels of this agriculturally-based social hierarchy. Late 19th century industrialization gave way to the mining and timber industries. Large portions of land were sold off to absentee corporations centralizing operations in larger urban centers (e.g., Pittsburgh, Cleveland, etc), who then imported timber and coal resources from adjacent Appalachian regions.

The relevant roots of the relationships between extractive industries, large-scale absentee land ownership, culture, ecosystem degradation and local socioeconomic disparities have been traced to Appalachia’s earliest days of colonial settlement in the late 17th century. Appalachia was colonial America’s first “western frontier” to be explored. Large-scale land speculation was conducted by wealthy families with ties to European royalty (e.g., Penn, Culvert, Fairfax, and Granville). Land licensing was concentrated in those with power - usually wealthy land owners in Eastern seaboard counties. Small land speculators or settlers had more difficult time acquiring land or licenses for land17. Land was commoditized on a large-scale as treasury and military land warrants were issued to officers in exchange for service in the American revolutionary war. The increased Appalachian land claims continuing into the late 18th century was accompanied by unregulated land surveying and haphazard land title documentation. Williams suggests that these early economic actions prefaced the Appalachian residents transfer of “…ownership of their land or rights to the timber on and minerals under it to nonresident corporations in the industrial era”17(p. 75). It was also during the 18th century when large land-holdings were first acquired by companies.
The successful Appalachian economy of the early to mid 19th century has been described as a farm-and-forest economy. Many families resided on subsistence farms. Hogs and cattle were open-range raised, foraging throughout the rich Appalachian forests. The forest ecology began to deteriorate in the early 19th century with natural increases in population density and livestock overgrazing. Land squatters were being pushed further into Appalachian mountain hollows and branch hollows as land nearest to coves, valleys, and rivers was most desirable to owners and speculators for both agricultural and commercial purposes. Subsequent clearing of the steeper land for farming purposes promoted erosion. Ironmaking and saltmaking were important industries within mid 19th century Appalachia. Both altered local ecology as they required large amounts of water, timber and coal.

The family-based culture and economy characterizing some within Central Appalachia has been attributed to both increasing susceptibility and resistance to poverty in the 19th and 20th centuries. Sociologist James S. Brown engaged several families of the Clay county, Kentucky area in an ethnographic study during the early 1940s (i.e., “Beech Creek study”). Brown recorded two pervasive cultural types within the region; ‘modern’ and ‘traditional’. The families adopting modern values (i.e., rational, purposeful and optimistic) were often more socioeconomically successful. Those who were ‘traditional’ displayed qualities of familialism (strict patriarchal family structure), fundamentalism (biblically-based code of conduct) and fatalism (passive acceptance of one’s societal position). Brown also noted that families of lower socioeconomic success were less likely to have descended from original Beech Creek settlers.
Refining Brown’s earlier work, Billings and Blee put forth the idea of a
*patriarchal moral economy of agriculture* to further explain the interdependencies of
culture, economy, and social relationships within Eastern Kentucky. They contend that
the farm-and-forest economy in the context of environmental constraints, a patriarchal
culture and capitalism within a neoliberal political system, were the seeds of local
poverty within Eastern Kentucky. Non-market based cultural strategies that influenced
family and regional economics were identified as familism and traditionalism. Familism
describes the cultural ethos that included kin-based relations of economic and social life
and a psychological sense of affectional kin ties. Three sets of social relations were
embedded in familism: 1) life course transitions were patterned by patriarchal family
strategy, not individual choice; 2) economic production was sustained through patrilineal
property transmission and cooperative labor among males; and 3) family networks were
linked in economic exchange and distribution. In these networks, adult males had the role
of redistribution of surplus to less fortunate families. The traditionalism previously
described by Brown was critiqued for its biased interpretation though the lens of a
capitalist system that promotes individualism, achievement motivation and universalism.
In contrast, the qualities ascribed to Brown’s familialism were demonstrated to offer
resistance to the influx of capitalist workplace control and authority; familism promoted
egalitarian norms. These kinship-based, egalitarian cultural ties allowed families of
Eastern Kentucky’s farm-and-forest economy to reproduce this culture and delay poverty
for a period of time.

Poverty encroached upon Central Appalachia with increased population density,
railroad activity and integration of local elites into the commodity chains. Increased
population density pressured the patriarchal moral economy as land was further divided into smaller tracts, disallowing subsistence farming. The increased railroad activity for the purpose of resource export contributed to increased land prices nearest to the lines. Propertied local elites vetted railroad companies to attract rail line construction as real estate became a lucrative industry\textsuperscript{17}. Many subsistence farmers only had physical labor to offer the increasingly market-based economy. The large pool of labor entering the new markets was leveraged by employers in the form of lower employee wages. Moreover, Central Appalachian miners earned lower wages than non-Appalachian miners, in part, as a strategy of offsetting increased costs incurred from shipping raw product large distances to non-Appalachian refineries\textsuperscript{17}. Transition to a wage-based labor economy allowed local elites to increase economic and political standing within Central Appalachia.

Counties – dubbed “little kingdoms” in Kentucky\textsuperscript{17(p. 192)} – were self-contained political units in 19\textsuperscript{th} century Central Appalachia\textsuperscript{17}. County governments controlled taxes, laws, ordinances and licenses. Men living part- or full-time within the county seat dominated political influence despite the county court’s membership being divided among leading families of county districts. Members of these families specialized in State or Federal offices, the circuit courts, or were lawyers or merchants. The power that Kentucky passed to its counties created fierce competition between sociopolitical elites of those counties. The sensationalized and oft misinterpreted Hatfield-McCoy feud originated in the Hatfield’s increased land acquisition and logging activities of Eastern Kentucky. The Hatfield’s increased economic activities were in direct response to the rising market value of timber and mineral resources resulting from the area’s connection to the railroads and fuller integration into the national economy. Contrary to 19\textsuperscript{th} and 20\textsuperscript{th}
century popular culture (see Algeo), feuding often took place between those of higher socioeconomic standing as opposed to between “degenerate hillbillies.” Indeed, the 19th century Central Appalachian land oligarchy can be traced from these prominent families (i.e., Hatfield, Floyd, McCoy, Vance, Rutherford and Chafin) back to original land owners of the 17th century.

Many mid to late 19th century Central Appalachian local economies were still barter and exchange based. The decrease in land value following the 1873 economic crisis led to decreased values of mortgage debts. This had differential affects on the local economy. Local elites who managed the bartering and exchange were more likely to sue those whose debt they held, forcing the indebted to auction their land. Billings and Blee demonstrate an inverse association between national prosperity and frequency of these lawsuits. Wealth accumulation by a few proceeded by defaults on debts allowing the auctioning of land at prices below market value. The land would then be purchased by wealthy land owners. This local exchange system produced little possibility to accumulate local capital. Despite the knowledge of local natural resource value, local developers had to seek outside capital to develop the infrastructure needed to extract the resources.

As previously noted in the Hatfield-McCoy feud, local elites attempted to capitalize on the rich timber and mineral resources of the region. During the 18th century trees were cut to clear land for agriculture and settlement. Timber exports were first conducted at this time when local entrepreneurs bought trees for milling. This process was later expanded and fully controlled by large timber corporations. Ensuing large-scale timbering resulted in forest fires and soil erosion. All but a few hundred of 10,000,000
acres of West Virginia’s old growth forest was gone by 1920\textsuperscript{17}. Mineral rights were purchased by outside company representatives and scouts during the 19\textsuperscript{th} century\textsuperscript{14}. Oftentimes, land owners were given 25\% of their land’s total value for the mineral rights with many land owners negotiating and receiving better rates. Absentee corporations did not always honor the terms of agreement for payment to land owners. Those affected by the broken contracts did not differ by SEP, however, individuals’ ability to legally battle the corporations did vary by SEP resulting in fewer payments to less wealthy landowners\textsuperscript{14}.

Local elites did not have as much control over the local economy as did outsiders who owned huge portions of mineral rights. Regardless, local elites still had more power relative to local farmers and land tenants. The transition period between the farm-and-forest and wage-based, natural resource economies left local economies undiversified and over-dependent on a few industries\textsuperscript{14,17,38,39}. This increased the dependency of many poorer residents on local elites for their economic livelihood\textsuperscript{14,15}. The economic transformation towards the natural resource economies exacerbated preexisting antagonistic relations among elites. Political economy researchers cite this increase in factionalism among the local elite due to changing economic times as a major contributor to the 20\textsuperscript{th} century political corruption that pervaded Eastern Kentucky\textsuperscript{14,15}.

The power wielded by local elites served to economically and politically subjugate non-elite individuals. This dependency and development ofclientelism (the exchange of goods or services for political support) between elites and non-elites divided communities based on whom they supported\textsuperscript{14}. These sociopolitical developments left a legacy of tying the political system to local elites, contributing to political corruption\textsuperscript{14,15}. 

138
Billings and Blee provide numerous examples demonstrating the persistence of these corrupt political relationships into the late 20th century. In analyzing his Eastern Kentucky ethnographies of the 1980s, the intersection of cultural kinship ties with clientelism and associated economic and political effects is noted by Allen Batteau, “Political knowledge in this community consists of knowing these networks, knowing how to do somebody a favor, and knowing how to get a favor done….Such favors, or ‘accommodations,’ include help in getting jobs, introductions to politicians, help in getting aid from government programs, or transportation to the hospital.”

Political sociologist John Gaventa expounds on these relationships in Central Appalachia while developing his community-based framework for identifying and addressing sociopolitical inequities.

John Gaventa’s *Power and Powerlessness: Quiescence and Rebellion in an Appalachian Valley* further notes the adverse effects placed on the community as a result of relationships between mining industries and politicians. Gaventa empirically demonstrates how coal company-affiliated and business-elite political candidates obtained the most votes regardless of party affiliation and contrary to the welfare of voters. As the book title suggests, Gaventa describes how historically-rooted, Central Appalachian power differentials between elite and non-elites can lead to quiescence and how this relationship is fostered by those in power. Analyses during the 1970’s have shown how large portions of Central Appalachian land were owned by mining companies. These mining lands were taxed at lower rates compared to privately owned land, resulting in decreased local government funds and decreased per capita government expenditures. This form of local government’s tax-based funding source is a
main reason cited by several plaintiffs in lawsuits brought against respective U.S. States for not providing equal educational opportunities across all areas (including Ohio, West Virginia, Kentucky, and Tennessee).\footnote{368}

In summary, several historically and spatiotemporally-specific, interdependent pathways contributing to persistent socioeconomic disparities among residents of Central Appalachian have been described. Present-day individual-level and regional socioeconomic disparities were borne out of or hastened by interactions of a patriarchal-based farm-and-forest local economy and culture, integration into the market-based national economy and under-regulation (i.e., land acquisition, employee compensation and lack of environmental protection). Despite the unsustainability of the patriarchal local economy and culture, it is important to note the resistance to poverty conferred by these cultural traits. Local barter and exchange economies in combination with national prosperity cycles allowed local elites to accrue wealth at the expense of the less wealthy. Further integration into the larger regional and national economies in the late 19\textsuperscript{th} and early 20\textsuperscript{th} centuries served to exacerbate socioeconomic disparities by inequitably benefiting those with power and privilege. Similarly, 20\textsuperscript{th} century Eastern Kentucky political corruption, public complacency with this corruption and resulting decreased government expenditures has been traced to processes of transition to wage-based economies, undiversified local economies, cultural kinship ties and clientelism. As disparities of ICCI strongly follow socioeconomic gradients, these interdependent processes and factors may also be responsible for ICCI disparities with women of Central Appalachia.
Discussion and Conclusions

The goal of this research was to illuminate relationships between historically-rooted social, economic, and political factors and processes, leading to a more comprehensive understanding of the creation and persistence of ICCI disparities in Central Appalachia. The integrative relationships with and pathways to socioeconomic deprivation and ICCI detailed here deepens the understanding of not only persistence, but by who and what are responsible for the disparities within Central Appalachia. Just as the development of socioeconomic disparities may be less the result of individual action and choice and more the product of sociopolitical, economic and ecologic processes playing out over the region’s history, then the same may be true for disparities of ICCI among women residing in Central Appalachia. This conceptualization of ICCI disparities and the resulting accountability placed on policies, economic conditions, cultural traits and those with large landholdings directs responsibility away from Central Appalachian females diagnosed with ICCI compared to a conceptualization that ignores these external processes.

Availability of highly efficacious preventive measures such as the HPV vaccine and Pap smear, could lead to the conclusion that the benefit of such complex ICCI conceptualization is dubious. Establishing causation, as has been suggested, may only be important for identifying a practical point at which to successfully intervene\textsuperscript{369}. Therefore, it may follow that the development of a vaccine that prevents the necessary cause of ICCI undermines the need for comprehensive theories that may identify other possible points of intervention. However, the very existence of salient points of healthcare intervention such as those preventing ICCI may be contributing to cancer
disparity production\textsuperscript{370,371}. A disparity may become apparent when these points of intervention become unequally distributed along recognizable societal gradients (i.e., Pap smear test and income or education). Further support for this hypothesis can be found in recent research suggesting that adolescent females of Central Appalachia have lower HPV vaccine uptake compared to the rest of the United States as well among all other Appalachian regions\textsuperscript{372}. Therefore, concerns of future ICCI disparity persistence may be best addressed through a focus on achieving equitable distribution \textit{and uptake} of the HPV vaccine and Pap smear.

Evidence demonstrating that the order of causation is directed from social, political, economic and cultural factors to individuals should be considered in any intervention or policy that may have implications for ICCI. Connecting various economic, political and cultural processes and factors with socioeconomic deprivation and ICCI in specific locales leads to the conclusion that each pathway’s importance – and most practical \textit{way} of intervention – may be dependent on several factors. It should not be assumed, for example, that the interdependent effects of local Eastern Kentucky politics, culture and socioeconomics on HPV vaccine uptake, Pap smear use and tobacco cessation will be the same in Southern Ohio. Determining the currently most important factors affecting local populations may require local knowledge, suggesting a need for community- and grassroots-based interventions and healthcare\textsuperscript{373-375}(p. 453-468). Disregard of local factors and processes could have unanticipated consequences to which a reflexive \textit{Ecosocial} theory cautions.

The unique perspective associated with an \textit{Ecosocial} conceptualization of a disparity may lead to conclusions that differ from more focused and less integrative
theories. The Appalachian trans-generational transmission of wealth and clout has been traced back to several original late 17th century land owners. Knowledge of this lineage has been documented since as late as Brown’s Beech Creek (1940’s)” but, ironically, undue accountability and agency have oftentimes been placed on impoverished Appalachians. A “culture of poverty” dominated explanations of Central Appalachians’ poverty until the 1970’s when political economy theorizing began to identify the responsibility attributable to other factors and processes. Despite an abundance of literature identifying structural processes producing Appalachian socioeconomic deprivation, authors of a 2012 economic strategy report commissioned by the ARC interpreted Pike County, Kentucky’s high percentage of residents receiving social security disability income (SSI) as, “…indicative of a culture of welfare reliance. Intergenerational welfare dependence is common and eligibility for benefits is almost regarded as a rite of passage for some families.” While this SSI disparity within Pike County, Kentucky is likely to be correct, it is interpretation of the underlying causes that ultimately direct the possible interventions of this disparity that are open to debate. Whether this disparity is believed to results from a sub-cultural feeling of entitlement acting within the region or due to larger influences such as differences in prevailing work type (e.g., labor-intensive versus professional occupations), employment availability or educational and job-training opportunities may have important implications for addressing the disparity.

The possible advantages of an integrative conceptualization offered by Ecosocial theory are associated with drawbacks. The level of detail and comprehensive information required – not only within epidemiology but crossing into other scientific disciplines –
seems to make *Ecosocial* theory intractable and recondite. An obvious and common criticism of *Ecosocial* theory is the magnitude of physical time needed to frame an epidemiologic issue as archival research is usually required\(^{303}(p.\ 233),^{376}(p.\ 214)\). The extra benefit to population health and health disparities attributable to the application of *Ecosocial* or similar theories may determine their viability into the future\(^{316}\). Directly addressing their complex characteristics, *Ecosocial* theory can be viewed as the most well positioned theoretical framework capable of explaining the causes and persistence of health disparities, as well as guiding interventions to address them.

Cervical cancer disparities have persisted in Central Appalachia for as long as cancer statistics have been collected. Reorienting ICCI from a largely biobehavioral occurrence to one that is the result of an individual’s embodiment of certain social, political and economic processes historically interacting is a large shift from traditional epidemiologic research. This conceptualization has not only underscored the responsibility attributable to processes and factors external to females diagnosed with ICCI, but the susceptibility and resistance conferred to individual and population health by constant integration into larger societal systems. Increased recognition of these contributors may be critical to reducing the disparities of ICCI as well as many other problems of population health.
CHAPTER 5: MULTILEVEL AND SPATIAL CHARACTERISTICS OF INVASIVE CERVICAL CANCER INCIDENCE IN OHIO

Abstract

Background: Despite wide-spread geographic disparities, few studies investigate area-level factors that may contribute to disparities of invasive cervical cancer incidence (ICCI). This study investigates the multilevel and spatial variation of ICCI in Ohio.

Methods: Individual-level demographics and ICCI data were from the Ohio Incidence Surveillance System. Cases were residents of Ohio diagnosed with ICCI from 1996-2009. Controls were sampled from a synthetic dataset representative of the Ohio female population. Socioeconomic, sociopolitical, built environment, and bio-behavioral data was gathered for census tract- school district- and county-levels from various Federal and State sources. Hierarchical logistic regression models were used to investigate and explain variation of ICCI at individual- and area-levels. Spatial data analysis techniques described the spatial variation of ICCI at each level and quantified the degree and location of clustering.

Results: In all models, individual-level age and race interact such that African Americans (compared to Whites) younger than 39 years have reduced odds of ICC, while African Americans older than 53 have an increased odds. County-level teen birth rate and percent without health insurance, and school district-level childhood poverty are associated with increased odds of ICC, while county-level Chlamydia rate, school district-level taxable
land value per a pupil, and census tract-level socioeconomic position (SEP) are associated with decreased odds of ICC. Model covariates accounted for area-level global spatial clustering of county- and census tract-level models. The spatial distribution of the school district-level source of variation was considerably changed with adjustment for model covariates. Percent of ICCI variation possibly due to county-, school district-, or census tract-level factors was 1.6%, 3.2%, and 4.3%, respectively.

**Discussion:** Despite large geographic disparities, a small percentage of the total ICCI variation in Ohio is possibly due to county-, school district-, or census tract-level factors. Taxable land value per a pupil, a present-day manifestation of historical sociopolitical processes, was inversely associated with ICC. These results may be used by public health practitioners to target interventions, by cancer researchers to develop future studies, or by those wishing to address disparities through policy action.

**Introduction**

Cervical cancer incidence and mortality rates have steadily decreased in the U.S. since the 1950s with uptake of the Papanicolaou (Pap) test. Universal distribution of the human papillomavirus (HPV) vaccine allows the possibility for cervical cancer rates to be substantially reduced further yet. Pre-HPV vaccine research has identified several factors that may contribute to ICCI risk, including: tobacco use, parity, oral contraceptive use, earlier age at coitarche (age of first sexual intercourse), infection with other sexually transmitted infections (STI), and immune suppression.

Despite progress that has led to large reductions in overall cervical cancer burden in the U.S., disparities of ICCI and related factors still exist. It has been known for decades that the burden of ICCI is patterned along distinct socioeconomic and geographic
gradients. Lower education and income measured at individual- and area-levels have consistently been associated with increased cervical cancer incidence. Data from cancer registries indicate that certain regions of the U.S. – Central Appalachia, along the Mississippi River in the South, Southern Texas, and the Eastern mid-Atlantic – have reported persistently elevated ICCI rates since the mid 20th Century. Moreover, evidence continues to mount suggesting that socioeconomic and geographic factors play important roles in Pap test and HPV vaccine utilization.

Appalachia is a contiguous region of the U.S. that stretches from Northeast Mississippi to Southwestern New York. The region’s characteristic lagging economic and social indicators that prompted Federal attention in the 1960’s have greatly improved. However, much of the Appalachian regions in Kentucky, West Virginia, and Southeastern Ohio (hereafter, ‘Central Appalachia’) still exhibit persistent patterns of low income and employment and high poverty. In addition to higher ICCI and lower socioeconomic indicators, Central Appalachian residents may engage in less healthy behaviors (including higher tobacco use and lower Pap test and HPV vaccine utilization) and have decreased access to healthcare compared to those not residing in Appalachia.

