IN THE WRONG PLACE?:
GEOGRAPHIC VARIATION IN U.S. OCCUPATIONAL INJURY / ILLNESS RATES

By:
Roni A. Neff, SM

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland
September, 2006

UMI Number: 3243310

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

DISSERTATION ABSTRACT

BACKGROUND: Around the world and across U.S. counties, workers and businesses operate in a diverse landscape of demographics, economy, culture, policy and industry. This dissertation presents four papers exploring geographic variation in U.S. occupational injury/illness rates.

METHODS: The literature on geographic variation in occupational injury/illness is reviewed and categorized. Three papers examine geographic variation in the OSHA Data Initiative (ODI), 1997-2001, a database of high injury/illness industries. The first presents surveillance tools including mapping, spatial statistics, and ranking. The second uses multilevel regression to examine social determinants of county-level variation in lost workday injury/illness rates (LWDII). Finally, a case study of the meat processing industry uses mapping and regression to explore risk factors associated with both establishment location and high-LWDII establishments.

RESULTS: 1) There is a small, uncoordinated literature using geographic methods to examine occupational injury/illness. 2) There is geographic variation in occupational injury/illness rates. The sample mean LWDII was 7.22 per 100 workers (county range: 0, 25.2). The five highest rate states were Vermont (9.77), West Virginia (9.76), Michigan (9.67), Maine (9.54) and Kentucky (8.99). Rates were low throughout the South.

3) Geographic variation was associated with social risk factors. In regressions, high rates were positively associated with county poverty, percent Caucasian, unionization, strong safety net, and industry hazard. Meat establishment locations were associated with county percent African American, non-college educated, longterm job gain, and urbanicity, plus

state-level anti-union policy, medium union membership, and slightly reduced OSHA inspections. By contrast, *high-LWDII* meat establishments were associated with county percent Caucasian, low income, high school education, and longterm job loss. 4) There is suggestive evidence of substantial, biased underreporting in the ODI.

CONCLUSIONS: Explanations for the findings are discussed. Recommendations focus on addressing underreporting, generating more county-level occupational injury/illness data, promoting county-level surveillance, increasing geographic research in occupational injury/illness, piloting programs for geographic targeting, and changing business and worker incentives and capacity for prevention.

Advisor:
Thomas A. Burke, PhD, MPH

Readers:
Susan Baker, MPH
Frank Curriero, PhD
Lynn Goldman, MD, MPH
Genevieve Matanoski, MD, DrPH

Alternates:
Margaret Ensminger, PhD
Mary Fox, PhD

ACKNOWLEDGEMENTS

I dedicate this dissertation in memory and in honor of those whose workplace was in "the wrong place." Hopefully, this work will contribute to reducing their numbers.

I have been so lucky to have Tom Burke as a mentor, advisor, and friend. It has been a pleasure to work with, learn from, and debate with someone so wise, committed to the public's health, and also funny. I have grown as a result of his subtle guidance, example, and ever-present support. Frank Curriero also played a critical role in this dissertation, from teaching me GIS and spatial statistics to writing R code to his patient and thoughtful reviews of paper drafts. I thank my other committee members, Lynn Goldman, Sue Baker, and Gene Matanoski; alternates Peg Ensminger and Mary Fox; and three who could not be on the committee for logistical reasons, Nga Tran, Jeff Johnson, and Tom Glass. Each has contributed valuable ideas to my methods and interpretations. It is a gift to share such intelligent, thoughtful feedback on another's work.

I am indebted to my funders: NIOSH (fellowship with tuition and stipend support, Pilot Project Grant); the Center for a Livable Future (CLF) (Innovation Grant); the Centers for Disease Control and Prevention (Environmental Public Health Tracking Fellowship); and the Department of Health Policy and Management (tuition support.) I am so grateful to Sue Baker and Gordon Smith for the initial decision to bring me onto the NIOSH grant. Thanks also to Peter Lees, Gary Sorock, the CLF staff, Beth Resnick, and Tom Burke.

In terms of influences, Jill Litt's 1999 APHA presentation stimulated my interest in mapping and a lecture by Ellen Silbergeld led to the initial idea for geographic analysis of occupational injury/illness. I thank Jordan Barab for his Confined Space blog, a vital source of news and perspectives on occupational injury and illness, and always a good reminder of why we do this work. http://www.spewingforth.blogspot.com/

Many others have contributed to this dissertation. First, I sincerely thank Robin Dranbauer for all her administrative support and efforts on my behalf through my years at Hopkins. Nga Tran mentored and advised me while she was at Hopkins and afterwards. Tim Shields provided training and GIS assistance. Tom Louis was helpful and patient in sharing ranking methods and R code. Suchih Su and Norma Kanarek shared information about CART and Ingo Ruczinski shared code. Donna Hesson and others in the library were always helpful. John Sestito of NIOSH and Joseph Fraumeni of NCI both took the time to share history and perspectives with me. Debra Nathans assisted with coding meat establishments. Mary Sewell and Richelle Lewis helped with administrative issues. The ERC advisory group gave very insightful feedback. And Joe Dubois and Dave Schmidt of OSHA provided me with data and helpful insights.

My friends and colleagues have made this experience enjoyable. Cheers to current/former officemates Ramya Chari, Beth Dederick, Erin Dreyling, Keeve Nachman, and Robin Streeter, as well as Beth Resnick and the rest of the tracking gang. And a long distance I-miss-you to Kristen Chossek Malecki for sharing offices, theses, pregnancy, babies, and friendship too. I also thank my fellow occ injury students for good discussion and ideas, and the dissertation support group for valuable feedback and fabulous brunches.

I could never have written this thesis if I didn’t so love and trust our child care providers. I have infinite thanks and respect for Ruhi Beg, Belle Jackson, and the teachers in Downtown Baltimore Child Care Rooms 1, 2, 4, and 5.

Finally, I thank my family. My parents, Joanne and Martin Neff, have *always* been there for me, to cheer me on, debate, understand, and offer material and other support. Their role in my work spans from inspiring my interest in public health and social issues, to encouraging me to stick with science so I could become a doctor (look! I finally did!), to driving me from way out in Queens to Manhattan so I could get my term paper in on time (I promised I'd never forget that one!), to proofreading every single comma in this dissertation. My grandfather earned his PhD in his sixties and my mother in her fifties, so 39 seems downright youthful. I thank my sisters, Debra Nathans and Sally Neff, my best girlfriends, for their friendship, understanding, and laughs.

My husband, John McGready, has contributed so much to this thesis, from countless extra child care pickups, daily gourmet cooking, and relieving me of household tasks, to listening to my rants, to always being willing to answer one more "little" stats question. I thank him for his infinite sensitivity, unwavering support, much hilarity, and his insistence that I learn to appreciate grindcore and learn about the "real" geography of public health through long driving tours of Baltimore neighborhoods. "It was such a perfect day, I'm glad I spent it with you."