A large body of literature is devoted to explaining Appalachia’s deep-rooted historical trends of lower socioeconomic factors with a special focus on the Central region. Accounts of Central Appalachia poverty often describe historical processes and factors that may have contributed to the present-day adverse conditions, including: ecologic degradation and environmental catastrophes, natural resource
and labor exploitation by external commercial interests, cultural and economic interactions, political corruption, and natural ecologic constraints of a mountainous region. This wealth of multidisciplinary, place-based research and literature presents the opportunity to examine Central Appalachian ICCI in a novel fashion.

Few studies have leveraged the extant social, economic, political and ecologic research to integratively investigate ICCI in Central Appalachia. Possibly due to the identification of HPV’s role, research on ICCI has overwhelmingly focused on individual-level effects but is less focused on neighborhood and spatial effects (exceptions noted). Despite their rarity in ICCI etiologic research, these analytic methods are well-suited and increasingly used together to investigate health disparities exhibiting spatial or neighborhood patterns.

This study aims to: 1) investigate and explain the variation of ICCI by partitioning it into individual- and various area-level sources; and 2) describe and summarize the spatial distribution of ICCI within each level. The comprehensive Ecosocial theory is used to guide variable selection, frame analyses and interpret findings. Ecosocial theory attempts to explain population-level patterns of health and disease (i.e., disparities) through critical examination of the historically-rooted and inter-related social, economic, political and ecologic contexts across spatio-temporal scales (i.e., historical and dynamic places).

The application of Ecosocial theory to disparities of ICCI in Central Appalachia has been previously detailed. Briefly, Central Appalachian socioeconomic and ICCI disparities may be causally and historically linked to interdependencies between a
patriarchal-based farm-and-forest economy and culture, integration into surrounding market-based economies, economic under-regulation, ecologic limits of the mountainous region, land-use policies, political corruption and public quiescence of this corruption. This theoretical work has identified school district taxable land value as a possible present-day manifestation of several of these interdependent historical processes. Commonly investigated factors affecting ICCI will be analyzed at common area-levels, including: individual-level age and race; and county- or census tract-level SEP (income, education, poverty and unemployment), tobacco use, risky sexual behavior indicators, healthcare access indicators, and rurality. Less commonly analyzed factors or area-levels that were chosen under the theoretical framework include: county-level physical activity indicators and industry-specific employment indicators; census tract-level manufacturing industry employment; and school district-level poverty, education expenditure indicators, and taxable land value per a pupil. When the data permitted, variable year was also selected under the theoretical framework.

**Methods**

**Study Sample**

**Study Design**

Invasive cervical cancer cases were from Ohio’s population-based state cancer registry, the Ohio Cancer Incidence Surveillance System (OCISS)\textsuperscript{26}. Females diagnosed with invasive cervical cancer (International Classification of Disease – Oncology codes C530-C539\textsuperscript{223}) from 1996 to 2009 were considered for inclusion (N=7124). Race, age at diagnosis and geographic coordinates of home address at diagnosis were also gathered from OCISS. Nine cases were excluded due to nonsensical or missing geographic
coordinates resulting in 7115 cases. A control sample was generated from an extant synthetic U.S. geospatial database\textsuperscript{27}. A total of 14,230 controls were randomly selected to achieve a 1 to 2 case to control ratio (21,345 cases and controls total). Each control’s estimated age, race and geographic coordinates of home address were also gathered from the synthetic dataset. The final individual-level dataset consisted of Ohio invasive cervical cancer cases and simulated controls all containing complete information on age, race and geographic coordinates of home address.

Measures

Area-level covariate data were gathered from the 2000 Census of Population and Housing (Summary File 4)\textsuperscript{28} and 1995 Small-Area Income and Poverty Estimates Program of the U.S. Census Bureau\textsuperscript{29}; the 2000 Rural-Urban Commuting Codes\textsuperscript{30}, 2003 Rural-Urban Continuum Codes\textsuperscript{31}, and 2004 County Typology Codes\textsuperscript{32} of the U.S. Department of Agriculture; the 2010 and 2012 County Health Rankings and Roadmaps Program of the University of Wisconsin’s Population Health Institute and the Robert Wood Johnson Foundation\textsuperscript{33}; 1990 and 2004 property tax data of the Ohio Department of Taxation\textsuperscript{34,35}; and 1994 and 2004 Local Report Card data of the Ohio Department of Education\textsuperscript{36} (see Appendix A for years of measurement, dataset, data source, description, classification, and geographic level of support for area-level variables).

County-level variables considered in modeling were: tobacco smoking (percent of adults identifying as current smokers and reporting at least 100 lifetime cigarettes, 2002-2008); Chlamydia rate (Chlamydia rate per 100,000 population, 2007); teen birth rate (teen birth rate per 1,000 female population, 15-19 years, 2000-2006); percent without health insurance (percent of the population less than 65 years without health insurance, 2006).
2005); primary care provider rate (primary care provider rate per 100,000 population, 2006); income (composite of income inequality at the household level, 2007, and percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 2005-2007); high school graduation rate (percent of 9th grade cohort that graduates in 4 years, 2005-2006); percent with a bachelor’s degree (percent of population of at least 25 years with 4-year college degrees or higher, 2005-2007); family and social support (composite of percent of single-parent households, 2005-2007, and percent of adults without social or emotional support, 2005-2008); childhood poverty (percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 2007); physical inactivity (percent of adults reporting no leisure time physical activity, 2009); recreational facility rate (recreational facilities per 100,000 population, 2009); urban (Rural-Urban Continuum Code, 2003); high mining employment (income earnings from mining industry is at least 15% of all income earnings, 2004); and high manufacturing employment (income earnings from manufacturing industry is at least 25% of all income earnings, 2004).

School district-level variables considered in modeling were: childhood poverty (percent of children living in families with incomes less than 100% the federal poverty level, 5-17 years, 1995); pupil support expenditures per pupil (pupil support services, guidance services, health services, psychological services, speech pathology, audiology services, social activities, work activities, and student instructional media services ÷ average daily membership, 1994 and 2004); instructional expenditures per pupil (salaries of teachers, teacher aides, or paraprofessionals, as well as materials, computers, books and other consumables used with students within the classroom ÷ average daily
membership, 1994 and 2004); total expenditures per pupil (administration, building operations support, staff support, pupil support, and instructional support ÷ average daily membership, 1994 and 2004); and taxable land value per pupil (value of all property subject to taxation and after application of assessment rates ÷ average daily membership, 1990 and 2004).

Census tract-level variables considered in modeling were: high mobility (Rural-Urban Commuting Codes, 2000); commute time (mean travel time to work among workers of at least 16 years, 2000); percent employed in manufacturing (percent of total civilian population of at least 16 years employed within the manufacturing industry, 2000); median household income (median of household incomes, 2000); percent unemployed (percent of total civilian population of at least 16 years in the labor force whom are not employed, 2000); percent impoverished (percent of total population with incomes less than 100% the federal poverty level, 2000); percent without a high school diploma (percent of population of at least 25 years without a high school diploma or equivalent, 2000); and percent without a bachelor’s degree (percent of population of at least 25 years with 4-year college degrees or higher, 2000).

Ohio’s county subdivision (year 2000), unified school district (year 2012), and census tract (year 2000) spatial boundary files were gathered from the U.S. Census Bureau’s TIGER program. A series of one to one spatial join operations enabled the linkage of area- to individual-level data. All of Ohio’s 88 counties, 611 of 615 school districts, and 2861 of 2940 census tracts were represented in the spatially joined dataset (i.e., the difference represent areas registering zero cases or controls).
Statistical Analysis

Variable Classification

Transformations, reclassifications or composites of both individual- and area-level variables were created prior to analyses. Individual-level race was collapsed into ‘White’ (N=18,184), ‘African American’ (N=2,506), and ‘Other’ (N=655). Appropriate age categorizations were investigated in response to previous research suggesting that age and race interact to affect ICCI risk\(^3,8,72\). Age- and race-specific (white or African American only) crude ICCI rates calculated using 19 age groups suggested age categories of: less than 39 years, between 39 years and 53 years, and greater than 53 years (Figure 3.3). All subsequent statistical model building included age, race and an age-race multiplicative interaction term.

Except for urban, high mining employment and high manufacturing employment all county-level variables were standardized and modeled as continuous Z-scores. County-level Rural-Urban Continuum codes were dichotomized into code=1 (part of a metropolitan statistical area with at least 1,000,000 population) versus codes 2 through 9 (all counties not part of a metropolitan statistical area with at least 1,000,000 population) to create the county variable ‘urban’. County-level high percent of mining employment and high percent of manufacturing employment were dichotomous by definition (see above or Appendix B). All school district-level variables were standardized and modeled as continuous Z-scores. Except for Rural-Urban Commuting Area codes, all census tract-level variables were standardized and modeled as continuous Z-scores. Rural-Urban Commuting Area codes were dichotomized into codes less than or equal to 5 (all metropolitan, micropolitan core, and micropolitan non-core/high commuting) and code=8
(small town high commuting) versus codes 5.1-7.4 and 8.1-10.6 (decreased commuting to urban areas or commuting from non-metropolitan origins) to create the census tract variable ‘high commuting’. Principal Component Analysis (PCA) of five variables commonly used to indicate area-based socioeconomic constructs – median household income, percent unemployed, percent impoverished, percent without a high school diploma, and percent without a bachelor’s degree – was utilized to reduce data dimensionality and possibility of collinearity between census tract-level variables. PCA analysis proceeded using the data correlation matrix and only components with Eigenvalues greater than 1.0 were considered for inclusion in models. The standardized components resulting from the PCA were used in subsequent modeling.

**Statistical Modeling**

Multilevel logistic regression models were developed to estimate associations between ICCI and covariates, explain variation in ICCI and to partition the odds of ICCI into individual- and area-level sources of variation. This partitioning, through the use of area-specific random intercepts, enabled the geographic investigation of individual- and area-level patterns in the odds of ICCI through the creation of maps. Cases and controls were assumed to follow a binomial distribution while the area-level random intercepts were assumed to follow a normal distribution. The probability of ICCI was related to the explanatory variables through the logit link function (see Appendix C for model notation). A compound symmetry random effects covariance structure was used in all models.

Model building was conducted separately for each of the three geographic levels. This approach allowed for comparisons of area-level variance across geographic levels.
Explanatory models were built in a backwards elimination fashion to yield a parsimonious model that explained the most variation possible in ICCI. The individual-level, fixed-effects covariates age, race, and an age-race multiplicative interaction term were included in all model building. Two-way cross-level interactions were judged for model inclusion by entering multiplicative interaction terms between any fixed-effects individual- and area-level main effect into the model. Those fixed-effect interactions associated with statistically significant type-III tests were included in the final model.

Sensitivity analyses were conducted using each of the three final models described above. Models calibrated using the entire Ohio dataset were stratified by Appalachian residence. This enabled the comparison of relationships between factors and ICCI throughout the State (the full dataset) to those in Appalachia, Ohio. Differences between estimates of the two models may suggest a possible interaction between that factor, Appalachian residence, and ICCI. Only results of models restricted to Appalachian residence are presented.

Exponentiation of any fixed-effect model coefficient yielded an odds ratio (OR). Odds ratios and 95% confidence intervals (CI) were calculated for all fixed-effect covariates of the final model. A statistical test of whether the log-OR = 0 was conducted and assessed using a t-test. Stratum-specific ORs and 95% CIs were reported for those covariates involved in an interaction. Odds ratios of covariates modeled as Z-scores were back calculated and reported in terms of the data and not the Z-score. Main-effect ICCI ORs estimated from Z-score variables were reported as a one standard deviation (SD) change in those variables. The original data of these Z-score variables were analyzed to ensure that a one SD change was not beyond the range of the data.
Changes in the level-specific distributions of ICCI were investigated through analyses of individual-level residuals and random intercepts estimated from multilevel null and final models. Multilevel null models are those only including an overall intercept and area-specific random intercepts (i.e., a model excluding explanatory variables). The residual variation of the null model serves as a baseline to which the residual variation of a final model can be compared. Any changes in the residual spatial distribution between null and final models of the same geographic-level can be attributed to the covariates within the final model.

Various items of information were reported from each of the two types (i.e., null and final) by three geographic-level explicit models (six sets of results). Model fit was estimated from the quantity two times the negative log-likelihood, conditional on the random intercepts (-2LLR|RI). A random intercept variance parameter, $\tau^2$, was estimated and used in calculations of the intraclass correlation coefficient (ICC). Random intercept variance and ICCs can be used to quantify the variation attributable to both individual- and area-level processes. The latent response method was applied in calculating the ICCs; the individual-level variance of the models was assumed to equal the constant, $\pi^2/3$. Therefore, model ICC was equal to, $\tau^2 / (\tau^2 + \pi^2/3)$. A statistical test of whether $\tau^2=0$ was conducted and assessed using a Chi-square mixture distribution. Tests of $\tau^2=0$ yielding a sufficiently high Chi-square value provide evidence that a model including random intercepts is better than one without. A statistical test of whether a given area’s associated random intercept=0 was conducted for all areas of each model and assessed using a t-test. Those tests yielding sufficiently high t-values provide
evidence that a given area’s ICCI odds were significantly different from the overall odds of the entire State.

Spatial Analysis

Basic exploratory spatial data analyses (ESDA) techniques were used to characterize the spatial pattern of ICCI in Ohio\textsuperscript{345,352,353}. A nonparametric estimate of the spatially-varying relative risk was calculated to produce a fine-grain, yet stable surface of individual-level ICCI probabilities\textsuperscript{353}. These probabilities were visualized in a map and represent the degree to which ICCI risk varies across the State. Details of the methods used to produce the surface are described elsewhere\textsuperscript{353}.

A global summary of the inter-point spatial dependence, or global clustering, of ICCI cases and controls was estimated using Ripley’s K-function\textsuperscript{352-354}. Details of the methods used in calculating the K-function are described elsewhere\textsuperscript{352-354}. In this application, the K-function is providing an estimate of the distances at which the density of cervical cancer cases may be statistically different than the density of controls. Evidence of statistically significant deviations between case and control density was assessed visually using significance bands. Significance bands represent the K-function null values at which to compare the K-function values estimated from the observed data.

In addition to characterizing any spatial clustering of ICCI in Ohio, the distance at which cervical cancer cases significantly clustered was used to inform subsequent definitions of a ‘neighborhood’ in later analyses.

Spatial patterns of individual-level residuals (i.e., Pearson-type residuals of the six multilevel models described above) were investigated the same way as the original case-
control data. Probability surfaces were calculated from the residuals and results were visualized in maps.

Spatial patterns of area-level residuals (i.e., random intercepts from the six multilevel models described above) were investigated through map visualization and calculation of the global and local version of Moran’s I\textsuperscript{355,356}. The global Moran’s I provides a single summary measure of the spatial clustering of an area-level variable (e.g., area-level random intercepts) throughout the entire study region (e.g., Ohio). A positive Moran’s I index indicates global clustering of area-level values, a negative index indicate global regularity or ‘repulsion’, and an index=0 indicates no global clustering\textsuperscript{356}. A statistical test of whether the index is significantly different from ‘0’ is also used. The local Moran’s I, or local indicator of spatial autocorrelation (LISA) measures the local “….similarity between each region’s associated value…and those of nearby regions”\textsuperscript{345}(p.237). Calculation of LISA statistics necessitates defining nearby regions or ‘neighborhoods’. The choice of neighborhood definition may have substantial influence on LISA results\textsuperscript{345}. As a sensitivity analysis, two different neighborhood definitions were used in each of the six sets of random intercepts: 1) edge and node contiguity (i.e., queen), and 2) zone of indifference (ZINDF). Queen contiguity defines a given region’s neighbors as those regions which share any border portion with that region’s border. ZINDF defines neighbors using a combination fixed distance and inverse distance weighting approach: neighbors within a region’s distance threshold contribute equal weight to the region, beyond this zone neighbors contribute a weight that is inversely proportional to the distance specified. The K-function results were used to determine the distance threshold value for ZINDF neighborhood definitions. LISA results are presented
in tables and visualized in maps (ZINDF LISA results will be presented as main findings while analyses using queen contiguity neighborhood definitions can be found in the Appendix E). It is important to note that evidence of spatial dependence within the random intercepts indicates that standard errors of area-level model estimates may be underestimated, resulting in inflated Type-I error. Conclusions regarding area-level covariates estimated from models exhibiting statistically significant spatial autocorrelation of area-level residuals should be interpreted cautiously\(^{345}\).

Data management and multilevel modeling were conducted in SAS using PROC GLIMMIX (SAS Institute Inc., version 9.3, Cary, North Carolina). SPSS was utilized for PCA analysis (IBM Corp., IBM SPSS, version 20.0.0, Armonk, NY). Spatial data management, visualization and LISA calculations took place in ArcGIS (Environmental Systems Research Institute, Inc. ArcMap version 10.0, Redlands, California). R statistical software (R Development Core Team, version 2.14.0, Vienna, Austria) with the Spatstat package (version 1.30-0) were utilized for probability density and Ripley’s K-function calculations.

**Results**

**Principal Component Analysis**

Principal component analysis of the census tract-level variables median household income, percent unemployed, percent impoverished, percent without a high school diploma, and percent without a bachelor’s degree resulted in a single component that accounted for 69.9% of the variance of the original covariates (Table 5.1). At least 72.5% of the original variance of median household income, percent impoverished and percent without a high school diploma was explained by the component. However, only 65.4%
and 54.3% of the original variance of percent unemployed and percent without a bachelor’s degree were explained, respectively. Accordingly, median household income, percent impoverished and percent without a high school diploma loaded most heavily on the component, followed by percent unemployed and percent without a bachelor’s degree (Table 5.1). The single component resulting from PCA analysis of these five variables is hereafter referred to as ‘socioeconomic position’ (SEP).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Original Variance</th>
<th>Factor Loading$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Household Income</td>
<td>0.758</td>
<td>-0.871</td>
</tr>
<tr>
<td>Percent Unemployed</td>
<td>0.654</td>
<td>0.808</td>
</tr>
<tr>
<td>Percent Impoverished</td>
<td>0.725</td>
<td>0.852</td>
</tr>
<tr>
<td>Percent without a High School Diploma</td>
<td>0.816</td>
<td>0.903</td>
</tr>
<tr>
<td>Percent without a Bachelor's Degree</td>
<td>0.543</td>
<td>0.737</td>
</tr>
</tbody>
</table>

$^1$The resulting component accounted for 69.9% of the variation of the original variables

Table 5.1. Summary results of the principal components analysis of socioeconomic-related census tract-level variables

**Multilevel Modeling**

**County**

The county null model estimated the statewide odds (95% CI) of ICCI to be to 0.52 (0.49-0.56) (Table 5.2). As a null model does not include covariates, these odds estimated the ratio of cases to controls (i.e., 7115 cases to 14230 controls). The county random intercept (RI) variance parameter of 0.0522 was significantly different than 0,
suggesting an improvement in fit for a model with county random intercepts compared to a model without. This county RI variance parameter was used to calculate an ICC of 0.0156, suggesting that 1.56% of ICCI variation within Ohio may be attributed to county-level effects.

Covariates included in the final county model were age, race, age-race interaction, Chlamydia rate, percent without health insurance, teen birth rate, and an age-teen birth rate interaction. The county final model estimated the statewide adjusted odds (95% CI) of ICCI to be to 0.37 (0.35-0.40) (hereafter, all final-model ORs were adjusted for every other model covariate and confidence intervals were at the level associated with Type-I error of 0.05; that is, 95% CIs). For the final model, this ICCI odds estimate was for those with a value of ‘0’ for all model covariates (i.e., White women, younger than 39 years residing in a county with an average Chlamydia rate, teen birth rate, and percent without health insurance). Among those younger than 39 years, the adjusted odds of ICCI among African Americans was 0.70 (95% CI: 0.59-0.83) the ICCI odds of Whites, while among those older than 53 years, African Americans had an adjusted ICCI odds that was 1.66 (95% CI: 1.43-1.93) that of Whites (no statistically significant differences comparing African Americans to Whites among those between 39 and 53 years). Among those less than 39 years, the adjusted odds (CI) of ICCI of those of other races was 0.57 (0.43-0.77) times that of Whites (no statistically significant differences comparing those of other races to Whites among those between 39 and 53 years or older than 53 years).

A county Chlamydia rate increase of 145.0 per 100,000 persons was associated with a 0.95 (CI: 0.92-0.98) change in ICCI odds. An increase of those without health insurance by 1.55 percent was associated with a 1.11 (CI: 1.06-1.17) increased odds of
<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual-level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.52</td>
<td>0.37</td>
</tr>
<tr>
<td>Age</td>
<td>0.49-0.56</td>
<td>0.35-0.40</td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.70</td>
<td>(0.59-0.83)</td>
</tr>
<tr>
<td>Other</td>
<td>0.57</td>
<td>(0.43-0.77)</td>
</tr>
<tr>
<td>At 39 years ≤ Age ≤ 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.99</td>
<td>(0.84-1.16)</td>
</tr>
<tr>
<td>Other</td>
<td>1.03</td>
<td>(0.75-1.41)</td>
</tr>
<tr>
<td>At Age &gt; 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.66</td>
<td>(1.43-1.93)</td>
</tr>
<tr>
<td>Other</td>
<td>0.87</td>
<td>(0.61-1.24)</td>
</tr>
<tr>
<td>County-level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia Rate</td>
<td>0.95</td>
<td>(0.92-0.98)</td>
</tr>
<tr>
<td>(per increase of 145.0 cases per 100,000 persons)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Uninsured</td>
<td>1.11</td>
<td>(1.06-1.17)</td>
</tr>
<tr>
<td>(per a 1.55 percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>increase of uninsured)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Birth Rate</td>
<td>1.15</td>
<td>(1.07-1.24)</td>
</tr>
<tr>
<td>(per increase of 12.9 births per 100,000 persons)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At 39 years ≤ Age ≤ 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Birth Rate</td>
<td>1.33</td>
<td>(1.24-1.43)</td>
</tr>
<tr>
<td>(per increase of 12.9 births per 100,000 persons)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At &gt; 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Birth Rate</td>
<td>1.31</td>
<td>(1.22-1.40)</td>
</tr>
<tr>
<td>(per increase of 12.9 births per 100,000 persons)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Teen Birth Rate = 50.9 per 100,000 persons and Race = White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.12</td>
<td>(1.92-2.35)</td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.69</td>
<td>(1.53-1.87)</td>
</tr>
<tr>
<td>At Teen Birth Rate = 34.1 per 100,000 persons and Race = White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>1.77</td>
<td>(1.60-1.94)</td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.44</td>
<td>(1.31-1.58)</td>
</tr>
</tbody>
</table>

Table 5.2. Effects of demographic and county-level factors on invasive cervical cancer incidence in Ohio, 1996-2009

cont’d
Table 5.2 (cont’d)

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Odds Ratio</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.00</td>
<td>(2.40-3.75)</td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>4.02</td>
<td>(3.23-4.99)</td>
</tr>
</tbody>
</table>

At Teen Birth Rate = 34.1 per 100,000 persons and Race = African American

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Odds Ratio</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.49</td>
<td>(1.97-3.15)</td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>3.41</td>
<td>(2.71-4.27)</td>
</tr>
</tbody>
</table>

At Teen Birth Rate = 50.9 per 100,000 persons and Race = Other

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Odds Ratio</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.80</td>
<td>(2.48-5.83)</td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>2.57</td>
<td>(1.63-4.06)</td>
</tr>
</tbody>
</table>

County-level Variance  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0522</td>
</tr>
<tr>
<td>Intra-class Correlation Coefficient</td>
<td>0.0156</td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>27102.5</td>
</tr>
</tbody>
</table>

The intercept may be interpreted as the odds of cervical cancer at the value of ‘0’ for all other covariates in the model.

Odds ratio involving age is dependent on individual-level race and county-level teen birth rate variables.

Statistical significance based on a mixture of chi-square distributions.

Statistical significance at p<0.001.

Evidence of county-level residual spatial autocorrelation suggests underestimation of county-level estimate’s standard errors.

ICCI. Generally, the odds of ICCI increased with increases in county teen birth rates.

However, a county teen birth rate increase of 12.9 was associated with 1.15 (CI: 1.07-1.24), 1.33 (CI: 1.24-1.43) and 1.31 (CI: 1.22-1.40) increased odds of ICCI among those younger than 39 years, between 39 and 53 years, and older than 53 years, respectively.

Odds ratios comparing different age groups at levels of race and teen birth rate (rates associated with the rank-ordered 25th and 75th percentiles) are presented due to the statistically significant age-race interaction and teen birth rate-age interaction. Compared
to those younger than 39 years, those between 39 and 53 years as well as those older than 53 years had statistically significantly elevated odds of ICCI at each of the race by teen birth rate levels. However, odds ratios of those between 39 and 53 years or those older than 53 years (versus younger than 39 years) were higher in counties with higher teen birth rates. For example, among African Americans living in a county with a teen birth rate of 50.9, those between 39 and 53 years had an odds of ICCI that was 3.00 (2.40-3.75) times that of those younger than 39 years. However, the same age-ICCI comparison among African Americans living in a county with a teen birth rate of 34.1 resulted in a smaller odds ratio; 2.49 (1.97-3.15).

The county final model RI variance parameter of 0.0135 was significantly different than 0, suggesting an improvement in model fit compared to a model without county RIs. The county RI variance parameter was used to calculate an ICC of 0.0041. This ICC suggested that, after adjustment for individual- and county-level covariates included in the final model, 0.41% of ICCI odds within Ohio may be attributed to additional county-level effects.

Restricting the data to only those residing in Appalachia, Ohio and refitting the county final model described above had little effect on the estimates (Appendix D). Likely the result of decreased power, several of the estimates found to be statistically significant using the full dataset became statistically insignificant when using the dataset restricted to those residing in Appalachia, Ohio.

School District

The school district null model estimated the statewide odds (CI) of ICCI to be 0.47 (0.45-0.49). The school district RI variance parameter of 0.1098 was significantly
different than 0, suggesting an improvement in fit for models with school district random
intercepts compared to a model without. This school district RI variance parameter was
used to calculate an ICC of 0.0323, suggesting that 3.23% of ICCI odds within Ohio were
attributed to school district-level effects.

Covariates included in the final school district model were age, race, age-race
interaction, percent of impoverished children, total taxable land value per a pupil (2004),
and a race-percent of impoverished children interaction (Table 5.3). The school district
final model estimated the statewide odds (95% CI) of ICCI to be to 0.32 (0.30-0.34). For
the final model, this ICCI odds estimate was for those with a value of ‘0’ for all model
covariates (i.e., White women, younger than 39 years residing in a school district with an
average percent of impoverished children and taxable land value per a pupil). Compared
to those younger than 39 years, those between 39 and 53 years as well as those older than
53 years had statistically significantly elevated odds of ICCI at each level of race.
However, the odds ratios (95% CI) comparing those between 39 and 53 years to those
younger than 39 years varied by race; OR_{39 and 53 years|White} = 1.98 (1.82-2.14), OR_{39 and 53
years|African American} = 2.83 (2.27-3.54), OR_{39 and 53 years|Other} = 3.33 (2.17-5.10). Odds ratios
(95% CI) comparing those older than 53 years to those less than 39 years also varied by
race; OR_{>53 years|White} = 1.60 (1.48-1.73), OR_{>53|African American} = 3.79 (3.05-4.70), OR_{>53|Other
= 2.31 (1.46-3.65).

A school district taxable land value (2004) increase of $176,401 per pupil was
associated with a 0.86 (95% CI: 0.75-0.98) change in ICCI odds. Generally, the odds of
ICCI increased with increases in school district childhood poverty rates. However, an 8.2
percentage point increase of childhood poverty was associated with 1.23 (CI: 1.18-1.28),
1.17 (CI: 1.09-1.25), and 0.95 (CI: 0.84-1.08) change in odds of ICCI among Whites, African Americans and those of other races, respectively. Odds ratios comparing races at levels of age and percent of childhood poverty (percents associated with the rank-ordered 25th and 75th percentiles) were presented due to the statistically significant age-race interaction and percent impoverished children-race interaction. Within each level of age, African Americans had consistent odds of ICCI (versus Whites) across levels of childhood poverty: among those younger than 39 years, OR_{Black|15.7\% Childhood Poverty} = 0.56 (0.46-0.68) and OR_{Black|5.9\% Childhood Poverty} = 0.59 (0.47-0.76); among those between 39 and 53 years, OR_{Black|15.7\% Childhood Poverty} = 0.80 (0.67-0.96) and OR_{Black|5.9\% Childhood Poverty} = 0.85 (0.68-1.07); and among those older than 53 years, OR_{Black|15.7\% Childhood Poverty} = 1.33 (1.11-1.58) and OR_{Black|5.9\% Childhood Poverty} = 1.41 (1.13-1.76). Within each level of age, those of other races, in contrast, had odds of ICCI (versus White) that varied across levels of childhood poverty: among those younger than 39 years, OR_{Other|15.7\% Childhood Poverty} = 0.56 (0.42-0.76) and OR_{Other|5.9\% Childhood Poverty} = 0.77 (0.56-1.06); among those between 39 and 53 years, OR_{Other|15.7\% Childhood Poverty} = 0.96 (0.70-1.32) and OR_{Other|5.9\% Childhood Poverty} = 1.29 (0.92-1.82); and among those older than 53 years, OR_{Other|15.7\% Childhood Poverty} = 0.82 (0.58-1.18) and OR_{Other|5.9\% Childhood Poverty} = 1.11 (0.75-1.64).

The school district final model RI variance parameter of 0.0505 was significantly different than 0, suggesting an improvement in model fit compared to a model without school district RIs. The school district RI variance parameter was used to calculate an ICC of 0.0151. This ICC suggests that, after adjustment for individual- and school district-level covariates included in the final model, 1.51% of ICCI odds within Ohio may be attributed to additional school district-level effects.
Restricting the data to only those residing in Appalachia, Ohio and refitting the school district final model described above had notable effects on some of the estimates (Appendix D). Contrary to the results from the full dataset, there was no evidence that taxable land value per a pupil was related ICCI among those residing in Appalachia, Ohio. Estimates of the odds of ICCI comparing those of the race category ‘Other’ to White, among those older than 53 years, may also vary by Appalachianian residence. Estimates of this association were statistically insignificant when using the full dataset. However, estimates using the Appalachia-restricted dataset indicate that those of the race category other have an odds of ICCI that was 0.11 (95% CI: 0.01-0.88) times that of Whites, among women older than 53 years and residing in school districts where 5.9% of children are impoverished. Similarly, among Appalachianian women older than 53 years and residing in school districts where 15.7% of children are impoverished, the odds of ICCI among those of the race category other was 0.17 (95% CI: 0.03-0.87) times that of Whites. It is important to note, however, the wide confidence intervals associated with these stratum-specific estimates.

Census Tract

The census tract null model estimated the statewide odds (95% CI) of ICCI to be to 0.49 (0.48-0.51) (Table 5.4). The census tract RI variance parameter of 0.1471 was significantly different than 0, suggesting an improvement in model fit compared to a model without census tract random intercepts. This census tract RI variance parameter was used to calculate an ICC of 0.0428, suggesting that 4.28% of ICCI odds within Ohio were attributed to census tract-level effects.
<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR 95% CI</td>
<td>OR 95% CI</td>
</tr>
<tr>
<td><strong>Individual-level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept(^a)</td>
<td>0.47 (0.45-0.49)</td>
<td>0.32 (0.30-0.34)</td>
</tr>
<tr>
<td>Race(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>1.98 (1.82-2.14)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.60 (1.48-1.73)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.83 (2.27-3.54)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>3.79 (3.05-4.70)</td>
<td></td>
</tr>
<tr>
<td>At Race= Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.33 (2.17-5.10)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>2.31 (1.46-3.65)</td>
<td></td>
</tr>
<tr>
<td><strong>School District-level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxable Land Value, 2004 (per increase of $176,401 per pupil)</td>
<td>0.86 (0.75-0.98)</td>
<td></td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Impoverished Children (per increase of 8.2 percent)</td>
<td>1.23 (1.18-1.28)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Impoverished Children (per increase of 8.2 percent)</td>
<td>1.17 (1.09-1.25)</td>
<td></td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Impoverished Children (per increase of 8.2 percent)</td>
<td>0.95 (0.84-1.08)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 15.7 and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.56 (0.46-0.68)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.56 (0.42-0.76)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 5.9 and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.59 (0.47-0.76)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.77 (0.56-1.06)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Effects of demographic and school district-level factors on invasive cervical cancer incidence in Ohio, 1996-2009

cont’d
Covariates included in the final census tract model were age, race, age-race interaction, SEP, and a race-SEP interaction. The census tract final model estimated the statewide odds (95% CI) of ICCI to be 0.37 (0.35-0.39). For the final model, this ICCI odds estimate was for those with a value of ‘0’ for all model covariates (i.e., White women, younger than 39 years residing in a census tract of average SEP). Compared to those younger than 39 years, those between 39 and 53 years as well as those older than 53

<table>
<thead>
<tr>
<th>At Percent Impoverished Children = 15.7 and 39 years ≤ Age ≤ 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.80 (0.67-0.96)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.96 (0.70-1.32)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At Percent Impoverished Children = 5.9 and 39 years ≤ Age ≤ 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.85 (0.68-1.07)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.29 (0.92-1.82)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At Percent Impoverished Children = 15.7 and Age &gt; 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.33 (1.11-1.58)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.82 (0.58-1.18)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At Percent Impoverished Children = 5.9 and Age &gt; 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.41 (1.13-1.76)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.11 (0.75-1.64)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School District-level Variance</th>
<th>0.1098</th>
<th>0.0505</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>0.0323</td>
<td>0.0151</td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>26988.8</td>
<td>26151.7</td>
</tr>
</tbody>
</table>

---


aThe intercept may be interpreted as the odds of cervical cancer at the value of ‘0’ for all other covariates in the model
bOdds ratio involving race is dependent on individual-level age and school district-level childhood poverty rate variables
cStatistical Significance based on a mixture of chi-square distributions
dp<0.001
eEvidence of school district-level residual spatial autocorrelation suggests underestimation of school district-level estimate's standard errors
<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual-level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept$^{a}$</td>
<td>0.49 (0.48-0.51)</td>
<td>0.37 (0.35-0.39)</td>
</tr>
<tr>
<td>Race$^{b}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.00 (1.84-2.17)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.60 (1.47-1.73)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.99 (2.39-3.75)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>3.95 (3.17-4.93)</td>
<td></td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.43 (2.24-5.26)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>2.40 (1.52-3.79)</td>
<td></td>
</tr>
<tr>
<td><strong>Census Tract-level Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Position (per 1 standard deviation increase)</td>
<td>0.63 (0.60-0.66)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Position (per 1 standard deviation increase)</td>
<td>0.78 (0.72-0.85)</td>
<td></td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Position (per 1 standard deviation increase)</td>
<td>0.96 (0.80-1.15)</td>
<td></td>
</tr>
<tr>
<td>At 75th % of Socioeconomic Position and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.57 (0.46-0.71)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.67 (0.49-0.91)</td>
<td></td>
</tr>
<tr>
<td>At 25th % of Socioeconomic Position and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.43 (0.36-0.52)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.71 (0.52-0.97)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4. Effects of demographic and census tract-level factors on invasive cervical cancer incidence in Ohio, 1996-2009.

cont’d
Table 5.4 (cont’d)

<table>
<thead>
<tr>
<th>At 75th % of Socioeconomic Position and 39 years ≤ Age ≤ 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.86</td>
<td>(0.71-1.04)</td>
</tr>
<tr>
<td>Other</td>
<td>1.15</td>
<td>(0.83-1.59)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At 25th % of Socioeconomic Position and 39 years ≤ Age ≤ 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.64</td>
<td>(0.55-0.76)</td>
</tr>
<tr>
<td>Other</td>
<td>1.22</td>
<td>(0.88-1.69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At 75th % of Socioeconomic Position and Age &gt; 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.42</td>
<td>(1.18-1.71)</td>
</tr>
<tr>
<td>Other</td>
<td>1.01</td>
<td>(0.70-1.45)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>At 25th % of Socioeconomic Position and Age &gt; 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.06</td>
<td>(0.91-1.25)</td>
</tr>
<tr>
<td>Other</td>
<td>1.07</td>
<td>(0.73-1.55)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Census Tract-level Variancec</th>
<th>0.1471d</th>
<th>0.07162d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>0.0428</td>
<td>0.0218</td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>27083.6</td>
<td>26090.6</td>
</tr>
</tbody>
</table>

*The intercept may be interpreted as the odds of cervical cancer at the value of ‘0’ for all other covariates in the model

*Odds ratio involving race is dependent on individual-level age and school district-level childhood poverty rate variables

*Statistical Significance based on a mixture of chi-square distributions

*p<0.001

*Evidence of census tract-level residual spatial autocorrelation suggests underestimation of census tract-level estimate’s standard errors

years had statistically significantly elevated odds of ICCI at each level of race. However, the odds ratios (95% CI) comparing those between 39 and 53 years to those younger than 39 years varied by race; OR39 and 53 years|White = 2.00 (1.84-2.17), OR39 and 53 years|African American = 2.99 (2.39-3.75), OR39 and 53 years|Other = 3.43 (2.17-5.10). Odds ratios (95% CI) comparing those older than 53 years to those less than 39 years also varied by race; OR>53 years|White = 1.60 (1.47-1.73), OR>53|African American = 3.95 (3.17-4.93), OR>53|Other = 2.40 (1.52-3.79).
Generally, the odds of ICCI decreased with increases in census tract SEP. However, a one SD increase of SEP was associated with a 0.63 (95% CI: 0.60-0.66), 0.78 (95% CI: 0.72-0.85), and 0.96 (95% CI: 0.80-1.15) change in odds of ICCI among Whites, African Americans and those of other races, respectively. Odds ratios comparing races at levels of age and SEP (scores associated with the rank-ordered 25th and 75th percentiles) were presented due to the statistically significant age-race interaction and SEP-race interaction. Within each level of age, those of other races had similar odds of ICCI (versus Whites) across levels of SEP: among those younger than 39 years, OR_{Other|75th\% SEP} = 0.67 (95% CI: 0.49-0.91) and OR_{Other|25th\% SEP} = 0.71 (95% CI:0.52-0.97); among those between 39 and 53 years, OR_{Other|75th\% SEP} = 1.15 (95% CI:0.83-1.59) and OR_{Other|25th\% SEP} = 1.22 (95% CI:0.88-1.69); and among those older than 53 years, OR_{Other|75th\% SEP} = 1.01 (95% CI:0.70-1.45) and OR_{Other|25th\% SEP} = 1.07 (95% CI:0.73-1.55). Within each level of age, African Americans, in contrast, had odds of ICCI (versus White) that may vary across levels of SEP: among those younger than 39 years, OR_{Black|75th\% SEP} = 0.57 (95% CI:0.46-0.71) and OR_{Black|25th\% SEP} = 0.43 (95% CI:0.36-0.52); among those between 39 and 53 years, OR_{Black|75th\% SEP} = 0.86 (95% CI:0.71-1.04) and OR_{Black|25th\% SEP} = 0.64 (95% CI:0.55-0.76); and among those older than 53 years, OR_{Black|75th\% SEP} = 1.42 (95% CI:1.18-1.71) and OR_{Black|25th\% SEP} = 1.06 (95% CI:0.91-1.25).

The census tract final model RI variance parameter of 0.07162 was significantly different than 0, suggesting an improvement in model fit compared to a model without census tract RIs. The census tract RI variance parameter was used to calculate an ICC of 0.0218. This ICC suggested that, after adjustment for individual- and census tract-level
covariates included in the final model, 2.18% of ICCI odds within Ohio may be attributed to additional census tract-level effects.

Restricting the data to only those residing in Appalachia, Ohio and refitting the census tract final model described above had little effect on the estimates (Appendix D). Likely the result of decreased power, several of the estimates found to be statistically significant using the full dataset became statistically insignificant when using the dataset restricted to those residing in Appalachia, Ohio.

**Spatial Analysis**

**Individual-Level**

A map of the probability surface generated from the original case-control data is displayed in Figure 5.1. All 88 counties, selected cities, and Appalachian counties are marked for reference purposes. Because cases represent exactly one-third of the case-control dataset, light green areas of the map where probability \((P) = 0.334\), provide evidence of local ICCI probabilities that do not differ from the State average. Those areas shaded yellow or beige may have higher probabilities, while those areas ranging from darker green to violet may have lower probabilities of ICCI compared to the State average. Generally, it appears as if areas of higher ICCI probability are focused in the Southeastern (Appalachia), Northwest, and West-central regions of Ohio. Areas possibly representing lower probabilities appear more localized and sporadic. Large regions of neither heightened nor lowered probability follow a general Southwest – Northeast gradient across the State. Much of the State’s Northern region also appears to be dominated by areas of ICCI probability that do not differ from the State average.
Figure 5.2 conveys identical information on the spatial distribution of ICCI probability as Figure 5.1. However, the input data used to generate Figure 5.2 are Pearson-type residuals of a single-level logistic regression (i.e., intercept only). The legend color scheme and range (-0.6 – 0.6) of the map in Figure 5.2 was used in maps

![Map of Ohio with color gradient indicating spatial variation in cervical cancer incidence probability.](image)

Figure 5.1. Spatially-varying probability of invasive cervical cancer incidence in Ohio (1:2 case:control), 1996-2009

of all subsequent logistic regression-based, individual-level, Pearson-type residuals enabling direct comparisons across maps (hereafter, all individual-level residuals resulted from a logistic regression model and are Pearson-type).