And my children, Micah and Emmet Neff-McGready. Their hugs and smiles nourish me. They have only known me as a graduate student. I hope they will learn that while work can be hazardous, it can also be fulfilling. And more importantly, I hope they will never stop playing – jumping, laughing, making art and music, and blowing raspberries (they can dispense with the pre-dawn rising, though.) Most importantly, I hope they will always know that they and their father are my rock, my heart.

TABLE OF CONTENTS

INTRODUCTION 1
I. CONCEPTUAL FRAMEWORK 1
II. LITERATURE REVIEW 4
III. RESEARCH GOALS 5
All Over the Map: 9
A Review of Literature on Geographic Variation in Occupational Injury & Illness . 9
ABSTRACT 9
I. INTRODUCTION 10
II. BACKGROUND 11
III. PRIOR RESEARCH 12
A. SURVEILLANCE REPORTS 13
B. HYPOTHESIS-GENERATING DESCRIPTIVE RESEARCH 19
C. RESEARCH CONTROLLING FOR GEOGRAPHY OR OCCUPATIONAL INJURY/ILLNESS 26
D. HYPOTHESIS-TESTING ANALYTIC RESEARCH 32
IV. CONCLUSION 38
"In the Wrong Place…?": 46
Spatial Tools for Occupational Injury/Illness Surveillance 46
ABSTRACT 46
I. INTRODUCTION 47
II. BACKGROUND 50
III. METHODS 53
A. METHODS - DATA 53
B. METHODS -- ANALYSES 57
IV. RESULTS 61
V. DISCUSSION 73
VI. CONCLUSION 81
Social Predictors of 84
County-Level Occupational Injury and Illness 84
ABSTRACT 84
I. INTRODUCTION 85
II. BACKGROUND 87
III. METHODS: 90
A. DATA 90
B) DATA ANALYSES 95
IV. RESULTS 98
V. DISCUSSION 107
VI. CONCLUSION 117
Occupational Injury and Illness in Meat Processing: 119
A Geographic Cut 119
ABSTRACT 119
I. INTRODUCTION 120
II. BACKGROUND 122

III. METHODS: 127
A. DATA 127
B) ANALYSIS: 131
IV. RESULTS 134
V. DISCUSSION 145
B. STRENGTHS AND LIMITATIONS 149
C. CONCLUSIONS 151
CONCLUSION 153
On Social Risk Factors 163
RECOMMENDATIONS: 166
REFERENCES 170
APPENDICES 194
CURRICULUM VITAE 242

LIST OF TABLES

Table 1: PubMed searches for articles about geography and occupational injury/illness. September 28, 2005 13
Table 2: Surveillance reports 17
Table 3: Hypothesis-generating/descriptive research 23
Table 4: Geography of Occupational Injury/Illness as Control Variable 28
Table 5: Theory-Based, Analytic Studies 34
Table 6: U.S. occupational injury/illness-related databases that might be used for geographic analysis 41
Table 7: OSHA Data Initiative database, 1997-2001: Establishment-level lost workday injury/illness (LWDII) rates in most high-rate industries 54
Table 8: Risk factors used in mapping and their distribution among ODI counties 56
Table 9: Table #: Highest- and lowest-LWDII rate states: Variations in top 10 industries (with the most ODI establishments.) 66
Table 10: States with the highest mean lost workday and injury rates in the ODI sample, 1997-2001. 67
Table 11: States with the lowest mean lost workday and injury rates in the ODI sample, 1997-2001 68
Table 12: Top 10 counties with highest LWDII based on direct means, 1997-2001 68
Table 13: Top 10 counties with highest mean LWDII based on Bayesian method adjusting ranks for variance of mean 69
Table 14: Top 10 counties with highest mean LWDII among those with >30 establishments in ODI 69
Table 15: OSHA Data Initiative database, 1997-2001: Establishment-level lost workday injury/illness (LWDII) rates in most high-rate industries 90

Table 16: Risk factors used in regression analysis and distribution among counties or states with observations in data set 94
Table 17: Effect of state and county risk factors on county mean LWDII, 1997-2001: bivariate, county-level and multilevel models 100
Table 18: Tabular description of CART regression tree based on risk factors in final model........ 107
Table 19: Meat Risk factor database........ 129
Table 20: OSHA Data Initiative meat industry data 134
Table 21: Significant predictors of meat industry locations based on logistic multi-level model*........ 140
Table 22: Associations between social risk factors and occupational injury illness rates based on county-level linear regression: All meat processing and each industry separately.* 142
Table 23: Indicators of underreporting: Reports of LWDII rates of "Zero"; Year-to-year LWDII drops of at least six points , By industry 143
Table 24: Indicators of underreporting: Reports of LWDII rates of "Zero"; Year-to-year LWDII drops of at least six points , By industry 155
Table 25: Summary of regression results from all-industry analyses. Outcome: mean county LWDII. 159
Table 26: Summary of regression results from the meat industry. Outcome: mean county LWDII. 159

LIST OF FIGURES

Figure 1: How geographic risk factors affect high risk employers and "all" employers.... 2
Figure 2: Conceptual model of how geographic risk factors affect occupational injury and illness rates........ 3
Figure 3: Moran's I test for global spatial dependence: Mean county LWDII compared to mean LWDII of surrounding counties 63
Figure 4: How geographic risk factors affect high risk employers and "all" employers.. 87
Figure 5: Conceptual model of how geographic risk factors affect occupational injury and illness 88
Figure 6: Semivariogram of standardized residuals from final GLLAMM model showing little remaining spatial dependence........ 103
Figure 7: CART regression tree based on the risk factors in the final regression model 106

LIST OF MAPS

Map 1: Mean county Lost Workday Injury/Illness Rate (LWDII), ODI 1997-2001, quintiles ... 6
Map 2: Mean state LWDII, ODI 1997-2001, quintiles ... 62
Map 3: Mean county LWDII, ODI 1997-2001, quintiles ... 63
Map 4: Index of Area Industry Hazard based on BLS data ... 64
Map 5: "Observed" (ODI) vs. "Expected" (Industry Hazard) LWDII ... 65
Map 6: High and Low LWDII Clusters (Local Moran's I) ... 67
Map 7: Mean county LWDII, Tertiles ... 70
Map 8: Distribution of LWDII in High Poverty Counties ... 71
Map 9: Distribution of LWDII in High % African American Counties ... 72
Map 10: Distribution of LWDII in High % White Counties ... 72
Map 11: Bureau of Labor Statistics, Survey of Occupational Injuries and Illnesses 2001 state LWDII rates (Left) ... 74
Map 12: ODI 1997-2001 mean LWDII rates (Right, tertile version of Map 2) ... 74
Map 13: Bureau of Labor Statistics 2001 state fatality rates (Left) ... 75
Map 14: Index of Area Industry Hazard [state version of Map 4] (Right) ... 75
Map 15: Mean county LWDII, Quintiles ... 99
Map 16: Standardized residuals from final model ... 104
Map 17: Meat processing establishments by industry ... 135
Map 18: Counties with meat processing establishments ... 136
Map 19: Top 1/3 of meat processing counties with highest mean injury/illness rates ... 136
Map 20: Distribution of longterm job loss and gain among counties with meat establishments. ... 137
Map 21: Distribution of longterm job loss and gain among high rate meat counties ... 138
Map 22: Distribution of percent African American among counties with meat establishments ... 139
Map 23: Distribution of % African American among meat counties with high mean injury/illness rates ... 139
Map 24: Standardized residuals from multilevel model measuring predictors of meat establishment locations ... 141
Map 25: Establishments reporting LWDII of “Zero” in one or more years ... 144
Map 26: Establishments reporting year-to-year drops >6 LWDII points ... 144
Map 27: Clusters of high and low rates in the OSHA Data Initiative 1997-2001 (Local Moran's I) ... 157
Map 28: Distribution of Lost Workday Injury/Illness Rates in High Poverty Counties 158

APPENDICES

APPENDIX 1: Top industries by number surveyed in OSHA Data Initiative (ODI) 1997-2001. 195
APPENDIX 2: Top industries by lost workday injury/illness rate (LWDII) in ODI, 1997-2001 196
APPENDIX 3: Maps of LWDII by year 197
APPENDIX 4: ODI sampling criteria by year and discussion/analysis of issue of dropping low rate establishments in subsequent years. 199
APPENDIX 5: Comparing Geocoded and Non-geocoded Records in the OSHA Data Initiative 201
APPENDIX 6: Index of Area Industry Hazard: Calculations 203
APPENDIX 7: Comparing the Index of Area Industry Hazard based on the Bureau of Labor Statistics, Survey of Occupational Injuries and Illnesses (SOII) data with an index based on ODI data itself 205
APPENDIX 8: Map of difference between ODI Index ("expected") and ODI LWDII ("observed") rates 207
APPENDIX 9: Steps for performing surveillance analyses 209
APPENDIX 10: List of risk factors and sources 214
APPENDIX 11: Risk factors discussed in regression results section, with: means; ranges; coefficients from bivariate regression versus mean county LWDII. Table of correlation coefficients. Variance Inflation Factors 218
APPENDIX 12: R Code for Calculating Semivariogram 223
APPENDIX 13: Linear regressions by year and at establishment level 224
APPENDIX 14: Regression using the outcome measure, "average county difference between observed and expected mean LWDII" (with expected mean defined as the Index of Area Industry Hazard.) 229
APPENDIX 15: Regression separately for Southern and non-Southern states and for top quarter high-LWDII industry vs. others 231
APPENDIX 16: Matching methods for meat database 234
APPENDIX 17: Meat: Establishments by size 235
APPENDIX 18: Meat: Establishments with LWDII >8 and others, by industry 237
APPENDIX 19: Correlations between variables used in meat analyses 239

1

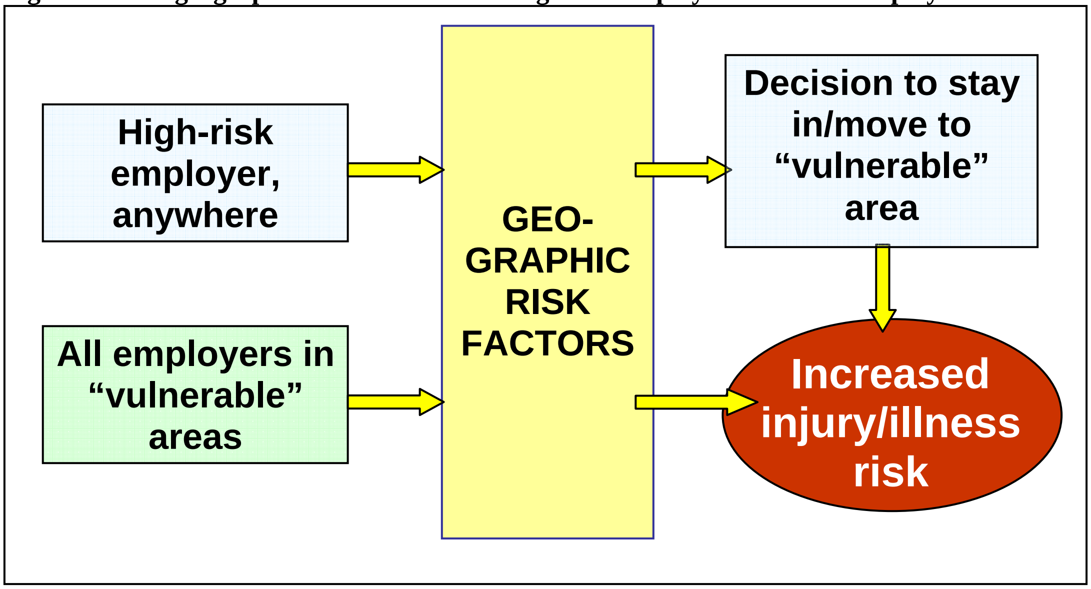
INTRODUCTION

When a worker's injury or illness is attributed to bad luck, it is often said that he or she was just "in the wrong place." But what if a place is "wrong" for a lot of workers? This thesis builds from the premise that most work-related injury and illness can be prevented, and that the more effectively public health programs can characterize the determinants, variation, and extent of health conditions, the better they will be at shaping, targeting and evaluating interventions and at communicating with the public and policymakers about needs and priorities. Four papers are presented exploring geographic variation in U.S. occupational injury and illness rates.