Surface probabilities of residuals estimated from county null and final models are displayed in Figures 5.3.A and 5.3.B, respectively. Comparing Figure 5.3.A to the
original spatial distribution (Figure 5.2) suggests that the individual-level residual variability is reduced (i.e., less yellow-beige and blue-violet regions) with inclusion of county-level random intercepts. As expected, the areas with average higher or lower probabilities of observed ICCI (several Southeastern counties of Figure 5.2) experience the largest absolute change in individual-level residuals with inclusion of county-level random intercepts. The only notable difference comparing residuals of the final to null models is the apparent increase in lower risk areas (i.e., more blue-violet areas) with final model covariate adjustment.

Figure 5.2. Spatially-varying probabilities of Pearson-type residuals from a null (intercept only), single-level logistic regression model of invasive cervical cancer incidence in Ohio, 1996-2009

175
Residual surface probabilities created from school district null and final models are displayed in Figures 5.4.A and 5.4.B. Individual-level variability is reduced with inclusion of school district random intercepts. The change in the spatial distribution of observed ICCI (Figure 5.2) when school district random intercepts are used appears to be similar to when county random intercepts are used (Figure 5.3.A); the areas with higher (lower) probability of ICCI experience the largest absolute change in residual ICCI probability. Also similar to the county model, individual-level residuals of the final school district model increase the number of low risk areas in the Appalachian region.

Surface probabilities of residuals estimated from census tract null and final models (Figures 5.5.A and 5.5.B) follow similar patterns to respective null and final models using county and school district random intercepts. Compared to the county and school district null models, the census tract null model may not do as well to account for the regional concentrations of higher ICCI probability in Southeastern Ohio (evidenced by the more widespread yellow-beige colors in Southeastern Ohio of Figure 5.5.A compared to Figures 5.3.A and 5.4.A). However, census tract final model covariate adjustment appears to account for the individual-level, regional ICCI trends just as well as final county and school district models (Figures 5.3.B-5.5.B). Similarly, census tract final model covariate adjustment appears to increase the frequency of low ICCI risk areas, especially in Appalachia Ohio.
Figure 5.3.A & B. Spatially-varying probability of Pearson-type residuals from county null and final multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009
Figure 5.4.A & B. Spatially-varying probability of Pearson-type residuals from school district null and final multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009
A. Individual-level residuals of census tract null model

B. Individual-level residuals of census tract final model

Figure 5.5.A & B. Spatially-varying probability of Pearson-type residuals from census tract null and final multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009
Ripley’s K-function results provide evidence of statistically significant global clustering of cervical cancer cases throughout the State (Figure 5.6). The distances of the observed function (solid line) that are higher than the upper bound of the null distribution (shaded area) are distances (≤ 12,324 meters) at which cases are in closer proximity to each other than that of controls. This suggests that cervical cancer cases are situated together more closely than controls at approximate distances of less than 12,000 meters, but situated together no more closely than controls at distances greater than 12,000 meters. Results of the K-function that seem to suggest cervical cancer case regularity or ‘repulsion’ as the observed function drops below the lower bound of the null distribution at distances greater than 27,223 meters may be an artifact and should be interpreted cautiously. These results may be more easily realized through a perspective density plot of observed ICCI probabilities (Figure 5.7). Spatial variation of ICCI probability is visually more pronounced in the perspective density plot as both color and perspective are used to convey variation. The perspective density plot is a stereoscopic version of the probability surfaces displayed in Figures 5.1 and 5.2. In terms of the perspective density plot, results of the K-function suggest that the average statistically significant ‘peaks’ of ICCI probability have average diameters of 12,000 meters. At distances greater than 12,000 meters away from the average case cases and controls are within statistically similar distances to one another.

Area-Level

A table summarizing the county random intercepts (RI) that resulted from the county null model can be found in Table 5.5. The model estimated a total of ten counties with RIs significantly different from zero. Evidence of statistically significant global and
Figure 5.6. Results of Ripley's K-function of invasive cervical cancer incidence case-control data in Ohio, 1996-2009
local spatial autocorrelation of the county-level random intercepts exists; global Moran’s $I (P) = 0.0596 \ (P<0.0001)$ and 14 counties with high (low) random intercepts that are statistically significantly clustered around neighboring counties with random intercepts of similar value.

County RIs resulting from the county null model are classified into quintiles and shaded from blue (20% of the lowest ICCI odds) to red (20% of the highest ICCI odds) in Figure 5.8. Hatched counties have odds that are statistically significantly different from zero. Stippled (dotted) counties have been identified by LISA analysis (using ZINDF
Table 5.5. Summary of spatial clustering of area-level random intercepts produced from multilevel logistic regression models of invasive cervical cancer incidence in Ohio, 1996-2009

<table>
<thead>
<tr>
<th>Level</th>
<th>Model</th>
<th>Neighborhood Definition</th>
<th>Number of R.I.(^a) P&lt;0.05</th>
<th>Moran's I P</th>
<th>Moran's I P</th>
<th>Number of LISA p&lt;0.05 Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>Null</td>
<td>Queen Contiguity</td>
<td>10</td>
<td>0.2749</td>
<td>&lt; 0.0001</td>
<td>15</td>
</tr>
<tr>
<td>County</td>
<td>Null</td>
<td>Zone of Indifference (12324m)</td>
<td>10</td>
<td>0.0596</td>
<td>&lt; 0.0001</td>
<td>14</td>
</tr>
<tr>
<td>County</td>
<td>Final</td>
<td>Queen Contiguity</td>
<td>4</td>
<td>0.0103</td>
<td>0.7331</td>
<td>4</td>
</tr>
<tr>
<td>County</td>
<td>Final</td>
<td>Zone of Indifference (12324m)</td>
<td>4</td>
<td>-0.0114</td>
<td>0.9930</td>
<td>0</td>
</tr>
<tr>
<td>School District</td>
<td>Null</td>
<td>Queen Contiguity</td>
<td>26</td>
<td>0.1471</td>
<td>&lt; 0.0001</td>
<td>48</td>
</tr>
<tr>
<td>School District</td>
<td>Null</td>
<td>Zone of Indifference (12324m)</td>
<td>26</td>
<td>0.1300</td>
<td>&lt; 0.0001</td>
<td>54</td>
</tr>
<tr>
<td>School District</td>
<td>Final</td>
<td>Queen Contiguity</td>
<td>4</td>
<td>0.0880</td>
<td>0.0004</td>
<td>36</td>
</tr>
<tr>
<td>School District</td>
<td>Final</td>
<td>Zone of Indifference (12324m)</td>
<td>4</td>
<td>0.1089</td>
<td>0.0007</td>
<td>38</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Null</td>
<td>Queen Contiguity</td>
<td>0</td>
<td>0.1277</td>
<td>&lt; 0.0001</td>
<td>228</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Null</td>
<td>Zone of Indifference (12324m)</td>
<td>0</td>
<td>0.0891</td>
<td>&lt; 0.0001</td>
<td>370</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Final</td>
<td>Queen Contiguity</td>
<td>0</td>
<td>0.0353</td>
<td>0.0023</td>
<td>138</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Final</td>
<td>Zone of Indifference (12324m)</td>
<td>0</td>
<td>0.0129</td>
<td>0.1048</td>
<td>208</td>
</tr>
</tbody>
</table>

\(^a\) R.I. = Number of area units that return statistical evidence suggesting random intercept ≠ 0

neighborhood definition) as being a statistically significantly high or low RI within a neighborhood of other high or low RIs, respectively. Random intercept values of those counties within 12,324 meters contribute equal weight to the LISA statistic calculated for the county in question. At distances greater than 12,324 RI values are weighted based on inverse distance (\(d^{-1}\)). Null model RIs of higher ICCI odds tend to spatially cluster in Southern and Eastern Ohio, while a large cluster of lower odds counties concentrate in Northeast Ohio.
Figure 5.8. County-level random intercepts of the invasive cervical cancer incidence null model
County-Level Random Intercepts of Cervical Cancer Final Model: age, race, age*race, chlamydia, teen birth, uninsured, age*teen birth

Figure 5.9. County-level random intercepts of the invasive cervical cancer incidence final model
Results in Table 5.5 and Figure 5.9 indicate that there is no evidence of global or local spatial autocorrelation of county-level RIs after adjustment for covariates in the final county model (global Moran’s I=-0.0114, P=0.993, no high or low local county clusters). However, the final county model estimated four counties with statistically significant random intercepts. The change in county RI spatial distribution from the null to final county models suggests that much of the observed county-level clustering could be explained by variables of the final county model (i.e., age, race, Chlamydia rate, teen birth rate, percent without health insurance, and interactions between age and race and age and teen birth rate).

There is strong evidence that school district RIs of the null school district model spatially cluster globally and locally (Table 5.5 and Figure 5.10). The global Moran’s I index of 0.13 is statistically significant (P<0.0001). LISA analysis indicates that 54 school districts have RIs that are locally clustered (identical neighborhood definition as described for county analyses above). The null school district model estimated 26 school districts with statistically significant random intercepts. Visual display of school district RIs in Figure 5.10 is cartographically similar to the display of county RIs; RIs are classified into quintiles, counties and Appalachia are outlined, school districts denoting statistically significant RI and LISA are hatched and stippled, respectively). In attempts to aid visualization, those school districts involved in spatial clusters are outlined electric blue. Nearly all spatially clustering, high ICCI odds RIs of the school district null model clustered in Southeastern Ohio. A few school district RIs resulting from the null model also spatially clustered within Northern and Northwestern Ohio. Significantly clustered
Figure 5.10. School district-level random intercepts of the invasive cervical cancer incidence null model
low odds school district RIs clustered in three or four groups around Southwestern, Central, and Northeastern Ohio.

Despite a reduction in school district RI measures of spatial clustering, evidence indicates that school district RIs are still clustered globally and locally after school district final model covariate adjustment (Table 5.5 and Figure 5.11). The global Moran’s I index was reduced (0.1089) but maintained statistical significance (P=0.0007). Similarly, the number of school districts involved in local spatial clusters was reduced from 54 to 38. Final model covariate adjustment has seemed to account for much of the school district-level spatial clustering previously noted in Southeastern (high odds), Southwestern (low odds) and Central Ohio (low odds). Generally, covariate adjustment did not affect the concentration of low ICCI odds school district RIs of Northeastern Ohio (Figure 5.11).

Spatial clustering analyses of census tract RIs produced from the census tract null model indicate strong evidence of spatial autocorrelation at the census tract level (Table 5.5 and Figure 5.12). The global Moran’s I of 0.0891 is statistically significant (P<0.0001) and 370 census tract RIs are considered to be statistically significantly part of local spatial clusters of similar values. However, no census tract RI was estimated to be statistically significantly different from zero. Figure 5.13 uses the same cartographic approach as described in the county and school district maps, in addition to map inserts that increase the resolution of urban areas containing small census tracts. It is difficult to discern regional RI patterns due to the large numbers (2861 Null model) and heterogenous sizes of census tracts. Clusters of high ICCI odds census tract RIs of the null model appear to concentrate in Southeastern Ohio. However, a large number of high
Figure 5.11. School district-level random intercepts of the invasive cervical cancer incidence final model
Figure 5.12. Census tract-level random intercepts of the invasive cervical cancer incidence null model
Figure 5.13. Census tract-level random intercepts of the invasive cervical cancer incidence final model
odds census tract RIs also spatially cluster in the Cleveland area (upper right inset). Also evident are groups of low odds RI clusters that occupy Northwestern, Central, and Northeastern Ohio (excluding Cleveland).

Results of Table 5.5 suggest that covariates of the census tract final model accounted for all of the census-tract level RI global spatial clustering observed in the null model (Moran’s I=0.0129, P=0.1048). However, RIs of 208 census tracts are considered to part of local clusters of similar RIs (Table 5.5 and Figure 5.13). Despite reductions in census tract RIs involved in local clusters, several high odds RIs within Southeastern Ohio remain significantly clustered. Census tract final model covariate adjustment appeared to have the effect of explaining much of the spatial clustering of low odds RIs in Southeastern and Central Ohio, while possibly increasing the spatial clustering of high odds RIs in the same areas (Figure 5.13). General patterns of low odds RIs within Northeastern Ohio (excluding Cleveland) were largely unchanged with census tract final model covariate adjustment. As in the census tract null model, no census tract RIs were found to be statistically significantly different from zero.

Discussion

This study investigated the multilevel and spatial variation of ICCI in Ohio. ICCI variation was partitioned into individual- and various area-levels (i.e., county-level, school district-level, and census tract-level). The spatial distributions of the individual- and area-levels of ICCI were described and summarized. Several findings from these analyses warrant further consideration and are discussed in detail below.
Multilevel Analysis

Though exploratory by design, several results of the multilevel analyses warrant further discussion and consideration. Individual-level age and race were found to interact in an unexpected fashion. Previous ICCI research has demonstrated similar logarithmic-shaped age-specific rates within various race/ethnic groups and other countries. Explanations of these patterns include differential screening utilization or cofactor risk profiles (i.e., tobacco use, parity, sexual behavior, etc). However, these are the first known results to demonstrate evidence of a pronounced race-specific, risk-reversal by age. African American women have substantially lower odds of ICCI compared to White women at ages younger than 39 years, no statistical difference in odds compared to White women at ages between 39 and 53 years, and substantially higher odds compared to White women at ages older than 53 years. The profile of ICCI odds among women of other races follows a similar pattern as African American women only among the younger than 39 and between 39 and 53 years age groups. Unlike African American women, women of other races do not experience a secondary spike in ICCI odds but maintain a profile statistically similar to White women. These estimates were adjusted for all other covariates in the final models (hereafter, all associations are adjusted for all other model covariates). Moreover, these results appear stable as similar relationships were observed across the three geography-specific models. Future research should focus on identifying the risk-modifying factors contributing to these relationships so that ICCI disparities by race may be properly addressed.

It is important to reiterate that caution should be exercised in concluding statistical significance of school district and census tract-level effect estimates as
evidence exists of local spatial autocorrelation among area-level residuals\textsuperscript{345}. In contrast to models involving census tract- and school district-level information, any county-level measure of SEP was not statistically significantly associated with ICCI in the final county model. One known study utilizing multilevel modeling found that county-level income, education and poverty were significantly associated with ICCI\textsuperscript{302}. However, the associations of that study were not adjusted for health insurance status, teen birth rate and chlamydia rate. County-level teen birth rate (dependently with age) and percent of adults without health insurance were significantly related to ICCI as anticipated. Teen birth rate may serve as an indicator for earlier age at coitarche while percent uninsured may be an indicator of Pap smear utilization – both are ICCI risk factors previously reported\textsuperscript{6,85,87,242,244,362}. Teen birth rate interacted with age suggesting that the increased odds of ICCI associated with a 12.9 births per 100,000 people increase among those younger than 39 years of age was a smaller increase compared to the same teen birth rate-ICCI comparison among those of the 39 to 53 and greater than 53 years age groups. It is unclear why age may modify the association between county teen birth rate and ICCI. If teen birth rate is serving as an indicator for earlier age at coitarche, then these associations may be due to the temporal decline in teen birth rate\textsuperscript{390390}. Those of older age groups would have had higher ‘exposure’ to counties of higher teen birth rates (i.e., increased probability that these older women would have engaged in sex at earlier ages) compared to women of younger age groups. This explanation also assumes: 1) that the geographic variability of year 2000 county teen birth rates is similar to that of previous decades (i.e., Ohio county teen birth rates uniformly decreased over time and did not
change rank-order position relative to one another), and 2) minimal inter-county
migration that would affect the age distribution between counties.

County-level Chlamydia rate was unexpectedly inversely associated with ICCI;
odds of ICCI significantly increased by 5% for every county chlamydia rate reduction of
145 cases per 100,000 (1 standard deviation). Chlamydia trachomatis has been identified
as a possible HPV cofactor, synergistically increasing ICCI risk among co-infected
individuals214. It is possible that the unexpected relationship between ICCI and
Chlamydia on such large geographic scales as counties may be explained by increased
cervical cancer screening. Counties of higher Chlamydia trachomatis infection rates may
be indicators of areas with an increased probability of detecting pre-cancerous cervical
neoplastic lesions among women who otherwise would have progressed to ICCI; a form
of diagnostic bias77.

SEP or SEP-related variables were significantly associated with ICCI in school
district (1990 percent impoverished children) and census tract (PCA-derived SEP
component) final models. Both of these variables are thought to affect ICCI risk through
their affects on ICCI-related risk factors including: tobacco use391,392, higher risk sexual
behavior393, or Pap smear utilization242-244. Interestingly, despite using different SEP
measures these results appear relatively stable across census tract- and school district-
levels. Both SEP measures interacted with race in similar fashions; among whites, a one
standard deviation (SD) increase in SEP (census tract) or one SD (8.2%) decrease in
childhood poverty (school district) was associated with significantly higher odds of ICCI
(OR=1.59 and 1.23, respectively); among African Americans, a one SD increase in SEP
or one SD decrease in childhood poverty was also associated with significantly higher
odds of ICCI (OR=1.28 and 1.17, respectively), but to a lesser degree than Whites; and, among those of other races SEP and child poverty were both not significantly related to ICCI. A study examining age-adjusted ICCI rates by race and census tract SEP-measures suggests that ICCI rates by levels of education among Asian Americans do not exhibit as steep of a gradient as that observed among both non-Hispanic Whites and Blacks\textsuperscript{300}. Though not directly comparable, results of that study may support our findings of a possible interaction between SEP and race.

At the school district-level, one SD ($176,401) increase in 2004 taxable land value per pupil was associated with 0.86 (95% CI: 0.75-0.98) reduction in odds of ICCI. School district-level taxable land value per pupil could be considered a proxy for other well-known SEP measures. Indeed, further analysis demonstrates a relationship between childhood poverty, taxable land value per a pupil and ICCI. In the school district main effects model childhood poverty confounds the relationship between taxable land value per a pupil and ICCI by -36.6%; \( \text{OR}_{\text{Crude}}=0.63 \) versus \( \text{OR}_{\text{Adjusted}}=0.86 \). In contrast, taxable land value per a pupil has little confounding effect on the relationship between childhood poverty and ICCI (\( \text{OR}_{\text{Crude}}=1.24 \) versus \( \text{OR}_{\text{Adjusted}}=1.22 \)) (crude estimates not included in results). Taxable land value per a pupil is a theoretically selected variable thought to reflect not only area-SEP but also local sociopolitical influences (e.g., land use, ecologic degradation, etc)\textsuperscript{15,23}. Historically, the weight of causation points from taxable land value towards area- and individual-level SEP measures. Although, within shorter time frames individual- and area-level SEP is likely to contribute to changes in taxable land value. This interpretation suggests that childhood poverty may be on the causal pathway of
taxable land value and ICCI, in which case effects of childhood poverty should not be adjusted in a model including taxable land value.

Previous public health research has used individual-level property value (both land and other assets) as a measure of SEP or the theoretically different construct, wealth\(^{394,395}\) (also, see a review by Pollack\(^{396}\)). Those studies examining such health outcomes as self-rated health, chronic conditions (excluding cancer), depression, body mass index and obesity-related factors oftentimes found the associations with property or asset value to be significantly related to the outcome, independently of education and income\(^{394-396}\). This is the first known study to evaluate the association between an area-level measure of property value and cancer incidence (non-public health research has evaluated associations with area-level property value\(^{41,397,398}\)).

The amount of ICCI variation explained by area-level processes was small (0.41%-4.28%) but statistically significant in each model. This area-level variation differed by expected patterns between each model type (i.e., null or final) and geographic level. At all three geographic levels, the amount of possible variation attributable to an (residual) area-level process was higher in null versus final models. This simply suggests that final model covariates accounted for variation in area-level ICCI variation. Within null or final models, the percent of ICCI variation attributed to the area-level process was in the order of census tract (highest), school district, and county (lowest). This sequence may uninformatively reflect the average size of the area units within each geographic level; those geographic levels with the smallest average sizes (census tracts) will capture more variation than those of larger average sizes.
Stratification on Appalachian residence did not appreciably change estimates of the county- or census tract-level models. However, restricting to Appalachian residence did change some estimates of the school district-level model. The association between taxable land value per a pupil and ICCI within Appalachia was statistically insignificant while associations between the race category other and ICCI among women older than 53 years and residing in Appalachia became statistically significant.