I. CONCEPTUAL FRAMEWORK

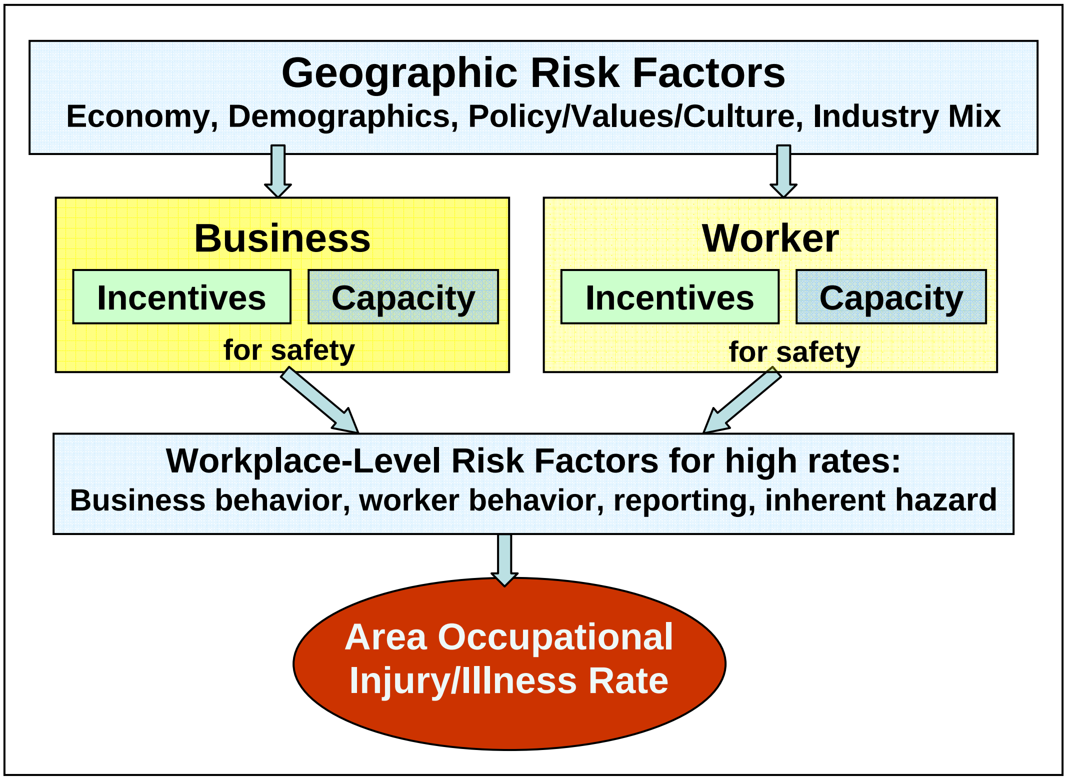
Around the world and across counties in the United States, workers and businesses operate in a diverse landscape of demographics, economy, culture, policy and industry. As shown in Figure 1, these local conditions play into business decisions about location, making risky and "bad actor" employers more likely to locate or stay in "vulnerable" areas where they are better able to get away with relatively low levels of worker protections. At the same time, *all* employers and workers are affected by geographic risk factors wherever they are. As described in Figure 2, these conditions often have implications for occupational injury/illness rates.

Figure 1: How geographic risk factors affect high risk employers and "all" employers



For this research, "geographic risk factors" are defined as factors measured at the area level (such as the county or state) rather than the individual or workplace level. They reflect both aggregated individual effects and those derived from the overall conditions in an area. Both within the U.S. and internationally, they include the economy; demographics; worker-friendliness of policy, values, and culture; and industry mix. These risk factors themselves are driven by policy choices that can be changed. Figure 2 presents my conceptual model for how geographic risk factors exert their effects. It shows that they drive business and worker incentives and capacity to avoid risk (by a variety of mechanisms including their impact on the balance of power between business and workers) (Robinson, 1988). In turn, business and worker incentives and capacity affect workplace-level behavior, which is the proximal cause of injury/illness.

Figure 2: Conceptual model of how geographic risk factors affect occupational injury and illness rates



Workplace level "behaviors" include a wide range of actions directly and indirectly related to occupational injury and illness, including industrial hygiene controls, safety and health programs, management commitment to safety, safety attitudes and carefulness, decisions about establishment location and what the establishment will produce or do, decisions about reporting – and also behaviors that unintentionally affect injuries/illnesses, including those related to work organization or job stress.

With adequate motivation, firms can change their injury/illness rates. For example, Shannon and Vidmar estimate that 42 percent of lost worktime injuries could be cut if all Ontario businesses matched the safety performance of the top 25th percentile for their industries (Shannon & Vidmar, 2004). Yet, it should be acknowledged that different establishment injury/illness rates may also reflect different levels of inherent hazard;

industry categories are broad markers for establishment activities. Establishment rates may also be tarnished by underreporting (Azaroff et al., 2002; Azaroff et al., 2004; Conway & Svenson, 1998; Leigh & Robbins, 2004; Leigh et al., 2004; Pollack & Keimig, 1987; Pransky et al., 1999; Rosenman et al., 2006; Smith, 2003; Smith et al., 2005). Analyses by Leigh et al (2004) and Rosenman et al (2006) estimate that nationally up to 2/3 of injuries and illnesses may be missed in surveillance systems, including due to underreporting. Azaroff et al (2004) documented powerful economic, demographic, political and social factors in underreporting, which may contribute to geographic variation in reporting rates.

II. LITERATURE REVIEW

Today, geographic analysis has become mainstream in public health (Cromley & McLafferty, 2002; Hillemeier et al., 2003; Krieger et al., 2002; Krieger et al., 2003; Waller & Gotway, 2004). The U.S. government has set the objective of “Increas[ing] the proportion of all major national, state and local health data systems that use geocoding [geographic identifiers that enable mapping] to promote nationwide use of geographic information systems (GIS) at all levels.” The target is 90 percent of all data systems (Objective 23-3) (U.S. Department of Health and Human Services, 2000).

Historically, the occupational injury/illness field was a leader in using geographic surveillance methods. But in the last two decades it has taken little advantage of the explosion of new tools, methods and databases. Few databases gather or make accessible geo-referenced data to facilitate such explorations, particularly below the state level.

Overall, the literature on this topic is small. Few examined articles use spatial statistics, and few evaluate more than the most basic geographic risk factors. Despite the consistent finding in geographic research that smaller aggregations improve the ability to detect associations, few publications in the past decade looked at aggregations below the state level. No identified article articulated a broad theory or conceptual framework for why occupational injury/illness risk would vary geographically. Finally, organized efforts to promote occupational safety and health surveillance have not viewed geography as a priority (Council of State and Territorial Epidemiologists (CSTE) in collaboration with NIOSH, 2004; CSTE, 1999; NIOSH, 2001).

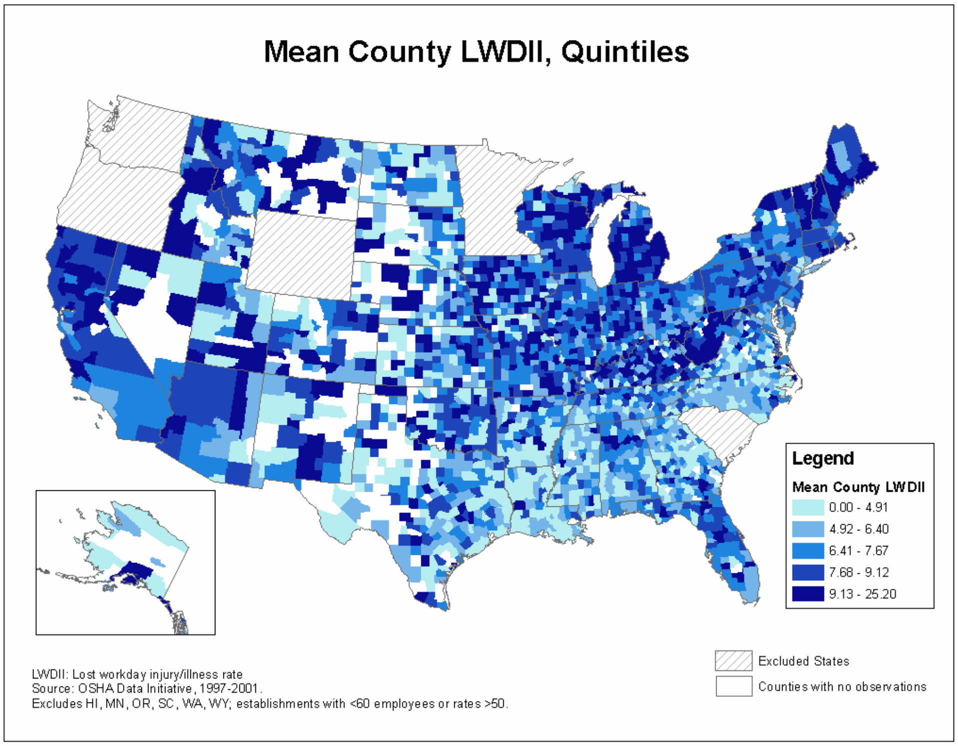
Past studies of geographic variation in occupational injury and illness have had real benefits. These include: critical insights in cancer and respiratory disease causation; effective programmatic intervention including improved targeting, clarifying and visually dramatizing information for policymakers and the public; supporting theory development; and contributing to understanding of social and other contributions to occupational injury and illness (Devesa, et al., 1999; Neff, 2006; J. Sestito, Personal communication, 2006; Smith, 2001).

The field is open for additional research and surveillance, particularly: developing indicators; using spatial statistics; theorizing; examining data at aggregations below the state level; and addressing questions specific to industries and occupations.

III. RESEARCH GOALS

This dissertation aims to stimulate additional surveillance and research on geography and occupational injury/illness and to present models for this work based on one database, the OSHA Data Initiative (ODI, 1997-2001, n=216,846) (Map 1). In the ODI, OSHA collects lost workday injury/illness rates (LWDII) and other data from establishments in high injury/illness industries, for use in targeting enforcement. The ODI is a rare national source of establishment-level injury/illness rates and no descriptions have been identified in the peer reviewed literature, so this investigation takes advantage of an important opportunity. The enforcement usage and changes in sampling strategy by year are limitations, and the thesis aims to evaluate their impact on the findings.

Map 1: Mean county Lost Workday Injury/Illness Rate (LWDII), ODI 1997-2001, quintiles



Following are the research questions guiding the four papers.

1) All over the map: A review of literature on geography and occupational injury and illness. This paper categorizes and critiques the small literature on geography and occupational injury/illness. Four categories are identified:

- Surveillance
- Hypothesis-generating research
- Research on either geography or occupational injury/illness, controlling for the other

- Hypothesis-testing research

2) "In the wrong place…?": Spatial tools for occupational injury and illness surveillance. This paper demonstrates methods for examining geographic data on occupational injury/illness through an analysis of county-level rates in the ODI. Methods including mapping, spatial statistics, and ranking are presented to address the following four questions:

- Does occupational injury/illness vary spatially at the county level?
- Does variation remain after accounting for industry hazard?
- Where are rates higher or lower than expected?
- What social risk factors seem to covary with occupational injury/illness?

Finally, this paper questions whether it is appropriate to use the ODI for surveillance.

3) Social predictors of county occupational injury and illness rates. This paper presents a multilevel regression analysis examining effects of county- and state-level risk factors on mean workplace occupational injury/illness rates by U.S. county. Outcome data come from the ODI and potential predictor variables from the U.S. Census 2000 and other sources. Spatial statistics and other analyses are used to evaluate the results.

4) Occupational injury and illness in meat processing: A geographic cut. The final paper presents a case study of one industry, because it is expected that geographic predictors will vary by industry. Focusing on a smaller sample also enabled matching database entries across years to help compensate for the ODI's lack of unique establishment identifiers. The case study examines the meat and poultry processing industries (n=1553 establishments), which are among the most hazardous in the United States. Research questions were:

- Where are processing plants located, and to what extent do these locations correspond to social risk factors?
- Is there a difference between the distribution of *high-injury/illness rate* plants and others? How do high-rate plant locations correspond to social risk factors?
- What are the associations between social risk factors and a) plant location; b) high rate plant location, as examined through linear and multilevel regressions?
- What evidence can be found regarding the possibility that underreporting plays a significant role in the ODI?

These investigations will provide new insights about geographic variation in work-related injury/illness and suggest a set of methods that could be developed in other databases or pursued in future years of the ODI. They will also contribute to the literature on underreporting of occupational injuries and illnesses. Based on this thesis, policy recommendations will be made in the following five areas: underreporting, surveillance, research, targeting, and evaluating the conceptual framework.

Each year, over 55,000 people die from work-related exposures in the U.S.; taken together, these events make occupation the equivalent of the eighth leading cause of death (Steenland et al., 2003). Further, 4.26 million U.S. workers reported nonfatal work-related injuries and illnesses in 2004, and this figure may exclude 2/3 of such events (Bureau of Labor Statistics, 2005; Leigh et al., 2004; Rosenman et al., 2006). Repercussions for workers' – and their families' - lives and livelihoods can be major (Boden et al., 2001; Dembe, 2001; Dorman, 2000; Leigh & Robbins, 2004). A public health approach to prevention involves not only searching for proximal causes but also stepping back to examine the context.

2

All Over the Map:

A Review of Literature on Geographic Variation in Occupational Injury & Illness

Neff, RA

ABSTRACT

Background: Assessment is the first core function of public health, and an understanding of geographic variation in health status and health risk factors is fundamental to prevention. In the occupational injury and illness field, there is great unrealized potential for using geographic tools.

Methods: A literature search was performed to identify surveillance reports and studies on occupational injury/illness and geography. Articles were categorized and reviewed.

Findings: Four literature categories were identified: surveillance; hypothesis generating research; research controlling for either factor; and hypothesis-testing research. Only seven of the identified papers were located through initial Medline searching. Few recent publications looked at aggregations below the state level, and virtually none used spatial statistics. No identified article articulated a broad theory about why occupational injury/illness risk would vary geographically. The literature generally did not build or gain cohesion across time. Yet, there is a proven track record of benefit from geographic examination.

Conclusions: To stimulate further research, I suggest research needs and databases that could be explored. Today, the literature on geographic variation in occupational injury/illness is "all over the map." A better understanding of the geographic spread and determinants of these events can help provide "driving directions" to streamline the path of prevention.

* * * * *

I. INTRODUCTION

In thirty-one years since the first published studies used geographic analysis to examine occupational injury/illness, geography has moved into the mainstream of public health and environmental tools. Geographic methods provide invaluable insights into disease mechanisms, drivers of disparities, and strategies for improving prevention, enforcement, and services (Cromley & McLafferty, 2002; Hillemeier et al., 2003; Krieger et al., 2002; Krieger, et al., 2003; Waller & Gotway, 2004). The U.S. government has set the objective of "Increas[ing] the proportion of all major national, state and local health data systems that use geocoding to promote nationwide use of geographic information systems (GIS) at all levels." The target is to have 90 percent of data systems using geocoding [geographic identifiers that enable mapping] by 2010 (Objective 23-3) (U.S. Department of Health and Human Services, 2000).