As is in the full dataset, the relationships between ICCI and the area-level measures childhood poverty and taxable land value should be considered in interpreting this possible interaction by Appalachian residence. Excluding childhood poverty from the stratified models indicates that an increase of $176,401 taxable land value per a pupil was associated with a 0.48 (95% CI: 0.20-0.67) reduced odds of ICCI, among those residing in Appalachia (results not tabulated). Among those not residing in Appalachia, an increase of $176,401 taxable land value per a pupil was associated with a 0.30 (95% CI: 0.19-0.40) reduced odds of ICCI (results not tabulated). However, inclusion of childhood poverty confounded the non-Appalachia and Appalachia stratified odds ratios associated with these reductions by 18% and 49%; $\text{OR}_{\text{Crude|Non-Appalachia}}=0.70$ versus $\text{OR}_{\text{Adjusted|Non-Appalachia}}=0.85$ and $\text{OR}_{\text{Crude|Appalachia}}=0.52$ versus $\text{OR}_{\text{Adjusted|Appalachia}}=1.00$, respectively. This suggests that taxable land value is associated with ICCI only in non-Appalachian regions of Ohio. If true, support for our theoretical framework explaining disparities of ICCI in Appalachia would be gained if it could be shown that: 1) the relationships between taxable land value, Appalachian residence, childhood poverty, and ICCI persists over time; and 2) ICCI risk has reduced more quickly in non-Appalachia compared to Appalachia over time. It is also possible that these factors relate to one another such that
Appalachian residence is an effect of both school district-level childhood poverty and taxable land value (e.g., due to residential sorting or intergenerational migration behavior). If so, conditioning on Appalachian residence may further induce a relationship between taxable land value and childhood poverty because it is a collider between taxable land value and childhood poverty\textsuperscript{399}. More robust analyses may be needed to disentangle the effects of school district-level taxable land value, measures of SEP, Appalachian residence, and ICCI.

**Spatial Analysis**

The spatial distribution of observed ICCI surface probabilities suggests that the areas of increased risk tend to concentrate mainly in the Southeastern and Northwestern regions of the State. The ICCI spatial distribution generated here generally resembles that produced within a report by the Ohio’s state cancer registry\textsuperscript{70}. However, the surface probabilities image provides further resolution suggesting that sub-regional geographic disparities may exist. Although several Southeastern counties with historically high ICCI rates exhibit little variation in ICCI probability (Figure 5.1), several other counties may contain areas of both increased and decreased probabilities. Appalachia Ohio is known to have higher rates of ICCI than non-Appalachia\textsuperscript{65,66}. Interestingly, the image of surface probabilities suggests that larger ICCI disparities may exist within Appalachia than non-Appalachia as concentrations of the lowest and highest probabilities are within Appalachia.

The spatial distribution of ICCI surface probabilities produced from individual-level residuals of the six models does not substantially vary across model type or
geographic levels. This suggests that the general spatial distribution of ICCI that is attributable to individual-level processes is independent of all individual- and area-level fixed-effects included in any of the models. Although not statistically tested, it appears that spatial structuring remains within the six individual-level residual spatial distributions of ICCI. All six images suggest local areas of increased ICCI risk in several Northwest, Southeast and Eastern counties. Three larger, local areas of decreased cervical risk also remain after final model covariate adjustment.

The graphical results of Ripley’s K function suggests that cervical cancer cases, on average, have more cases surrounding them than expected for distances up to 12.32 km (7.66 mi). The multitude of possible causal factors affecting ICCI risk complicates any deeper discussions of this result. In terms of ICCI causal factors, the distance of 12.32 km may be interpreted as the distance at which the joint causal relationships affecting ICCI spatially structure so as to statistically significantly increase overall clustering of ICCI throughout the region. Specific ICCI causal relationships, however, may result in ICCI average cluster sizes smaller or larger than 12.32 km (or no spatial clustering).

Individual- and area-level residual ICCI distributions within each geographic level and model type combination appear to follow similar spatial distributions. For example, null model school district RIs spatially distribute in patterns similar to that of individual-level residuals estimated from the school district null model. Spatial patterns are similar within individual- and area-level residuals of final models of the same geographic level as well. This suggests that the process(es) affecting ICCI in Ohio do not
impact the spatial patterning of individual-level ICCI variation differently than the spatial patterning of area-level sources of ICCI variation.

The distributions of area-level RIs appeared to vary greatly between null and final models of the same geographic level. Within each geographic level, RIs from the null model had a larger variance, a higher index of overall spatial clustering, and more areas involved in local clusters. Together, this suggests that final model covariates explained area-level variation in such a way that also reduced the overall and local spatial structuring of ICCI. For county models the entire statistically significant global and local spatial structure of county-level ICCI was explained by final model covariates. Similarly, the statistically significant overall spatial clustering of census tract-level ICCI variation was explained by census tract final model covariates. However, 208 (7.3%) final model census tract-level RIs were still found to be involved in local spatial clusters. Although RI variance, overall spatial clustering, and local spatial clustering of school district null models were all reduced with final model covariate adjustment, all these measures remained statistically significant. This suggests that statistically significant overall and local spatial structuring remains at the school district level after adjustment for individual-level age and race (and their interaction), school district-level childhood poverty, taxable land value per a pupil, and an interaction between race and childhood poverty.

Sensitivity analyses of the neighborhood conceptualization used in area-level spatial analyses generally confirms the original results and conclusions. A few exceptions and patterns comparing results using the ZINDF and queen contiguity conceptualizations are of note. Indices of RI overall spatial clustering using ZINDF neighborhood
conceptualization were more conservative at each model type by geographic level comparison with the exception of school district final models (Table 5.5). The number of RIs involved in local spatial clusters under a ZINDF neighborhood definition were fewer for both county model types, but greater for both model types at the school district- and census tract-level. The difference in local spatial cluster identification by neighborhood conceptualization may be a function of area-level size and extent of local spatial clustering. The short distance (12.3 km) associated with the ZINDF conceptualization at the county level may lead to fewer counties being considered neighborhoods to one another than the queen contiguity conceptualization. In contrast, the 12.3 km zone of the ZINDF neighborhoods in the comparatively smaller school district and census tract analyses may allow for a larger number of neighbors compared to queen contiguity neighborhood conceptualization. The increased number of statistically significantly clustered RIs with increased neighbors of school district and census tract analyses suggests that local spatial clustering may operate, on average, on a larger scale than a school district or census tract’s immediate neighbors. Generally, comparisons of spatial distribution of statistically significant local spatial clusters are similar. Notably, the increased number of statistically significant locally clustered RIs associated with ZINDF conceptualization appears largely concentrated in the more urban regions with smaller area units (for similar reasons provided explaining the increase in number of units involved in local spatial clusters under ZINDF conceptualization).

Limitations
This exploratory multilevel and spatial analysis is not without limitations. Possibly most importantly is the accuracy and representativeness of the control dataset. The control dataset was synthesized from several data sources and using various methods. This dataset’s design has varying magnitudes of inaccuracies associated with the spatial location and demographic assignments made at those locations. If these inaccuracies are distributed equally across all exposure levels, then we may assume non-differential misclassification of ICCI status, leading to results biased towards the null\textsuperscript{77}. However, if these inaccuracies are associated with any variables considered in modeling, then differential misclassification of our outcome may exist, biasing our estimate either towards or away from null\textsuperscript{77}.

Demographic data used in the control dataset creation was from 2005-2009\textsuperscript{27}. While these dates are included in the range of cervical cancer cases diagnosis years, they do not represent all years. It is possible that population dynamics shifted from 1996 (first possible diagnosis) to 2009, in which case the control dataset may not adequately reflect the distribution of cases. This possible bias cannot be quantified without knowing how the female population of Ohio may have changed in the years not included in the control dataset (1996-2004).

Exposure misclassification bias due to cervical cancer case migration may also affect our results\textsuperscript{27,77,258,296,400}. Latency periods, or the period of time in which effects of area-level factors are thought to influence ICCI risk, vary between area-level covariates\textsuperscript{77}. For example, the effects of school district-level taxable land value per a pupil is thought to affect ICCI risk primarily through decreased area-level SEP and community services and resources (i.e., temporally indefinite latency), and decreased educational quality (i.e.,
latency during years of education). In contrast, census tract-level SEP may partially reflect individual-level SEP or community services and resources (i.e., temporally indefinite latencies). Obviously, lack of residential histories precludes any investigation of this bias. For area-level covariates, if we could assume that migration patterns are similar across levels of area-level factors and that the area-level covariate did not serve as a proxy for a truly individual-level factor, then we could conclude non-differential misclassification bias.

Though all variables were considered according to the *a priori* defined theoretical framework, many are to be considered variables of convenience. For example, percent population who smoke, teen birth and Chlamydia rates, and percent of population who are physically inactive were only meant to capture variation in their individual-level analogues. Ideally, much more individual-level data would have been available to avoid these convenience variables. Inclusion of individual-level SEP would have led to clearer conclusions of any significant area-level SEP-related variable. Lack of these individual-level data serves to greatly limit the strength of these results and any sound conclusions that may be drawn. Moreover, the relatively small proportion of variability due to area-level effects underscores the possibility that any statistically significant area-level associations with ICCI may be due to unmeasured individual-level confounding.

**Strengths**

The strength of these analyses focus on the integrative theoretical and methodological approach. All variables considered in models were chosen based upon previous research or the theoretical framework. This approach to variable consideration
allowed qualitative ascription of a possible causative link to any of the statistically significant correlations. Integrating multilevel and spatial analyses to investigate possible neighborhood and spatial effects is an approach that has gained popularity in recent years. As noted by others, the systematic examination of individual- and area-level sources of variation along with possible spatial structuring of that variation allows for a more comprehensive understanding of the health outcome.

Several other methodological approaches add strength to these exploratory analyses, including: Ripley’s K-function to define neighborhood conceptualization; estimated probability surfaces for individual-level variation visualization; neighborhood sensitivity analyses; overall and local spatial clustering computations; and visualization and comparison of area- and individual-level sources of variation and spatial clustering results. Ripley’s K-function allowed an objective conceptualization of an area’s neighborhood. Computation of surface probabilities allowed for a high-resolution visualization of ICCI that was not constrained by administrative boundaries. Sensitivity analyses allowed the comparison of results and conclusions that followed from spatial analyses using the common queen contiguity neighborhood. Global and local Moran’s I indices provided objective measures describing the magnitude and location of any statistically significant area-level spatial clustering. Lastly, a clearer picture of the spatial relationships was realized though the use of maps to simultaneously convey information on: the magnitude, variability and spatial distribution of ICCI; the spatial clustering tendencies of area-level ICCI; and display of common geographic reference features to better contextualize the outcome being displayed.
Although exploratory, these results may have impact for public health practitioners, cancer prevention and control researchers, or development of health policy. Possibly most obvious is the benefit of these results to public health practitioners in Ohio. The identification of several regions with an increased risk of ICCI could be targeted by practitioners. Cancer prevention and control researchers may use these results to guide future research. The areas in Southeastern and Northwestern Ohio that exhibit elevated ICCI risk after adjustment for covariates may warrant closer investigation. Equally interesting for cancer prevention and control research may be the regions of consistently low ICCI risk in Northeastern Ohio. The race-specific ICCI risk reversal by age group is an unexpected result that also deserves further consideration. These results may also lay the groundwork for future studies explicitly investigating the effects of land tax policies, school district revenue or other possible associations with ICCI that may be best addressed through policy.

Conclusions

The spatial patterns of ICCI in Ohio suggest that geographic disparities may persist after controlling for several factors, including various SEP measures. Identifying factors associated with ICCI at various levels is important for the development of public health interventions and health policies. Integrating spatial analyses and results visualization with multilevel analyses allows the potential for further explanation of ICCI disparities in Ohio. Recent attention directed towards not only multilevel statistical models but multilevel interventions suggest that these analyses may become increasingly useful. Adequate theories enabling the integrative explanation of these multilevel
factors and processes across space are just as important as the analyses that the theories motivate and frame. The approach to explicitly integrate neighborhood and spatial theories and methodologies has great potential to identify, intervene upon, and ultimately reduce ICCI disparities.
CHAPTER 6: PATHWAYS LEADING FROM SOCIOECONOMIC AND BIOBEHAVIORAL FACTORS TO CERVICAL ABNORMALITIES AMONG WOMEN IN APPALACHIA OHIO

Abstract

Background: Socioeconomic and geographic factors affect the distributions of invasive cervical cancer incidence (ICCI) and ICCI-related risk factors. This study aims to investigate the relationships between biologic, behavioral, socioeconomic, and sociopolitical constructs hypothesized to contribute to risk of ICCI and ICCI-related factors.

Methods: Structural equation modeling (SEM) was used on a sample (N=800) of Appalachian women from the Community, Awareness, Resources and Education (CARE) Project. Relationships between the latent variables taxable land value per a pupil, area-level deprivation, individual-level socioeconomic position (SEP), age, tobacco use, risky sexual behavior, human papillomavirus (HPV) infection and persistence, and cervical abnormalities were investigated. Mediating effects were investigated.

Results: The final model contained relationships between individual-level SEP, age, tobacco use, risky sexual behavior, HPV infection and persistence, and cervical abnormalities. Increased risky sexual behavior and HPV infection and persistence were associated with increased cervical abnormalities (both P <0.05). Age was inversely associated with risky sexual behavior and tobacco use (both P<0.05). There was marginal
evidence for the inverse relationships between individual-level SEP and risky sexual behavior and HPV infection and persistence (both $P=0.07$). Tobacco use decreased with increased individual-level SEP ($P<0.05$).

**Discussion:** A theoretical model linking sociopolitical and area-level socioeconomic factors with ICCI-related factors was not supported. Age and individual-level SEP were related to several ICCI-related factors and may be important targets or points of intervention in efforts to reduce ICCI disparities in Appalachia Ohio.

**Introduction**

Cervical cancer incidence and mortality rates have steadily decreased in the U.S. since the 1950s with uptake of the Papanicolaou (Pap) test$^{6,5}$. Universal distribution of the highly efficacious human papillomavirus (HPV) vaccine allows the possibility for cervical cancer rates to be reduced by another 70%$^{6,7}$. Several biologic and behavioral factors that may contribute to risk of ICCI have also been identified: persistent high risk (HR)-HPV infection (a sexually transmitted infection [STI] and necessary ICCI cause), age, tobacco use, parity, oral contraceptive use, infection with other STIs, and immune suppression$^{6,8,5,36}$. Disparities of ICCI may develop and persist when these risk or protective factors become disproportionately distributed across segments of the population.

The burden of cervical cancer is patterned along distinct demographic, geographic, historical, and social gradients. African Americans compared to whites$^{2,8}$ and Hispanics compared to non-hispanics$^{2,8}$ have elevated risk of ICCI. Lower education and income measured at individual-and area-levels have consistently been associated with increased cervical cancer incidence$^{276,300,301,378-382}$. Certain regions of the U.S., including
Central Appalachia, have reported persistently elevated ICCI rates for decades (only mortality rates were reported prior to the 1990s)\textsuperscript{11,12}.

Evidence continues to mount suggesting that socioeconomic and geographic factors (i.e., suitable transportation, geographic access to healthcare, rural/Appalachian residence, etc) also affect the distributions of ICCI risk and protective factors, including: tobacco use\textsuperscript{49,65,66,69}, Pap test\textsuperscript{235-237,242,243}, and HPV vaccine utilization\textsuperscript{385}. Central Appalachia is a region known to have lagging economic and social indicators since the late nineteenth and early twentieth centuries\textsuperscript{13-17}. Previous theoretical research has sought to integrate the social, political, and economic processes historically interacting in Central Appalachia with known risk factors of ICCI. A subsequent empirical application of this theoretical research found that school district taxable land value per a pupil – a present-day manifestation of historically acting sociopolitical and economic processes – was inversely associated with odds of ICCI in Ohio.

Despite the large body of literature demonstrating consistent relationships between ICCI and ICCI-related factors with socioeconomic and geographic factors, no known study has attempted to quantify the relationships between these factors in a single analysis. This study aims to further test the previously mentioned theory by investigating the relationships between cervical abnormalities (a necessary ICCI precursor), biologic and behavioral ICCI risk factors, and socioeconomic and sociopolitical risk factors of both ICCI and ICCI-related factors among Appalachian women of the Community, Awareness, Resources and Education (CARE) Project.
Methods

Study Sample

The CARE Project is a Center for Population Health and Health Disparities, funded by the National Cancer Institute (CA105632, 2003-present)^235,318. The purpose of CARE I (2003-2007) was to utilize transdisciplinary approaches to address disparities of cervical cancer incidence among females of Appalachia Ohio. Data used in this analysis are from CARE I project 3, a case-control study investigating the social, behavioral, and biologic variables that may contribute to increased risk of abnormal cervical cytology. A two-stage sampling scheme was utilized to achieve a sample of females that would allow estimation of the effects of area-level socioenvironmental (e.g., county rurality) and individual-level biobehavioral and socioeconomic (e.g., tobacco use, age at coitarche [age of first sexual intercourse], educational attainment) factors on abnormal cervical cytology. At the time of this study’s conception, Ohio Appalachia consisted of 29 counties (all present-day counties but Ashtabula, Trumbull, and Mahoning) with females being recruited only from these 29 counties. Participants were sampled using a stratified sampling plan to enable a heterogeneous sample of females with respect to urban/rural residence.

A census of clinics within these counties was generated (N=171 counties). A total of 22 clinics met the following inclusion criteria: performed Pap smear tests, reported at least 200 Pap screenings per month, and served a demographically diverse population. These clinics were approached and asked to participate in the study of which 14 agreed (63.6%). An additional three clinics that were approached following enrollment of the initial 14 also participated in the study.
Patient listings were gathered from each of the 17 enrolled clinics and females were randomly selected on a monthly basis. Study personnel reviewed medical records of 6028 randomly selected females to deem eligibility. Females were eligible if at least 18 years of age, a resident of Appalachia Ohio, not pregnant, no history of hysterectomy or invasive cervical cancer, and seen within a participating clinic within the last two years. Potentially eligible participants were approached and asked to participate in person by a study nurse on the day of the female’s routinely scheduled pelvic examination. Of the 4043 total eligible females, 2394 agreed to participate (59.2%). Participants were informed of the study, signed a written informed consent, completed a self-administered questionnaire, and provided blood and saliva samples before undergoing the pelvic examination.

The self-administered questionnaire gathered all demographic, socioeconomic, behavioral, and biologic information considered for inclusion in the current study. After completion of the self-administered survey, participants proceeded to undergo the routine pelvic examination and Pap smear. Pap smear results were reported using the Bethesda system\textsuperscript{95}. Pap smear results were broadly classified into three categories: 1) negative for intraepithelial lesion or malignancy (NILM), 2) epithelial cell abnormalities, and 3) other. Results deemed as representing epithelial cell abnormalities were further interpreted and classified into cell types (i.e., squamous or glandular). Squamous cell abnormalities, the most common type, were classified into the following: 1) atypical squamous cells (ASC), which are often further classified and presented as, of undetermined significance (ASC-US), or cannot exclude high-grade squamous intraepithelial lesion (ASC-H); 2) low-grade squamous intraepithelial lesion (LSIL); 3) high-grade squamous intraepithelial
lesion (HSIL); and 4) squamous cell carcinoma. An additional cervical sample was assayed for HPV typing (AMPLICOR: Roche, Indianapolis, IN) at the Centers for Disease Control and Prevention (CDC).

A total of 284 females presented with epithelial cell abnormalities leaving 2081 with normal Pap smear tests (29 had an unsatisfactory or missing sample). Females with normal Pap smear tests were matched to those with abnormal Pap smear tests based on clinic site and date of Pap smear (within a three month window) and randomly selected to achieve a three to one, normal (n=847) to abnormal Pap smear test ratio. HPV typing was conducted for these 1131 females and classified into: 1) HPV negative, 2) HR-HPV positive (types 16, 18, 31, 33, 39, 45, 51, 52, 56, 58, 59, 68, 73, or 82), and 3) low risk (LR) HPV positive (any other type).

Participants’ addresses were extracted from the medical record and geocoded to the point-level. Of all participants for which a mailing address was provided, 66 (5.8%) were post office box only, and five (0.4%) addresses could not be assigned a latitude or longitude due to mismatches with address location methods.

**Measured Variables Considered for Structural Equation Modeling**

Relevant demographic data included age, race (“White”, “Black of African American”, “American Indian or Alaskan Native”, “Asian”, “Native Hawaiian/Other Pacific Islander”, “Mixed”, and “Other”), and marital status (“Married”, “Divorced”, “Widowed”, “Separated”, “Never been married”, “Member of unmarried couple”, and “Don’t know”). Age was centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. All race categories besides “White” were collapsed to create a race dichotomy of “White” and “Non-White”. Three marital status categories
were created through the following combinations: 1) “Married” and “Member of unmarried couple”, 2) “Divorced”, “Widowed”, and “Separated”, and 3) “Never been married”.