But while other sectors of public health have advanced their use of geography over time, the occupational safety and health field has been far less active. To help stimulate exploration of geography and occupational injury/illness, this paper illustrates the literature gap, categorizing and reviewing the history of publication on this topic. Further papers will present geographic tools for describing a national county-level occupational injury/illness database and geographic risk factors, evaluate quantitatively a set of area-level risk factors suggested by this paper, and present a geographic case study of the meat processing industry.

II. BACKGROUND

In this paper, the terms, "geographic" and "spatial" are used interchangeably to refer to the variation across space in risk factors or health outcomes. "Area-level" variation refers to variations between discrete geographic areas such as counties. Area-level characteristics can reflect the aggregated effect of individual characteristics (such as the summed experience of workers in a high hazard industry that dominates an area) or the holistic effect of being in an area with particular conditions (such as a cultural view that occupational injury is normal). Some characteristics such as inequality and segregation only make sense when examined at the area level.

The geographic tools used in papers reviewed here could be considered to occupy a hierarchy of analytic detail. 1) Stratifying by geographic units; 2) Mapping data points or area distributions, either alone or in conjunction with pertinent risk factors; 3) Using spatial statistics to a) test whether maps show an overall pattern of geographic variation in the outcome (spatial dependence), b) identify and examine clusters of high or low rates, and/or c) conduct regression.

Most studies focus at the bottom of this hierarchy. Research is improved when it takes better account of the spatial nature of data. Specifically, due to the fact that risk factors often do not respect geographic boundaries, nearby areas are often more similar to each other than to those further away, a phenomenon known as positive spatial autocorrelation. For regressions, this autocorrelation violates the assumption of independence. If areas are not independent but are analyzed as if they are, the uncertainty of regression effects could be underestimated, leading to findings that overstate statistical significance.

III. PRIOR RESEARCH

There is a small and challenging-to-access literature examining occupational injury and illness in geographic context. Table 1 lists the numbers of responses appearing on Pubmed searches using simple search strategies. While the most productive search term, "occupational injury AND geographic" yielded 28 studies, abstract review indicated that only seven of these actually addressed the topic, in English.[1] Further searching, particularly using NIOSH's NIOSHTIC database,(NIOSH, 2006b) the OSHROM Medline subset database (available by license), and "snowballing" from cited

[1] Studies were excluded because: geography was not studied but was mentioned in the abstract as part of sample selection (4) or discussion (2); geographic and work causation of a condition were studied in separate analyses (4); "occupational" referred to "occupational therapy" (1) ; an exposure or disease was associated with the term, "occupational" in a MESH heading, even if it was not studied in this context (5); papers were not in English (5).

references and the Science Citation Index, has led to compilation of a larger bibliography.

Table 1: PubMed searches for articles about geography and occupational injury/illness. September 28, 2005

Search strategy	# returned	# *actually addressing* occupational injury/illness and geography
occupational injury AND geography	9	0
occupational illness AND geography	2	0
occupational illness AND spatial	9	1 (Lange et al., 2002)
occupational injury AND spatial	19	1 (Lange et al., 2002)
occupational illness AND geographic	9	1 (Lange et al., 2002)
occupational injury AND geographic	28	7 (Cattledge et al., 1996; Lange et al., 2002; D. Loomis et al., 2003; McNabb et al., 1994; Miller & Levy, 1999; D. Richardson et al., 2004; Watson & White, 1984)

Reviewed publications are divided into 4 groups:

1. Surveillance reports describing variation in risk by area

2. Hypothesis-generating or descriptive research

3. Research focused on *either* geography *or* occupational injury/illness, controlling for the other

4. Hypothesis-testing research.

A. SURVEILLANCE REPORTS

The Centers for Disease Control and Prevention (CDC) defines public health surveillance as "the ongoing systematic collection, analysis and interpretation of outcome-specific data for use in the planning, implementation and evaluation of public health practice."(Thacker, 2000).

While formal surveillance of occupational injury and illness existed in the U.S.

since the early part of the 20th century,(Sundin & Frazier, 1989) the first identified data sources enabling geographic analysis became available in the mid 1970's. The National Occupational Hazard Survey (NOHS, 1974) yielded a list of almost 9,000 exposure agents by industry and occupation, enabling researchers to generate estimates of potentially exposed workers by hazard and industry (Sundin & Frazier, 1989). The data were placed in geographic context as "hazard" and "exposure" by linking the file with Dun and Bradstreet's commercial file listing establishments, counties, industry, and workforce size,(Frazier & Sundin, 1986) or the 1963 Census of Manufacturers (Stone et al., 1978). Another newly available database listed toxic effects of a subset of these industrial chemicals, enabling researchers to map health outcomes of concern based on industrial locations (Sundin & Frazier, 1989).

In 1975 the National Cancer Institute (NCI) published the first national county-level Atlas of Cancer Mortality. This resource and the follow-up version for non-white populations provided county-level data on numerous cancer outcomes, often divided by demographics. It also spurred efforts to control cancer in many states (Mason et al., 1975; Mason et al., 1976). Updated Cancer Atlases were published in 1987, 1990 and 1999, (Devesa, et al., 1999; Pickle et al., 1987; Pickle et al., 1990) and the cancer mapping project has been replicated in over 30 countries (Fraumeni, J, Personal communication, April 10, 2006). Today, the U.S. Cancer Atlas has an on-line interface allowing users to construct their own maps based on area, age, gender, race, and time period.

An additional disease-specific data source that has been used geographically is NIOSH's National Occupational Respiratory Mortality System (NORMS), a component of the National Occupational Mortality Surveillance (NOMS) system. NIOSH publishes state-level maps of several respiratory mortality outcomes and provides an online interface allowing users to generate national county-level maps by gender, race/ethnicity, age, and year. A 1998 atlas mapped diseases at the health services area level (Kim, 1998; NIOSH, 2006a).

The Injury Fact Book, first published in 1984, includes a chapter on occupational injury, with state maps of fatalities in general and by cause. The information is enriched by the ability to compare it with data on non-occupational injury (Baker et al., 1992).

Probably the most comprehensive compilation of mapped occupational injury/ illness data is the NIOSH *Worker Health Chartbook*, which presents state-level distributions of occupational injury and illness outcomes, including by industry, drawn from a variety of surveillance data sources (NIOSH, 2004). The chartbook was started in 2000 on paper; the 2004 edition uses electronic technology to make graphics and underlying data available to users in a variety of formats. The Bureau of Labor Statistics also published a useful chartbook on fatal occupational injuries and illnesses in 2003, including separate sets of charts by state describing fatality demographics, occupations, industries, and event causes (US Department of Labor, Bureau of Labor Statistics, 2003).

NIOSH's 1993 chartbook on National Traumatic Occupational Fatality data was an early presentation of fatality data (NIOSH, 1993). Its maps showing Alaska's fatality rate to be almost five times the U.S. rate led to a multiagency collaborative to address occupational injuries/illnesses there; today the collaborative is considered a major success story for the field (Smith, 2001). Another type of state-level surveillance report is the AFL-CIO's annual "Death on the Job" report, describing occupational safety and health issues with a focus on enforcement gaps (for example, (AFL-CIO, 2006)).

These surveillance reports are useful to policymakers and program planners, and can support hypothesis-generation for research. Unfortunately, due to data limitations, very few maps showing *county-level* distributions are available. The online NORMS interface and the Cancer Atlas are the only national county-level maps identified since the 1980's. At least two states, however, have presented county-level maps (New Jersey Department of Health and Senior Services, 2005; Stanbury et al., 2004). Michigan's report is an especially useful example of how geographic tools can show county variations that suggest needs for followup, and can contribute to hypothesis-generation.

Other surveillance data sources (described in Table 2) include the International Labour Organization's national-level surveillance materials using geographic comparison, and Hazardous Substances Emergency Events data by state (Horton et al., 2004; International Labour Organization, 2004; Takala & International Labour Organization, 2005).

Overall, while these examples of geographic tools in occupational injury/illness surveillance reports can be found, it is still not the mainstream to analyze data geographically, and as described below, probably will not become so in the near future. NIOSH underwent a strategic planning process on surveillance starting in 1997. The resultant report emphasizes some priorities relevant to geographic surveillance, such as creating data useful for localities, but does not recommend mapping or gathering georeferenced data (NIOSH, 2001). The Council on State and Territorial Epidemiologists (CSTE) designed and piloted a set of occupational injury/illness indicators for state surveillance. The group did not recommend geographic data display; indeed, the issue was not discussed during planning (CSTE in collaboration with NIOSH, 2004; CSTE, 1999; Sestito, J., Personal communication, 2006). John P. Sestito, J.D., M.S., NIOSH Surveillance Program Coordinator, participated in both processes and suggests that this is because data at relevant aggregations is not available in many states, and because in general geographic display and analysis has been perceived as a second-level issue that might be addressed after the basic priorities are covered (Sestito, J., Personal communication, 2006).

Table 2: Surveillance reports

Citation	Geography, aggregation	Methods
Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF, Jr. 1975. **Atlas of cancer mortality for U.S. counties: 1950-1969.** DHEW Publ. No. (NIH) 75-780. *(See also updates 1976, 1987, 1990, 1999*)	U.S., by county	Maps of all cancers and many specific cancers by year, race, gender. 1950-1969 initially; later versions up to 1994.
Stone BJ, Blot WJ & Fraumeni JF. **Geographic patterns of industry in the United States: An aid to the study of occupational disease**. *J. Occup Med* **20**, 472-7 (1978).	U.S. counties, 1963 Census of Manufacturers, 1960 Census	Calculated county percent employment by 18 major manufacturing industries, mapped distribution. Described average county % urban, median schooling, and median income for each industry.
Centers for Disease Control. **Annual Summary 1984: Reported morbidity and mortality in the United States. "Occupational Hazards".** *MMWR* **33**, 97-104 (1986).	U.S., by county	Mapped facilities that might use inorganic lead. 1972-1974
NIOSH. 1991. **Work-related lung disease surveillance report.** DHHS (NIOSH) Pub. Number 91-113. *(See also updates 1992, 1994, 1996, 1999, 2002)*	U.S., by state	Presents state (and sometimes county) data on multiple work-related lung diseases in table form. Later update reports include maps. For example, the 2002 version maps the following respiratory outcomes and related exposures: asbestosis, pneumoconiosis, silicosis, byssinosis, malignant mesothelioma, hypersensitivity pneumonitis.
Baker SP, O'Neill B, Ginsburg MJ, Li G. ***The Injury Fact Book, Second Ed.*** Oxford University Press: New York. (1992) *(See also, first edition.)*	U.S., by state (and one analysis by county)	Maps state occupational fatality rates, both in general and those due to machinery, falling objects, electric current, and explosions. Compares death rates from these causes in low income versus high income counties.
NIOSH. 1993**. Fatal injuries to workers in the United States, 1980-1989: A decade of surveillance.** Report 93-108.	U.S., by state	Maps average annual state traumatic occupational fatalities. Also maps fatalities for six causes of death and four industries separately. Map categorization was based on standard

Citation	Geography, aggregation	Methods
		deviation.
Multiple authors. 1998. **Adult blood lead epidemiology and surveillance [ABLES]—United States, third quarter, 1997**. MMWR 47(4):77-80. *(See also subsequent issues of MMWR including 3/19/99, 12/13/02, 7/9/04)*	U.S., by state	Analyses of ABLES adult blood lead data are reported regularly in MMWR. For example, the February 6, 1998 issue mapped % change in reports of blood lead levels >=25 ucg/dL, by state." Years vary.
Kim JH, 1998. **Atlas of respiratory disease mortality, United States: 1982-1993.** DHHS (NIOSH) Number 98-157.	U.S., by Health Service Area	Maps age-adjusted death rate and area death rate compared to U.S. death rate for numerous respiratory conditions. 1982-1993.
Windau JA, J. T., & Toscano GA, . **State and Industry Fatal Occupational Injuries, 1992-96.** *Compensation and Working Conditions Online* (Summer 1998).	U.S., by state	Describes occupational fatalities by state, industry. 1992-1996.
Adekoya N, Pratt SG. 2001. **Fatal unintentional farm injuries among persons less than 20 years of age in the United States: Geographic profiles.** DHHS (NIOSH) Publication No. 2001-131.	U.S., by census region, census division, state	Presents numerous tables and charts comparing unintentional farm fatalities among youths by geographic area. Leading causes are presented for each state separately. Maps are presented at the state level. 1982-1996.
NIOSH. **Worker Health Chartbook 2004.** 2004-146 (2004). *(the earlier version, 2000, had far fewer geographic analyses*(NIOSH, 2000))	U.S., by state	Chartbook presents state-level maps, tables covering: fatal occupational injuries, nonfatal injury/illness, blood lead, musculoskeletal disorders, disorders due to physical agents, poisonings, hypersensitivity pneumonitis, malignant mesothelioma, asbestosis, byssinosis, pneumoconiosis, silicosis, occupational dust diseases, respiratory conditions, skin diseases, agriculture/forestry/fishing industry events. Multiple years.
U.S. Department of Labor, Bureau of Labor Statistics. **Fatal Occupational Injuries in the United States, 1995-1999: A Chartbook.** BLS Report 965 (2003).	U.S., each state separately	Chartbook describes occupational fatalities: events leading to fatality, occupations/industries with the most fatalities, state breakdowns by wage/salary vs. self-employed, gender, age, race of fatal/nonfatal occupational injury/illness. 1995-1999.
New Jersey Department of Health and Senior Services. 2005. **Fatal occupational injuries in New Jersey: Ten year report 1993-2002.** 2006(April 17)	New Jersey, including by county	State surveillance of occupational injuries/illnesses includes maps of counts of fatal occupational injuries, fatal occupational roadway injuries, workplace homicides. 1993-2002.
Stanbury M, Largo TW, Granger J, Cameron L, Rosenman K. 2004. **Profiles of occupational injuries and diseases in Michigan.** 2005(September 28)	Michigan, including by county.	State surveillance of occupational injuries/illnesses includes maps showing: occupational fatality (2001), hospitalization for work-related injury (1999-2001), work-related asthma (1988-2001), work related elevated