Relevant socioeconomic data included previous year’s total household income (“Less than $10,000”, “$10,001 to $15,000”, “$15,001 to $20,000”, “$20,001 to $25,000”, “$25,001 to $35,000”, “$35,001 to $50,000”, “Over $50,001”, “Don’t know”, and “I prefer not to answer”) and educational attainment (“None”, “Preschool/Kindergarten”, “1st grade”, … , “12th grade”, “GED”, “Some college, technical or trade school, but less than 1 year and no degree”, “1 or more years of college, technical or trade school, but not degree”, “technical or trade degree or certification”, “Associate degree”, “Bachelor’s degree”, “Master’s degree”, “Doctorate degree or other advanced degree”, and “Don’t know”). Income categories were combined to create the following groups: 1) $20,000 or less, 2) $20,001-$50,000, and 3) at least $50,001.

Education was centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. Non-responses or responses of “Don’t know” or “I prefer not to answer” for either socioeconomic-related question was treated as missing.

Current cigarette smoking (“Former smoker”, “Current smoker”, “Never smoker”, and “Don’t know”) and various sexual health questions were used to characterize participants’ relevant behaviors that may affect risk of abnormal cervical cytology. Sexual health data included age at coitarche, total number of male partners with whom a female had sexual intercourse within the previous 12 months, ever used condoms (“Yes”, “No”, “Don’t know”, and “I prefer not to answer”), and previous history of an abnormal Pap smear (“Yes”, “No”, and “Don’t know”). Age at coitarche was centered and
standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. Due to major normality violations number of recent male partners was grouped to create the following categories: 1) zero partners, 2) one partner, 3) two partners, and 4) more than two partners (maximum of twelve). Non-responses or responses of “Don’t know” or “I prefer not to answer” were treated as missing.

Area-level covariate data were gathered from the 2000 Census of Population and Housing (Summary File 4) of the U.S. Census Bureau (census tract)\textsuperscript{28} and 2004 property tax data of the Ohio Department of Taxation (school-district level)\textsuperscript{34,35}. Census tract-level variables were: median household income (median of household incomes, 1999); percent unemployed (percent of total civilian population of at least 16 years in the labor force whom are not employed, 2000); percent impoverished (percent of total population with incomes less than 100% the federal poverty level, 1999); percent without a high school diploma (percent of population of at least 25 years without a high school diploma or equivalent, 2000); and percent without a bachelor’s degree (percent of population of at least 25 years without a 4-year college degree, 2000). The school district-level variable was taxable land value per pupil (value of all property subject to taxation and after application of assessment rates ÷ average daily school membership). All six area-level covariates were centered and standardized (mean=0, standard deviation=1) and analyzed as a continuous variable. Centering and standardization was conducted only for those census tracts and school districts represented in the sample (i.e., omitting those regions that did not have a study participant residing within). Due to heavy skewness log transformations were used for percent unemployed, percent impoverished, percent without a high school diploma, and taxable land value per pupil.
Ohio’s census tract (year 2000) and unified school district (year 2012) spatial boundary files were gathered from the U.S. Census Bureau’s TIGER program\textsuperscript{330}. A series of one to one spatial join operations enabled the linkage of area- to individual-level data through the point-level geocoded participant’s address. Eighty-seven of 615 school districts, and 201 of Ohio’s 2940 census tracts were represented in the spatially joined dataset.

**Latent Variables Considered for Structural Equation Modeling**

Seven latent variables were constructed from one or more measurement variables and used in structural modeling: 1) taxable land value, 2) area-level deprivation, 3) individual-level SEP, 4) risky sexual behavior, 5) HPV infection and persistence, 6) tobacco use, and 7) age (Table 3.1). Taxable land value per a pupil served as the sole indicator of the taxable land value latent variable. Area-level deprivation was indicated by the five census tract-level covariates listed above. Previous year’s total household income and educational attainment were indicators of individual-level SEP. Choice of risky sexual behavior indicators was partially informed by previous work conducted by Reiter, et al\textsuperscript{357} and included: number of recent male sexual partners, condom use, age at coitarche, and marital status. HPV infection and persistence was indicated by current HPV infection status and previous abnormal Pap smear test (a possible indicator of persistent infection). Current cigarette smoking was the sole indicator of tobacco use. Current age was the sole indicator of age. The hypothesized model with measured variables, latent variables, and directed relationships between measured and latent variables is displayed in Figure 3.4 (Race was excluded from all model consideration when discovered that the sample was overwhelmingly White). As is common practice,
one indicator of each latent variable was held constant to set the scale of the latent variable\textsuperscript{339}. Those paths in Figure 3.4 marked with a “*” are to be estimated while those marked by a “C” are to be held constant.

**Statistical Analysis**

Descriptive statistics (means and standard deviations for continuous variables and totals and percents for categorical variables) were computed for all measured variables using SAS (SAS Institute Inc., version 9.3, Cary, North Carolina).

**Structural Equation Modeling**

Structural equation modeling (SEM) was used to estimate the relationships of Figure 3.4 and test how well the data fit the hypothesized model. Because SEM techniques allow examination of relationships between multiple independent and dependent variables, both direct and indirect effects were investigated. Relationships were reported as standardized coefficients (i.e., interpreted as standard deviation changes). Relationship statistical significance was assessed through calculation of scaled $z$-tests (accounting for non-normality). Type-I error was held at 0.05. The total variation of cervical abnormalities explained by the latent variables was estimated and reported.

Similar to linear regression, SEM requires that many assumptions be true, including: multivariate normality and absence of outliers; relationship linearity; absence of multicollinearity and singularity; and residual normality, homoskedasticity and independence\textsuperscript{358}. The categorical natures of several measured variables of the hypothesized model complicated the examination of these assumptions. Multicollinearity and singularity were assessed though examination of the determinant of the model correlation matrix; small determinants indicating assumption violation. It was necessary
that all categorical variables represented ordered categories of an underlying latent continuous distribution (i.e., “No HPV”, “LR-HPV”, and “HR-HPV” are categorized as zero, one, and two from an unobserved distribution of HPV types that continuously vary with respect to risk of cervical abnormalities). All categorical variables were ordered and could be said to have been grouped from a continuous distribution except marriage. It was assumed that marital status would be coded by the following: “Never been married” = zero (highest risk for cervical abnormalities), “Divorced”, “Separated”, or “Widowed” = one (moderate risk for cervical abnormalities), and “Married” and “Member of an unmarried couple” = two (lowest risk for cervical abnormalities).

Due to the use of categorical variables, model estimation was performed using the correlation matrix. Due to the expected non-normality of the data, scaled Satorra-Bentler maximum likelihood estimation techniques were used to estimate parameter standard errors \(^{359}\). Accordingly, model fit was assessed using the Satorra-Bentler scaled chi-squared, comparative fit index (CFI), and root mean-square error of approximation (RMSEA)\(^{339,359}\). As in other goodness of fit indices, the chi-square test statistic measures the degree to which the observed correlation matrix differs from that of the models, and thus a nonsignificant chi square test is desired. Because the p-value of the chi-square test statistic is dependent on the sample size, the CFI and RMSEA will be the primary indices used to determine fit. Acceptable fit is indicated by a CFI value greater than 0.92 and RMSEA value less than 0.05. SEM analyses were conducted using EQS (Multivariate Software, Inc., version 6.2, Encino, California).
Results

Sample Description

Numerous participants (N=331, 29.3%) were missing information on one or more covariates of the hypothetical model. Those covariates associated with the largest missing data include: household income (N=274, 24.2%), any area-level covariate (N=71, 6.3%), number of recent male sexual partners (N=45, 4.0%), previous abnormal Pap smear tests (N=40, 3.5%), and condom use (N=36, 3.2%). Individual- and area-level characteristics for participants with complete data (N=800) are shown in Table 6.1. Seventy five percent of the final analytic dataset had normal cervical cytology while the remaining 25% was split between those with ASCUS (12.9%) or SIL and squamous cell carcinoma (12.1%). The majority of participants were White (96.0%), attained educational levels of technical or trade degrees or less (79.7%), were member to households with annual incomes less than $50,000 (80.5%), were married or in a marriage-like relationship (51.1%), had no more than one male sexual partner in the past year (81.5%), used condoms (76.1%), were a former or current smoker (57.8%), did not have a prior history of abnormal Pap smear tests (53.1%), and were not positive for any HPV type (56.6%). The census tracts in which participant’s resided averaged a percent unemployment of 6.9%, percent impoverished of 11.0%, median household income of $32,321, and percent less than a high school diploma or equivalent of 21.1%. Participant’s mean school district taxable land value per a pupil was $55,225.
<table>
<thead>
<tr>
<th>Variable</th>
<th>N/Mean</th>
<th>%/S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervical abnormalities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>600</td>
<td>75.0%</td>
</tr>
<tr>
<td>ASCUS</td>
<td>103</td>
<td>12.9%</td>
</tr>
<tr>
<td>SIL or more severe</td>
<td>97</td>
<td>12.1%</td>
</tr>
<tr>
<td>Age (years)</td>
<td>34.7</td>
<td>12.7%</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>768</td>
<td>96.0%</td>
</tr>
<tr>
<td>Non-White</td>
<td>25</td>
<td>3.1%</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or equivalent</td>
<td>341</td>
<td>42.6%</td>
</tr>
<tr>
<td>Some college or technical/trade</td>
<td>297</td>
<td>37.1%</td>
</tr>
<tr>
<td>Associate degree of higher</td>
<td>162</td>
<td>20.2%</td>
</tr>
<tr>
<td>Household income ($)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; $20,001</td>
<td>381</td>
<td>47.6%</td>
</tr>
<tr>
<td>$20,001-$50,000</td>
<td>271</td>
<td>33.9%</td>
</tr>
<tr>
<td>&gt;$50,000</td>
<td>148</td>
<td>18.5%</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never been married</td>
<td>233</td>
<td>29.1%</td>
</tr>
<tr>
<td>Single, divorced/separated, widowed</td>
<td>158</td>
<td>19.7%</td>
</tr>
<tr>
<td>Married, marriage-like couple</td>
<td>409</td>
<td>51.1%</td>
</tr>
<tr>
<td>Coitarche (Age in years)</td>
<td>16.6</td>
<td>2.4%</td>
</tr>
<tr>
<td>Recent male sexual partners</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>72</td>
<td>9.0%</td>
</tr>
<tr>
<td>1</td>
<td>580</td>
<td>72.5%</td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>11.9%</td>
</tr>
<tr>
<td>&gt;2</td>
<td>53</td>
<td>6.6%</td>
</tr>
<tr>
<td>Condom use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>191</td>
<td>23.9%</td>
</tr>
<tr>
<td>Yes</td>
<td>609</td>
<td>76.1%</td>
</tr>
<tr>
<td>Tobacco smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>338</td>
<td>42.2%</td>
</tr>
<tr>
<td>Former</td>
<td>139</td>
<td>17.4%</td>
</tr>
<tr>
<td>Current</td>
<td>323</td>
<td>40.4%</td>
</tr>
<tr>
<td>Previous abnormal Pap smear test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>425</td>
<td>53.1%</td>
</tr>
<tr>
<td>Yes</td>
<td>375</td>
<td>46.9%</td>
</tr>
</tbody>
</table>

Table 6.1. Characteristics of participants in the CARE study with complete data on covariates of the hypothetical model

Cont’d
Table 6.1 (Cont’d).

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>%/S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HPV status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>453</td>
<td>56.6%</td>
<td></td>
</tr>
<tr>
<td>Low-risk</td>
<td>99</td>
<td>12.4%</td>
<td></td>
</tr>
<tr>
<td>High-risk</td>
<td>248</td>
<td>31.0%</td>
<td></td>
</tr>
<tr>
<td><strong>Census tract-level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent less than a bachelors degree</td>
<td>89.0</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Percent less than a high school diploma</td>
<td>21.1</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>Percent unemployed</td>
<td>6.9</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>Percent impoverished</td>
<td>11.0</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>Median household income ($)</td>
<td>32,321</td>
<td>6,628</td>
<td></td>
</tr>
<tr>
<td><strong>School district level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxable land value per a pupil ($)</td>
<td>55,225</td>
<td>23,106</td>
<td></td>
</tr>
</tbody>
</table>

1 Total race frequencies and percents do not equal that of other groups because race was omitted from model consideration upon realizing low variability (i.e., participants included in the analytic dataset have missing race information).

2 None was HPV negative for all types; High-risk (HR) HPV types were: 16, 18, 31, 33, 39, 45, 51, 52, 56, 58, 59, 68, 73, or 82; low-risk HPV types were any non-HR type

Model Development

The hypothesized model depicted in Figure 3.4 would not converge. This model was systematically simplified by removing paths between variables involved in error messages and those that had high correlations indicated within the correlation matrix. The first model to converge (with errors) is identical to the initial hypothesized model excluding the path between tobacco use and HPV infection and persistence. Maximum likelihood results of several variations of the hypothesized model indicate that the determinant of the correlation matrix may not be positive definite, suggesting multicollinearity or singularity issues (oftentimes leading to very large standard errors, variances constrained at zero by the program, nonsensical fit indices, or error messages indicating a source of singularity). Model development proceeded until the program did not provide indications of any of the following: issues with singularity, program-initiated
variances constrained at zero, nonsensical fit indices, or very large estimate standard errors.

Final Model

Model Fit

The final model with standardized path coefficients is shown in Figure 6.2. The Satorra-Bentler scaled chi-square for goodness of fit is highly significant, suggesting a lack of fit (chi-square=124.6, degrees of freedom [df] =36, P<0.0001). However, the CFI of 0.944 and RMSEA of 0.056 (90% CI: 0.045, 0.066) both indicate good fit. The determinant of this model’s correlation matrix, 0.045, is rather small but an order of magnitude larger than most other models that were fit. Because model development led to the fitting of several models, a correlation between parameter estimates of the final model and the first hypothesized model that converged was calculated; r(30) = 0.99, P < 0.01. The high correlation of the thirty parameter estimates shared between the two models indicate that despite model modifications the shared estimates are highly consistent with one another.

Direct Effects

Of the three direct paths leading to cervical abnormalities, only HPV infection and persistence is statistically significant; cervical abnormalities increases as HPV infection and persistence increases (standardized coefficient = 0.88, P < 0.05). However, several other statistically significant direct relationships are of note: individual-level SEP increases with older age (standardized coefficient = 0.38, P < 0.01); tobacco use (standardized coefficient = -0.50, P < 0.01) decreases with increased individual-level SEP; tobacco use (standardized coefficient = -0.46, P < 0.01) and risky sexual behavior
Figure 6.1. Final structural equation model with standardized coefficients of directed paths estimated between measured variables and latent variables among 800 participants in the CARE study

(standardized coefficient = -0.82, P < 0.01) decrease with increased age; and HPV infection and persistence increase with increased risky sexual behavior (standardized coefficient = 0.64, P < 0.001). There was marginal support that increases in individual-level SEP was associated with decreases in risky sexual behavior (standardized coefficient = -0.34, P = 0.068).

Indirect Effects

Indirect effects can be examined for both latent and measured variables. Age, individual-level SEP, and risky sexual behavior may be investigated for possibly
indirectly affecting other latent variables as these relationships have indirect pathways leading to risky sexual behavior, HPV infection and persistence, cervical abnormalities, or tobacco use. Increased age is indirectly associated with decreased HPV infection and persistence (standardized coefficient = -0.51, \( P < 0.05 \)) and decreased cervical abnormalities (standardized coefficient = -0.42, \( P < 0.05 \)). Increased risky sexual behavior is associated with increased cervical abnormalities (standardized coefficient = 0.56, \( P < 0.05 \)). There is little evidence of indirect effects between age and risky sexual behavior (standardized coefficient = 0.64, \( P > 0.1 \)), age and tobacco use (standardized coefficient = 0.64, \( P > 0.1 \)), or individual-level SEP and cervical abnormalities (standardized coefficient = 0.64, \( P > 0.1 \)). Marginal evidence exists suggesting that increased individual-level SEP is associated with HPV infection and persistence (standardized coefficient = -0.18, \( P = 0.067 \)). Individual-level SEP, age, tobacco use, risky sexual behavior, and HPV infection and persistence explained 63.9% of the total variation in cervical abnormalities.

As all measured variables are dependent variables, several indirect effects between latent and measured variables can be examined. The following are the statistically significant indirect relationships between latent and measured variables: increased age was associated with increased age at coitarche (standardized coefficient = 0.41, \( P < 0.05 \)), decreased condom use (standardized coefficient = -0.29, \( P < 0.05 \)), decreased number of recent male sexual partners (standardized coefficient = -0.51, \( P < 0.05 \)), marital status (standardized coefficient = 0.66, \( P < 0.05 \)), decreased infection with HPV types (standardized coefficient = -0.51, \( P < 0.05 \)), and decreased reporting of prior history of abnormal Pap smear test (standardized coefficient = -0.23, \( P < 0.05 \)). Indirect
relationships between increased individual-level SEP and increased age at coitarche (standardized coefficient = 0.15), decreased condom use (standardized coefficient = -0.10), decreased number of recent male sexual partners (standardized coefficient = -0.18), marital status (standardized coefficient = 0.23), decreased infection with higher risk HPV types (standardized coefficient = -0.18), and decreased reporting of prior history of abnormal Pap smear test (standardized coefficient = -0.08) were all marginally statistically insignificant (0.06 < P < 0.08).

Discussion

Findings

This study has demonstrated that several factors were statistically significantly directly or indirectly related to increases in cervical abnormalities. Increased cervical abnormalities were directly associated with increases in HPV infection and persistence – a latent variable measured by infection with HPV type and prior history of abnormal Pap smear tests. Increases in the latent variable risky sexual behavior – measured by marital status, recent male sexual partners, condom use, and age at coitarche – was associated with increases in cervical abnormalities through HPV infection and persistence; increased risky sexual behavior directly increased HPV infection and persistence which, in turn, was associated with increased cervical abnormalities. Older age was indirectly associated with decreases in cervical abnormalities through pathways involving individual-level SEP, tobacco use, risky sexual behavior, and HPV infection and persistence. Both individual-level SEP and tobacco use were not statistically significantly related to cervical abnormalities, directly or indirectly.
This study has also found several significant relationships between ICCI-related risk factors. Tobacco use was inversely associated with both individual-level SEP and age. Risky sexual behavior and HPV infection and persistence were not associated with individual-level SEP, but were both inversely associated with age. Increases in risky sexual behavior were associated with increases in HPV infection and persistence.

Several associations found in this study are consistent with previous findings, including the associations between: HPV infection and persistence and increased risk of cervical dysplasia\textsuperscript{93,94}, increased risky sexual behavior and increased risk of HPV infection and persistence\textsuperscript{88,90,93}, older age and decreased HPV infection and persistence\textsuperscript{88,124,275,405}, older age and decreased cervical dysplasia\textsuperscript{124,405}, older age and decreased tobacco use\textsuperscript{391}, increased individual-level SEP and decreased tobacco use\textsuperscript{365,391}, older age and decreased risky sexual behavior\textsuperscript{393}, and older age and increased individual-level SEP\textsuperscript{406-408}.

This study also failed to replicate some well-established findings found between the following: increased tobacco use and increased cervical dysplasia\textsuperscript{93,148}, and increased individual-level SEP and cervical dysplasia\textsuperscript{275,387}. Previous studies report mixed associations between increased tobacco use and HPV infection and persistence\textsuperscript{90,275,409}. These findings may add to the body of evidence suggesting that tobacco use does not increase risk of HPV infection or persistence. Strictly speaking, this study failed to replicate previous research suggesting that increased individual-level SEP is associated with decreased risky sexual behavior\textsuperscript{393} and decreased HPV infection and persistence\textsuperscript{275}. However, the p-values of these associations were both less than 0.07 and in the expected direction. Other researchers have noted the difficulty in disentangling the relationships.
between age, tobacco use, SEP, risky sexual behavior, HPV infection and persistence, and cervical dysplasia because of shared risk factors and the possibilities of mediation among these factors and cervical dysplasia\textsuperscript{124,275}. Interestingly, Khan, et al. found an unadjusted association between individual-level educational attainment and HR-HPV infection\textsuperscript{275}. However, this association was insignificant with adjustment for age. Therefore, these findings may be consistent with those of Khan, et al, which together suggest that individual-level SEP may serve as mediator between age and HR-HPV infection and persistence.

Lack of model convergence with inclusion of the sociopolitical and area-level SEP constructs may suggest that the hypothetical model was not a good fit and that the theoretical framework prompting this model is incorrect. However, the modeling issues encountered when fitting the hypothetical model may have been the result of: 1) inadequate power (i.e., small sample size), 2) a lack of variability in measured variables, or 3) incompatible data to test the theory. First, the original study was not designed for a structural analysis of covariance and may have been underpowered to detect associations between certain variables. Second, cervical abnormalities were overwhelmingly comprised of those with normal or the mildest form of cervical abnormalities – ASCUS. Several other measured variables were categorical and, therefore, have a reduced variability of possible responses. It is possible that this reduced variability led to model convergence instability and small correlation matrix determinants. Third, the majority of ASCUS and LG-SIL naturally regress to normal cytology without the need for medical intervention and are not suitable indicators of ICCI. However, the theoretical model was developed based on factors that contribute to the risk of ICCI.
Limitations

There are several limitations of this study. As previously noted, the data used in these analyses may not be as compatible as needed to adequately test the theoretical framework with the techniques employed. Future studies attempting to replicate these findings may benefit from more detailed data collection and increased variability in certain responses (i.e., more specific income level, cigarette use frequency/duration, lower proportions of less severe cervical abnormalities, etc.).

Despite hypothesizing on the directed relationships between variables, the data used in this study were largely cross sectional. This severely inhibits assignment of causative claims to the estimated associations. Equally important in ascribing causation to these associations are the strong “no unmeasured confounding” assumptions being made in this model\textsuperscript{410,411}. Assertions of causality would be bolstered by a longitudinal study design and use of analytic techniques that better control for confounding, such as marginal structural modeling\textsuperscript{412}.

The external validity of this study may also be questionable as the characteristics of this sample of women may systematically differ from the general Appalachian Ohio population of women. These women all had usual sources of health care (as part of the inclusion criteria), were former or current smokers at frequencies higher than all Appalachian Ohio women\textsuperscript{65}, and were younger than all Appalachian Ohio women\textsuperscript{58}.

Strengths

Despite the limitations, several strengths underscore the importance of this research. The analytic approach taken allowed the modeling of theoretically informed constructs and relationships in ways not allowed by traditional regression analyses. The
investigation of direct and indirect effects has illuminated previous research by suggesting that individual-level SEP may serve as a mediator between age and HR-HPV infection and persistence. Several variables of interest in the investigation of ICCI disparities are latent and traditionally measured by multiple variables (i.e., risky sexual behavior, individual-level SEP, and HPV persistence). The modeling of latent variables as opposed to single variables allowed by SEM may have better captured the relationships between variables of interest, improving construct validity.

Few studies have attempted to quantify these associations within a sample of women from Appalachia. This is the first known study to report on the associations between age and individual-level SEP, age and HPV infection and persistence, and individual-level SEP and HPV infection and persistence among a sample of women from Appalachia who may be at an increased risk of developing ICCI.

Conclusions

A theory hypothesizing relationships between sociopolitical factors, area-level SEP, and individual-level SEP or cervical abnormalities was not supported by these data. However, these results do suggest that several relationships of public health importance exist between a sample of Appalachian women who may be at an increased risk for ICCI: older age is positively associated with individual-level SEP, which, in turn, is significantly associated with decreased cigarette use as well as decreased HPV infection and persistence; higher individual-level SEP may be associated with decreases in risky sexual behavior, which, in turn, is positively associated with increased HPV infection and persistence; risky sexual behavior and HPV infection and persistence are associated with cervical abnormalities; and smoking is not associated with HPV infection or persistence.
Efforts to reduce ICCI disparities in Appalachia Ohio guided by these results may benefit by targeting younger women of lower SEP for risk reduction behavior and HPV vaccine uptake. Age and individual-level SEP were shown to have strong effects on several well-established ICCI risk constructs in this study and may be important targets for intervention.
CHAPTER 7: CONCLUSIONS AND IMPLICATIONS

The goals of this research were to theoretically integrate current knowledge of the causes of socioeconomic disparities with the causes of ICCI disparities within Central Appalachia, to describe the multilevel and spatial characteristic of ICCI in Ohio, and to investigate relationships between sociopolitical, economic, and biobehavioral factors together with the prevalence of cervical abnormalities.

There are several potentially important results of this integrated approach to the investigation of invasive cervical cancer incidence (ICCI) and ICCI-related disparities. Findings from the application of the integrated theoretical framework to persistent disparities of ICCI in Central Appalachia demonstrated that: 1) relationships may exist between historically dependent sociopolitical, environmental, built environment, socioeconomic, and ICCI or ICCI-related factors; and 2) explanations of factors contributing to the development or persistence of ICCI disparities may be incomplete if the causes of socioeconomic disparities are not also addressed. Important results from the multilevel and spatial characteristics of ICCI in Ohio were: 1) despite large geographic disparities, a small percentage of the total ICCI variation is possibly due to county-, school district-, or census tract-level factors; 2) Whites may experience greater reductions of ICCI with increases in school district- and census tract-levels of socioeconomic position (SEP) compared to African Americans or those who are non-White and non-
African American; and 3) taxable land value per pupil, a present-day manifestation of historical sociopolitical processes, was inversely associated with ICCI. Finally, the investigation of the pathways leading from socioeconomic and bio-behavioral pathways to cervical abnormalities found: 1) a test of the theoretical model described above, linking sociopolitical and area-level socioeconomic factors to ICCI and ICCI-related factors, was not supported; and 2) age and individual-level SEP were related to cervical abnormalities and ICCI-related factors. The implications of these finding are further detailed in turn.

**Research Aim 1**

It has been demonstrated that relationships may exist between historically dependent sociopolitical (i.e., regional economic integration, sociopolitical clientelism and factionalism, land-use/land-tax policy, public education funding system), environmental (i.e., natural ecologic limits of a mountainous region, rurality), built environment (i.e., physical activity facilities and resources), socioeconomic (i.e., farm-and-forest economy, community resources, individual-level SEP), and ICCI or ICCI-related factors (i.e., human papillomavirus [HPV] infection and persistence, biologic stress, immune suppression, tobacco use, Pap smear utilization) in Central Appalachia. Verification of these relationships through empirical studies would have important implications for both integrative theories and explanations of disparities of ICCI in Central Appalachia. Empirical validation would strengthen the utility and applicability of this integrative theory. ICCI disparities within Central Appalachia may be an ideal ‘case-study’ to validate this theory because of the multitude of research conducted on ICCI. However, if the theory’s validity is dependent on the degree to which inter-related, context-dependent relationships can be verified, then it is likely that this theory has
limited applicability to intervene and reduce the absolute burden of ICCI (or many other health outcomes). It is likely—as is the case with ICCI—that more efficient interventions will be developed (e.g., Pap smear test and HPV vaccine), largely devoid of social context, as empirical evidence substantiates relationships between factors, processes, and the outcome under study. Therefore, the utility of this integrative theorizing may lie not in its ability to develop interventions, but rather to highlight the possibility that present-day factors and processes may have future consequences for health disparities (including ICCI).

Explanations of persistent ICCI disparities in Central Appalachia that do not include explanations of socioeconomic disparities may be incomplete. Previous research demonstrates consistent associations between ICCI and socioeconomic measures as well as socioeconomic disparities and social, economic, political, environmental, and cultural processes. Therefore, research concerned with the causes of ICCI disparities in Central Appalachia may yield more insightful findings if causes of socioeconomic disparities (e.g., taxable land value, land-use/land-tax rates, etc.) are also considered. Because of the longer temporal windows of exposure associated with the development of regional socioeconomic disparities, explanations of ICCI disparities that address socioeconomic disparities are less likely to decrease ICCI disparities in the short-term. This suggests that the present-day causes of ICCI health disparities may not always be the best points to intervene as the exposure latency period contributing to the disparity may have passed. As noted above, emphasizing the possibility that present-day factors and processes may have future consequences for ICCI disparities suggests that prevention of ICCI disparities may be addressed, in part, by addressing socioeconomic disparities.
Research Aims 2 and 3

Despite widespread geographic disparities, very little of the total ICCI variation in Ohio is accounted for by county-, school district-, or census tract-levels (1.6%, 3.2%, and 4.3%, respectively). This suggests that a fuller explanation of geographic disparities of ICCI in Ohio will require information other than that operating at county-, school district-, or census tract-levels. It may be inaccurate to conclude that any factors or processes operating at levels larger than an individual are unimportant. These three area-levels were chosen out of convenience as they have attributed information that serves as useful indicators of the constructs investigated (i.e., sociopolitical, socioeconomic, built environment, etc). Knowledge of these small percentages of variation explained could benefit future research that is focused on defining constructs that operate at levels not confined to an individual.

Both childhood poverty at the school district-level and a socioeconomic composite at the census tract-level were related to ICCI in a similar fashion. Both of these measures interacted with race to suggest that increases in area-level SEP may have more favorable effects for Whites as compared to African Americans. Moreover, the odds of ICCI among those who are non-White and non-African American may be unaffected by area-level SEP. The implication of this finding is complicated by the dependent associations between age and race of the same models. These results suggest that decreased ICCI risk among African Americans (compared to Whites) younger than 39 years may be less pronounced in areas of high SEP, while the increased risk among African Americans older than 53 years may be more pronounced in areas of high SEP. If true, interventions aimed at reducing racial ICCI disparities in Ohio may benefit by
targeting White women in areas of low SEP and African American women in areas of high SEP.

Taxable land value per pupil, a present-day manifestation of historical sociopolitical processes, was inversely associated with ICCI in Ohio. This association partially supports the integrative theory described above. If true, this indicates that cervical cancer disparities in Ohio may be related to the public education funding system of Ohio. This is the first known use of this variable in studies of any health outcome and cannot be compared to previous research. Therefore, this finding will need to be replicated by future research concerning health disparities that follow socioeconomic gradients.

**Research Aim 4**

The path analysis failed to support the theoretical model described above, linking sociopolitical and area-level socioeconomic factors to ICCI and ICCI-related factors. However, this may have been due to: 1) inadequate power (i.e., small sample size), 2) a lack of variability in measured variables, or 3) incompatible data to test the theory. It is also possible that the theory used is inappropriate for describing ICCI disparities in Appalachia Ohio. Future studies investigating ICCI disparities in Central Appalachia by using the methods should consider larger samples sizes, samples with more subject variability or more sensitive measurements, and currently accepted indicators of ICCI as an outcome.

Results of the path analysis suggested that age and individual-level SEP were related to cervical abnormalities and ICCI-related factors within a sample of Appalachian women that may be at an increased risk of ICCI, including: older age was positively
associated with individual-level SEP, which, in turn, was significantly associated with cigarette use; individual-level SEP may have mediated the relationship between age and HPV infection and persistence; and individual-level SEP may be associated with risky sexual behavior and HPV infection and persistence. If true, efforts to reduce ICCI disparities in Appalachia Ohio may benefit by targeting younger women of lower SEP for risk reduction behavior and HPV vaccine uptake.

Summary

The approach taken to investigate disparities of ICCI in Central Appalachia was broad and integrative in scope and methodology. This approach has addressed numerous aspects of cervical cancer disparities that cut across several scientific disciplines. Advantages of this approach include: 1) a deeper understanding of factors and processes that may contribute to the cause and persistence of cervical cancer disparities in Central Appalachia, 2) empirical evidence of a previously unconsidered association with cervical cancer disparities that has theoretical connections to historically relevant sociopolitical processes, and 3) detailed descriptions of the distribution of ICCI, and factors associated with ICCI, measured at multiple geographic levels. The time-intensive nature and possible compromises of discipline-specific depth for the sake of multidisciplinary breadth are important disadvantages of this integrative approach. However, the insight gained and future avenues of research identified may offset these disadvantages. Persistent ICCI disparities in Central Appalachia may be the result of complex relationships between historically-dependent social, economic, political, cultural, and environmental processes. Interventions may need to consider several of these processes in order to successfully reduce ICCI disparities in Central Appalachia.
APPENDIX A: Schematic Used to Geocode Participants of the Community Awareness Resources and Education Project 3 Study
The following is a description of the steps taken in geocoding participants of the Community Awareness Resources and Education study. These women were enrolled asked to provide their address. Issues arose one of two general ways: reporting of addresses which carried multiple names (e.g., US 23 or High street), or when the information was transcribed with mistakes (e.g., 43235 becomes 43233, Moody street becomes Moddy, etc).

I. Visually scan database of addresses for any blatant errors (e.g., missing address components, P.O. boxes, addition of apartment information to street address field, database errors [column shift, ‘Null’, etc])
   A. Subset any addresses with errors known to prevent geocoding which cannot be manually corrected from visual scan*
   B. Subset separately PO boxes for discussion of proper handling
II. Process initially checked addresses in geocoder (used ArcGIS 9.3, year 2000 Ohio street file, built address locator using alphanumeric street and zip zone for address identification)
III. Subset the unmatched (or unacceptably matched) addresses
   A. Manually enter these addresses into the United States Postal Service (USPS) ‘Zip Code Lookup’ @
      1. The USPS zip code lookup is helpful in the following:
         a. Verifying accurate zip codes
         b. Finding and correcting 1 or 2 alphabet character errors (e.g., ‘Moddy’ will be corrected to ‘Moody’, ‘Ventura’ corrected to ‘Ventoura’, etc) – not very robust at correcting any numeric errors
         c. Correcting street suffixes (e.g., ‘McGhee Ln’ corrected to ‘McGhee Dr’)
         d. Verifying street numbers even though they may be within a possible range (will return an error of “Address may be undeliverable” if street number is within range but not valid address* )
   2. Divide the USPS verified addresses into 1 of 3 categories: Perfectly matched, Corrected, Unmatched
      a. Rerun the USPS-Corrected addresses through geocoder
         i. Subset the unmatched (or unacceptably matched) addresses
      b. Verify the USPS-Perfectly matched (III-A-2), USPS-Unmatched(III-A-2), and USPS-Corrected but ungeocoded(III-A-2-a-i) together using the following techniques simultaneously:
         i. A search engine’s mapping abilities (e.g., http://maps.google.com/) → helpful with the following
            1. Most everything USPS is helpful with (III-A-1-a,b,c. Not III-A-1-d. Will ‘correctly’ locate & identify undeliverable/nonexistent addresses if they are within a correct numeric range of a street)
2. Providing alternative names to roads (e.g., zooming into address will yield various names of road [US 23 or High street])

3. Providing alternate similar addresses in neighboring cities/zip codes (i.e., ‘123 Smith Rd, 43210’ yielding google-suggestion of ‘123 Smith Run Rd, 43201’)

4. Zooming into located address to:
   - Discern building footprint (helpful to couple with any county auditor website information)
   - Find location of local business with ‘known’ addresses to relatively compare to location proposed by google of observation address
   - If satisfied with google-located address as ‘true’ accurate location (usually if following criteria are satisfied: 1) google locates address without much suggestion for change from original input 2) Verified county auditor match with observation name and owner of parcel – or – 1) google locates approximate address [observation address is just out of range] 2) Query of road file [III-A-2-b-iii] returns valid address or address alternative suggested by google [III-A-2-b-i-2,3] for observation-provided or USPS-corrected zipcode but close, non-included address range) then, right-clicking on address/parcel will prompt option of ‘What’s here’ which returns location’s latitude and longitude.

ii. County auditor search \(\rightarrow\) helpful with the following

1. Locating parcel by name (if name information is available & parcel is owned by individual) to determine address
   - Providing building ‘sketch’ or more accurate (usually) remotely sensed image to determine building footprint (helpful in coupling with search engine mapping abilities [III-A-2-b-i])

2. Locating parcel by address to determine ownership
   - Determine if address is legitimate within auditor website
   - Determine if address is a rental property (useful in rechecking of address to determine if individual is a renter or who the landlord is for indirect geocoding to landlord’s rental property)
• Providing building ‘sketch’ or more accurate (usually) remotely sensed image to determine building footprint (helpful in coupling with search engine mapping abilities [III-A-2-b-i])

3. Those websites with interactive GIS’s provide similar capabilities to google (III-A-2-b-i-4)
   • Zoomed in panning of area
   • Determination of parcel boundaries for subsequent placing of google-lat/long
   • Identification of nearby possible addresses to conceivably match with observation-provided or google-suggested address

iii. Query of road file used in geocoder → helpful with the following
   1. Ordering the street names (not numbers) and subsetting the whole file (to ease visual search since original file will likely be large [~800,000 observations for ohio road file]) by observation-provided or USPS-corrected zipcode for subsequent visual search of observation-provided street names or google-suggested alternatives (useful when coupled with information from III-A-2-b-i)
      • Allows one to visually see how the road file (and geocoder) is naming a given road (e.g., hyphenated, spaced, spelling, street name variant in use, etc)
      • Allows one to visually see the street number ranges attributed to each road name
   2. Ordering the street names (not numbers) and subsetting the whole file by zipcode, then subsetting again by any address range with the observation-provided street numbers or google-suggested alternatives for subsequent visual search of observation-provided street names or google-suggested alternatives
   3. Subsetting the file by observation-provided or USPS-corrected zipcode to determine if road file even has any roads for this zipcode

   c. Collate the alternative addresses found by the combined efforts of III-A-2-b-i,ii,iii for subsequent regeocoding
      i. Purpose of regeocoding instead of simply utilizing google-latitude/longitude (III-A-2-b-i-4) to reduce bias as much as possible between differences of lat/long attribution to a given address in geocoder compared to that of given address in google
ii. Subset those unmatched still*
d. Subset those addresses without acceptable valid combinations of techniques in III-A-2-b
   (successful google-suggestion with unsuccessful auditor search/road file query; unsuccessful
   returns of any valid III-A-2-b techniques; etc)*
APPENDIX B: A summary of the Variables Considered in any Modeling Investigating Area-level Effects
<table>
<thead>
<tr>
<th>Level</th>
<th>Data Set(s)</th>
<th>Data Source(s)</th>
<th>Variable</th>
<th>Year(s)</th>
<th>Classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>Behavioral Risk Factor Surveillance System</td>
<td>Tobacco Smoking</td>
<td>2002-2008</td>
<td>Z-Score</td>
<td>% of adults reporting ≥ 100 lifetime cigarettes and currently smokers</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>Centers for Disease Control and Prevention, National Center for Hepatitis, HIV, STD, and TB Prevention</td>
<td>Chlamydia Rate</td>
<td>2007</td>
<td>Z-Score</td>
<td>Chlamydia rate per 100,000 population</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>Vital Statistics, National Center for Health Statistics</td>
<td>Teen Birth Rate</td>
<td>2000-2006</td>
<td>Z-Score</td>
<td>Teen birth rate per 1,000 female population, (15-19 years)</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>U.S. Census Bureau/Current Population Survey—Small Area Health Insurance Estimates</td>
<td>Percent without Health Insurance</td>
<td>2005</td>
<td>Z-Score</td>
<td>% of population &lt; 65 years without health insurance</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>Health Resources and Services Administration, Area Resource File</td>
<td>Primary Care Provider Rate</td>
<td>2006</td>
<td>Z-Score</td>
<td>Primary care provider rate per 100,000 population</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>National Center for Education Statistics</td>
<td>High School Graduation Rate</td>
<td>2005-2006</td>
<td>Z-Score</td>
<td>Averaged Freshman Graduation rate (% of 9th grade cohort that graduates in 4 years)</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>American Community Survey</td>
<td>Percent with College Degrees</td>
<td>2005-2007</td>
<td>Z-Score</td>
<td>% of population ≥ 25 years with 4-year college degrees or higher</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>Behavioral Risk Factor Surveillance System and American Community Survey</td>
<td>Family and Social Support</td>
<td>2005-2008; 2005-2007</td>
<td>Z-Score</td>
<td>UWPHI/RWJF-created composite of % of adults without social or emotional support and % of households that are single-parent</td>
</tr>
<tr>
<td>County</td>
<td>2010 County Health Rankings and Roadmaps Program</td>
<td>Current Population Survey and Small Area Income and Poverty Estimates</td>
<td>Percent of Impoverished Children</td>
<td>2007</td>
<td>Z-Score</td>
<td>% of all children (5-17 years) living in families with incomes &lt; 100% the federal poverty level</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>County</td>
<td>2012 County Health Rankings and Roadmaps Program</td>
<td>National Center for Chronic Disease Prevention and Health Promotion, from Behavioral Risk Factor Surveillance System</td>
<td>Physical Inactivity</td>
<td>2009</td>
<td>Z-Score</td>
<td>% of adults reporting no leisure time physical activity</td>
</tr>
<tr>
<td>County</td>
<td>2012 County Health Rankings and Roadmaps Program</td>
<td>U.S. Census Zip Code Business Patterns</td>
<td>Recreational Facility Rate</td>
<td>2009</td>
<td>Z-Score</td>
<td>Recreational facilities per 100,000 Population</td>
</tr>
<tr>
<td>County</td>
<td>Rural-Urban Continuum Codes</td>
<td>U.S. Department of Agriculture</td>
<td>Urban</td>
<td>2003</td>
<td>Dichotomy</td>
<td>Code=1 (Part of Metropolitan Statistical Area with ≥ 1,000,000 Population) versus codes 2-9 (i.e., all nonmetro and smaller metro)</td>
</tr>
<tr>
<td>County</td>
<td>County Typoogy Codes</td>
<td>U.S. Department of Agriculture</td>
<td>High Mining Employment</td>
<td>2004</td>
<td>Dichotomy</td>
<td>1998-2000 average annual earnings from mining industry ≥ 15% total earnings</td>
</tr>
<tr>
<td>County</td>
<td>County Typoogy Codes</td>
<td>U.S. Department of Agriculture</td>
<td>High Manufacturing Employment</td>
<td>2004</td>
<td>Dichotomy</td>
<td>1998-2000 average annual earnings from manufacturing industry ≥ 25% total earnings</td>
</tr>
<tr>
<td>School District</td>
<td>Small Area Income and Poverty Estimates</td>
<td>U.S. Census Bureau</td>
<td>Percent of Impoverished Children</td>
<td>1995</td>
<td>Z-Score</td>
<td>% of all children (5-17 years) living in families with incomes &lt; 100% the federal poverty level</td>
</tr>
<tr>
<td>School District</td>
<td>interactive Local Report Card</td>
<td>Ohio Department of Education</td>
<td>Pupil Support Expenditures per a Pupil</td>
<td>1994</td>
<td>Z-Score</td>
<td>Pupil support expenditures (i.e., support services, guidance services, health services, psychological services, speech pathology, audiology services, social/work activities, and student instructional media services) / Average Daily Membership</td>
</tr>
<tr>
<td>School District</td>
<td>interactive Local Report Card</td>
<td>Ohio Department of Education</td>
<td>Pupil Support Expenditures per a Pupil</td>
<td>2004</td>
<td>Z-Score</td>
<td>Pupil support expenditures (i.e., support services, guidance services, health services, psychological services, speech pathology, audiology services, social/work activities, and student instructional media services) / Average Daily Membership</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------</td>
<td>------------------------------</td>
<td>---------------------------------------</td>
<td>------</td>
<td>---------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>School District</td>
<td>interactive Local Report Card</td>
<td>Ohio Department of Education</td>
<td>Instructional Expenditures per a Pupil</td>
<td>1994</td>
<td>Z-Score</td>
<td>Instructional expenditures (i.e., salaries of teachers, teacher aides, or paraprofessionals, as well as materials, computers, books and other consumables used with students within the classroom) / Average Daily Membership</td>
</tr>
<tr>
<td>School District</td>
<td>interactive Local Report Card</td>
<td>Ohio Department of Education</td>
<td>Instructional Expenditures per a Pupil</td>
<td>2004</td>
<td>Z-Score</td>
<td>Instructional expenditures (i.e., salaries of teachers, teacher aides, or paraprofessionals, as well as materials, computers, books and other consumables used with students within the classroom) / Average Daily Membership</td>
</tr>
<tr>
<td>School District</td>
<td>interactive Local Report Card</td>
<td>Ohio Department of Education</td>
<td>Total Expenditures per a Pupil</td>
<td>1994</td>
<td>Z-Score</td>
<td>Total expenditures (i.e., Administration, building operations support, staff support, pupil support, and instructional support) / Average Daily Membership</td>
</tr>
<tr>
<td>School District</td>
<td>interactive Local Report Card</td>
<td>Ohio Department of Education</td>
<td>Total Expenditures per a Pupil</td>
<td>2004</td>
<td>Z-Score</td>
<td>Total expenditures (i.e., Administration, building operations support, staff support, pupil support, and instructional support) / Average Daily Membership</td>
</tr>
<tr>
<td>School District</td>
<td>Taxable Property Values by School District, Taxes Levied and Tax Rates for Current Expenses, and Average Property Values per Pupil</td>
<td>Ohio Department of Taxation</td>
<td>Total Land Value per a Pupil</td>
<td>1990</td>
<td>Z-Score</td>
<td>Value of all property subject to taxation and after application of assessment rates / Average Daily Membership</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>School District</td>
<td>Taxable Property Values by School District, Taxes Levied and Tax Rates for Current Expenses, and Average Property Values per Pupil</td>
<td>Ohio Department of Taxation</td>
<td>Total Land Value per a Pupil</td>
<td>2004</td>
<td>Z-Score</td>
<td>Value of all property subject to taxation and after application of assessment rates / Average Daily Membership</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Rural-Urban Commuting Areas</td>
<td>U.S. Department of Agriculture</td>
<td>High Mobility</td>
<td>2000</td>
<td>Dichotomy</td>
<td>Code≤5 (i.e., all metropolitan, micropolitan core, and micropolitan high commuting) and code=8 (small town high commuting) versus codes=5.1-7.4,8.1-10.6 (i.e., decreased commuting to urban areas or commuting from non-metropolitan origin)</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>Commute Time</td>
<td>2000</td>
<td>Z-Score</td>
<td>Mean work travel time (minutes) among workers ≥ 16 years</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>Percent Employed in Manufacturing</td>
<td>2000</td>
<td>Z-Score</td>
<td>% of total civilian population ≥ 16 years employed within the manufacturing industry</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>Percent Unemployed</td>
<td>2000</td>
<td>Z-Score</td>
<td>% unemployed among population ≥ 16 years in the civilian labor force</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>Percent Impoverished</td>
<td>2000</td>
<td>Z-Score</td>
<td>% of total population with incomes &lt; 100% Federal Poverty Level, 1999</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>------</td>
<td>---------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>Percent without a High School Diploma</td>
<td>2000</td>
<td>Z-Score</td>
<td>% of population ≥ 25 years without a high school diploma or equivalent</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>Percent without a Bachelor's Degree</td>
<td>2000</td>
<td>Z-Score</td>
<td>% of population ≥ 25 years without a 4-year college degree</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Census of Population and Housing, Summary File 4</td>
<td>U.S. Census</td>
<td>SEP Composite</td>
<td>2000</td>
<td>Z-Score</td>
<td>Principal component analysis of median household income, percent unemployed, percent impoverished, percent not attaining a high school diploma, and percent not attaining a bachelor's degree (as z-scores)</td>
</tr>
</tbody>
</table>
APPENDIX C: Generic Notational Representation of the Multilevel Logistic Regression Equation
\[ p(y|\beta, \lambda) = \prod_{j=1}^{J} \prod_{i=1}^{N_j} p(y_{ij}|\beta, \lambda_j) \quad \text{where} \quad p(y_{ij}|\beta, \lambda_j) \sim \text{Binomial}(n,p) \]

\[ \ln(p_{ij} / [1-p_{ij}]) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \lambda_j \]

\[ p(\lambda|\beta^C, \tau^2) = \prod_{j=1}^{J} p(\lambda_j|\beta^C, \tau^2) \quad \text{where} \quad p(\lambda_j|\beta^C, \tau^2) \sim \text{Normal}(\mu, \tau^2) \]

\[ \mu_j = \beta_{1}^C X_{j1} + \beta_{2}^C X_{j2} + \ldots + \beta_{n}^C X_{jn} \]

Generic notational representation of the multilevel logistic regression equations, where within the individual-level (superscript ‘I’) \( p_{ij} \) is the probability of invasive cervical cancer incidence (ICCI) for the given covariate pattern \( i \) within area \( j \); \( \beta_0 \) is the overall intercept of the model (i.e., global odds of ICCI); \( \beta_1 - \beta_3 \) are the regression coefficients of variables \( X_1 - X_3 \); \( X_1 \) is the variable age (<39 years, between 39 years and 53 years, >53 years), \( X_2 \) is the variable race (White, Black, other), \( X_3 \) is the variable resulting from a multiplicative interaction between age and race. Within the area-level (superscript ‘C’) \( \lambda_j \) is the area-level random effect, \( \beta_1, \beta_2, \ldots, \beta_n \) are the regression coefficients of the area-level variables, and \( \tau^2 \) is the variance of the area-level effects. ICCI odds ratios can be calculated from resulting model estimates by exponentiating the regression coefficients or differences between random intercepts (i.e., \( e^{[\lambda_j + \lambda^2]} \)– exponentiation of any intercept-associated estimate (e.g., \( \beta_0 \) or \( [\beta_0 + \lambda_j] \)) simply yields an estimate of ICCI odds.
APPENDIX D: Model Estimates Resulting from a Sensitivity Analysis of the Multilevel and Spatial Investigation of Invasive Cervical Cancer Incidence in Ohio that Restricted to the Appalachian Region
<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual-level</strong></td>
<td>OR 95% CI</td>
<td>OR 95% CI</td>
</tr>
<tr>
<td>Intercept(^a)</td>
<td>0.62 (0.55-0.69)</td>
<td>0.39 (0.33-0.46)</td>
</tr>
<tr>
<td>Age(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.50 (0.24-1.03)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.47 (0.21-1.08)</td>
<td></td>
</tr>
<tr>
<td>At 39 years ≤ Age ≤ 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.82 (0.50-1.43)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.64 (0.18-2.24)</td>
<td></td>
</tr>
<tr>
<td>At Age &gt; 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2.10 (1.21-3.63)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.23 (0.05-1.02)</td>
<td></td>
</tr>
<tr>
<td>County-level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlamydia Rate (per increase of 145.0 cases per 100,000 persons)</td>
<td>0.93 (0.77-1.12)</td>
<td></td>
</tr>
<tr>
<td>Percent Uninsured (per a 1.55 percent increase of uninsured)</td>
<td>1.08 (0.99-1.18)</td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Birth Rate (per increase of 12.9 births per 100,000 persons)</td>
<td>1.13 (0.95-1.34)</td>
<td></td>
</tr>
<tr>
<td>At 39 years ≤ Age ≤ 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Birth Rate (per increase of 12.9 births per 100,000 persons)</td>
<td>1.29 (1.07-1.55)</td>
<td></td>
</tr>
<tr>
<td>At &gt; 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Birth Rate (per increase of 12.9 births per 100,000 persons)</td>
<td>1.25 (1.06-1.48)</td>
<td></td>
</tr>
<tr>
<td>At Teen Birth Rate = 50.9 per 100,000 persons and Race = White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.26 (1.86-2.73)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.60 (1.33-1.92)</td>
<td></td>
</tr>
<tr>
<td>At Teen Birth Rate = 34.1 per 100,000 persons and Race = White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>1.91 (1.48-2.46)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.40 (1.10-1.78)</td>
<td></td>
</tr>
</tbody>
</table>

Table D.1. Effects of demographic and county-level factors on invasive cervical cancer incidence in Appalachia, Ohio, 1996-2009

cont’d
Table D.1 (cont’d)

<table>
<thead>
<tr>
<th>At Teen Birth Rate = 50.9 per 100,000 persons and Race = African American</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.72 (1.51-9.16)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>6.74 (2.75-16.56)</td>
<td></td>
</tr>
<tr>
<td>At Teen Birth Rate = 34.1 per 100,000 persons and Race = African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.15 (1.28-7.78)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>5.92 (2.41-14.55)</td>
<td></td>
</tr>
<tr>
<td>At Teen Birth Rate = 50.9 per 100,000 persons and Race = Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.05 (0.69-13.52)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>0.78 (0.14-4.23)</td>
<td></td>
</tr>
<tr>
<td>At Teen Birth Rate = 34.1 per 100,000 persons and Race = Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.58 (0.57-11.62)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>0.68 (0.12-3.75)</td>
<td></td>
</tr>
</tbody>
</table>

| County-level Variance | 0.0507^d | 0.0298^d |
| Intraclass Correlation Coefficient | 0.0152 | 0.0090 |
| -2 Log Likelihood | 5256.2 | 5135.6 |

^aThe intercept may be interpreted as the odds of cervical cancer at the value of ‘0’ for all other covariates in the model
^bOdds ratio involving age is dependent on individual-level race and county-level teen birth rate variables
^cStatistical significance based on a mixture of chi-square distributions
^dp<0.05
^eEvidence of county-level residual spatial autocorrelation suggests underestimation of county-level estimate’s standard errors
<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual-level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.57 (0.53-0.63)</td>
<td>0.32 (0.27-0.37)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.18 (1.83-2.60)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.55 (1.31-1.83)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.59 (1.47-8.77)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>6.56 (2.63-16.36)</td>
<td></td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.64 (0.61-21.81)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>0.88 (0.15-5.14)</td>
<td></td>
</tr>
<tr>
<td>School District-level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxable Land Value, 2004 (per increase of $176,401 per pupil)</td>
<td>1.01 (0.62-1.62)</td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Impoverished Children (per increase of 8.2 percent)</td>
<td>1.27 (1.16-1.39)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Impoverished Children (per increase of 8.2 percent)</td>
<td>1.10 (0.82-1.46)</td>
<td></td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Impoverished Children (per increase of 8.2 percent)</td>
<td>1.76 (0.92-3.36)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 15.7 and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.47 (0.19-1.15)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.29 (0.09-0.97)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 5.9 and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.56 (0.18-1.70)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.20 (0.03-1.23)</td>
<td></td>
</tr>
</tbody>
</table>

Table D.2. Effects of demographic and school district-level factors on invasive cervical cancer incidence in Appalachia, Ohio, 1996-2009

*cont’d*
<table>
<thead>
<tr>
<th>At Percent Impoverished Children = 15.7 and 39 years ≤ Age ≤ 53 years</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.92 (0.34-2.50)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.49 (0.12-2.01)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 5.9 and 39 years ≤ Age ≤ 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.92 (0.34-2.50)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.33 (0.06-1.74)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 15.7 and Age &gt; 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2.00 (1.06-3.77)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.17 (0.03-0.87)</td>
<td></td>
</tr>
<tr>
<td>At Percent Impoverished Children = 5.9 and Age &gt; 53 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2.37 (1.01-5.55)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.11 (0.01-0.88)</td>
<td></td>
</tr>
<tr>
<td>School District-level Variance</td>
<td>0.1152d</td>
<td>0.0689d</td>
</tr>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>0.0338</td>
<td>0.0205</td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>5253.5</td>
<td>5098.7</td>
</tr>
</tbody>
</table>

*aThe intercept may be interpreted as the odds of cervical cancer at the value of ‘0’ for all other covariates in the model

*bOdds ratio involving race is dependent on individual-level age and school district-level childhood poverty rate variables

*cStatistical Significance based on a mixture of chi-square distributions

*d\(p<0.05\)

*eEvidence of school district-level residual spatial autocorrelation suggests underestimation of school district-level estimate’s standard errors
<table>
<thead>
<tr>
<th>Individual-level</th>
<th>Null Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.58</td>
<td>0.34</td>
</tr>
<tr>
<td>OR</td>
<td>0.54-0.62</td>
<td>0.30-0.39</td>
</tr>
<tr>
<td>Race&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.21 (1.85-2.64)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>1.56</td>
<td>1.32-1.85</td>
</tr>
<tr>
<td>At Race=African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>3.78 (1.53-9.36)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>7.37</td>
<td>2.92-18.59</td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age &lt; 39 years</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>39 years ≤ Age ≤ 53 years</td>
<td>2.75 (0.56-13.61)</td>
<td></td>
</tr>
<tr>
<td>Age &gt; 53 years</td>
<td>0.66</td>
<td>0.11-3.79</td>
</tr>
<tr>
<td>Census Tract-level Interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=White</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Position (per 1 standard deviation increase)</td>
<td>0.59 (0.52-0.68)</td>
<td></td>
</tr>
<tr>
<td>At Race=African American</td>
<td>0.77 (0.50-1.19)</td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Position (per 1 standard deviation increase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Race=Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic Position (per 1 standard deviation increase)</td>
<td>1.11 (0.42-2.94)</td>
<td></td>
</tr>
<tr>
<td>At 75&lt;sup&gt;th&lt;/sup&gt;% of Socioeconomic Position and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.43</td>
<td>0.15-1.23</td>
</tr>
<tr>
<td>Other</td>
<td>0.85</td>
<td>0.22-3.29</td>
</tr>
<tr>
<td>At 25&lt;sup&gt;th&lt;/sup&gt;% of Socioeconomic Position and Age &lt; 39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.30</td>
<td>0.14-0.65</td>
</tr>
<tr>
<td>Other</td>
<td>0.93</td>
<td>0.22-3.99</td>
</tr>
</tbody>
</table>

Table D.3. Effects of demographic and census tract-level factors on invasive cervical cancer incidence in Appalachia, Ohio, 1996-2009

cont'd
Table D.3 (cont’d)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>At 75th % of Socioeconomic Position and 39 years ≤ Age ≤ 53 years</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.73 (0.30-1.77)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.06 (0.27-4.11)</td>
<td></td>
</tr>
<tr>
<td><strong>At 25th % of Socioeconomic Position and 39 years ≤ Age ≤ 53 years</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0.52 (0.29-0.93)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.16 (0.28-4.73)</td>
<td></td>
</tr>
<tr>
<td><strong>At 75th % of Socioeconomic Position and Age &gt; 53 years</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2.02 (0.93-4.42)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.36 (0.07-1.87)</td>
<td></td>
</tr>
<tr>
<td><strong>At 25th % of Socioeconomic Position and Age &gt; 53 years</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.43 (0.82-2.51)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.39 (0.07-2.16)</td>
<td></td>
</tr>
<tr>
<td>Census Tract-level Variance^c</td>
<td>0.1223^d</td>
<td>0.0850^d</td>
</tr>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>0.0358</td>
<td>0.0252</td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>5268.3</td>
<td>5101.0</td>
</tr>
</tbody>
</table>

^a The intercept may be interpreted as the odds of cervical cancer at the value of ‘0’ for all other covariates in the model
^b Odds ratio involving race is dependent on individual-level age and school district-level childhood poverty rate variables
^c Statistical Significance based on a mixture of chi-square distributions
^d p<0.05
^e Evidence of census tract-level residual spatial autocorrelation suggests underestimation of census tract-level estimate’s standard errors
APPENDIX E: Maps Resulting from Neighborhood Definition Sensitivity Analysis of the Multilevel and Spatial Investigation of Invasive Cervical Cancer Incidence in Ohio
Figure E.1. County-level random intercepts of the cervical cancer county null model. Random intercepts are shaded by quintiles. Random intercepts of hatched county are statistically significantly different from zero. Random intercepts of stippled counties are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.
Figure E.2. County-level random intercepts of the cervical cancer county final model. Random intercepts are shaded by quintiles. Random intercepts of hatched county are statistically significantly different from zero. Random intercepts of stippled counties are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.
Figure E.3. School district-level random intercepts of the cervical cancer school district null model. Random intercepts are shaded by quintiles. Random intercepts of hatched school districts are statistically significantly different from zero. Random intercepts of stippled counties are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.

Random Intercepts are shaded by quintiles. Random intercepts of hatched school districts are statistically significantly different from zero. Random intercepts of stippled and blue-outlined school districts are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.

Figure E.4. School district-level random intercepts of the cervical cancer school district final model. Random intercepts are shaded by quintiles. Random intercepts of hatched school districts are statistically significantly different from zero. Random intercepts of stippled and blue-outlined school districts are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.
Figure E.5. Census tract-level random intercepts of the cervical cancer census tract null model. Random intercepts are shaded by quintiles. Random intercepts of hatched census tracts are statistically significantly different from zero. Random intercepts of stippled and blue-outlined census tracts are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.
Figure E.6. Census tract-level random intercepts of the cervical cancer census tract final model. Random intercepts are shaded by quintiles. Random intercepts of hatched census tracts are statistically significantly different from zero. Random intercepts of stippled and blue-outlined census tracts are statistically significantly high (low) values within a neighborhood (queen contiguity) of high (low) values.
Bibliography


(34) Ohio Department of Taxation. 2004 School District Taxable Property Values, Taxes Levied and Tax Rates for Current Expenses, and Average Property Values per Pupil. Ohio Department of Taxation; Columbus, Ohio, 2005; Available at: http://www.tax.ohio.gov/portals/0/tax_analysis/tax_data_series/school_district_data/sd1/sd1cy04.xls, 2013.

(35) Ohio Department of Taxation. 1990 School District Taxable Property Values, Taxes Levied and Tax Rates for Current Expenses, and Average Property Values per Pupil.
Ohio Department of Taxation; Columbus, Ohio. Available at: http://www.tax.ohio.gov/portals/0/tax_analysis/tax_data_series/school_district_data/sd1/sd1cy90.wk1, 2012.


(61) Lane B. Still Breathing. The Columbus Dispatch 2008 April 13, 2008;Home.


(64) Brems B. Percent of Uninsured Adults in Ohio. 2009(Ohio).


(234) United States Behavioral Risk Factor Surveillance System (2010). Prevalence Data: Women aged 18+ who have had a pap test within the past three years, Centers for Disease Control.


(335) OhioLINK. OhioLINK Library Catalog. Unknown; Available at: http://olc1.ohiolink.edu/search/, 2013.


(390) Hamilton BE, Ventura SJ. Birth rates for US teenagers reach historic lows for all age and ethnic groups. : US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2012.


(394) Cohen-Cline H. Patterns of obesogenic neighborhood features and residential property values. 2012.

(395) Rehm CD, Moudon AV, Hurvitz PM, Drewnowski A. Residential property values are associated with obesity among women in King County, WA, USA. Soc.Sci.Med. 2012.