Citation	Geography, aggregation	Methods
		blood lead (1998-2001), work-related silicosis (1985-2001), malignant mesothelioma (1985-2000).
Horton DK, B. Z., & Kaye WE, . **Surveillance of Hazardous Materials Events in 17 States, 1993–2001: A Report From the Hazardous Substances Emergency Events Surveillance (HSEES) System.** *American Journal of Industrial Medicine* **45**, 539-48 (2004).	17 U.S. states, by state	Review of Hazardous Substances Emergency Events data showing industry, industry sub-category, and substances released. 1993-2001.
Council of State and Territorial Epidemiologists. **Putting data to work: Occupational Health Indicators from Thirteen Pilot States for 2000.** (2005).	13 U.S. states, each state separately	Presents indicator data covering the following: employment in industries/occupations with high risk for occupational morbidity/mortality; occupational health and safety professionals; OSHA enforcement activities; non-fatal injuries/illnesses; hospitalizations; fatal injuries; amputations; burns; musculoskeletal disorders; pneumoconiosis; pesticide poisonings; mesothelioma; elevated blood lead. 2000.
Takala J, International Labour Organization. 2005. **Introductory report: Decent work -- Safe work.** 2005(September 23)	World, by I.L.O. region and country	Report includes data by region on fatal/nonfatal accidents, multiple types of illness, occupational death as % of total death and of disability adjusted life years. Also comparative analysis of country competitiveness vs. occ, fatalities. 2002.

B. HYPOTHESIS-GENERATING DESCRIPTIVE RESEARCH

This category is distinguished from “surveillance” in that studies are aimed at questions of causation, whereas surveillance is practice-oriented and aimed at improving interventions. Clearly, there is much overlap between the two. The studies in this category, described in Table 3, commonly focus on particular outcomes or particular industries/occupations.

The late 1970’s and early 1980’s saw a flowering of county-level descriptive research using the above-described new surveillance data sources (Blot et al., 1977; Brinton et al., 1976; Hoover & Fraumeni, 1975; Hoover et al., 1975; Stone et al., 1978).

Researchers generally compared sets of maps or compared standardized mortality ratios between counties with high prevalence of particular diseases or industries and those without. They also developed creative methods. For example, Hoover et al (1975) used high male versus female cancer rates to suggest occupationally-related cancers; and excluded counties with high lung cancer as a way to control for the effect of smoking prevalence. Researchers also brought together and compared geographic patterns of hazard, exposure, and health outcome – the full circle of public health surveillance (Froines et al., 1986; Thacker et al., 1996). An example of this data triangulation is Frazier et al's (1983) paper presenting maps of county-level number of worksites using formaldehyde, number of exposed workers, and nasal cancers.

In its 1999 Cancer Atlas, NCI summarized key scientific findings that came from earlier atlases. Occupational findings include associations between lung cancer and both smelter workers with arsenic exposure, and shipyard workers with asbestos exposure; nasal cavity cancers and furniture workers; and bladder cancer and truck drivers. They also cited evidence that bladder cancer seems to be associated with chemical industry counties, and that agriculture may contribute to prostate cancer, non-Hodgkins lymphoma, and leukemia (Devesa, et al., 1999).

Interestingly, most of this body of hypothesis-generating geographic literature was created at the National Cancer Institute (NCI), with a few colleagues doing related work at NIOSH and initially funded by NCI (Frazier et al., 1983). Perhaps extramural researchers lacked the data access, computing and software to perform similar analyses.

During this period, public health as a whole was taking little advantage of geographic information systems (GIS) (Cromley & McLafferty, 2002). By contrast, since that time GIS research has flourished throughout public health, but since the late 1970's only a few hypothesis-generating geographic studies have been published on occupational injury/illness. These are generally disconnected from one another, examining a variety of industries, occupations, and outcomes, at varying aggregations (Cattledge et al., 1996; Frazier & Sundin, 1986; Froines et al., 1986; Waehrer et al., 2004; Watson & White, 1984). The above-referenced studies examined exposures to silica, formaldehyde, being a female coal miner, construction falls, and industry hazard.

Of particular general interest, Waehrer et al (2004) estimated state costs per worker for nonfatal and fatal occupational injuries/illnesses, finding almost a 3-fold range. They found that 73 percent of the variation in combined state costs could be explained by state industrial mix alone (Waehrer et al., 2004).

When asked why NCI researchers had cut back this line of research into geography and occupational cancers, Dr. Joseph Fraumeni, Jr., currently Director of NCI's Division of Cancer Epidemiology and Genetics, stated that "in the 1970's, cancer was more of a black box," so researchers would take advantage of every opportunity to glean information that would provide clues to etiology. Fraumeni said that while there is still interest in cancer mapping and GIS approaches at NCI, many more research tools are now available with more precision. The finding that cancer rates vary geographically was initially explored primarily based on environmental and occupational hazards, whereas today lifestyle and genetic determinants are being increasingly emphasized in epidemiologic research. (Fraumeni, J. Personal communication, April 10, 2006)

I argue that there remains a place for geographic hypothesis-generating research. Given that Census and other area data for comparison are high quality and easily accessible, where geo-referenced occupational injury/illness data already exist, these studies have the benefit of being relatively easy/inexpensive ways to look quickly at a variety of risk factors to get ideas. Further, these are often the best methods for generating hypotheses about social risk factors, both because the risk factors often have most meaning at the area level and because individual-level data on them is rare. Geographic tools can also be used to consider the *interactions* of genetic and lifestyle determinants with environmental and social ones.

Today there are many more geographically referenced databases available than ever before. Some of the data sources used in earlier studies have been updated, making this hypothesis-generating research more feasible than ever.

Table 3: Hypothesis-generating/descriptive research

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
Hoover R, Mason TJ, McKay FW & Fraumeni JF. **Cancer by county: New resource for etiologic clues.** *Science* **189**, 1005-1007 (1975).	U.S., by county	Bladder & stomach cancers were mapped and risk factors examined as case studies of use of geographic methods for cancer. In bladder cancer study, a set of counties with increased bladder cancer, high m:f ratio, and not-high lung cancer were compared to others, in terms of industrial distribution and other risk factors. 1950-1969.	Non-electrical and electrical manufacturing, motor vehicle manufacturing, and urbanicity were risk factors for bladder cancer. Occupational risk factors were not suggested for stomach cancer.	N	N
Hoover R & Fraumeni JF. **Cancer mortality in U.S. counties with chemical industries**. *Environmental Research* **9**, 196-207 (1975).	Chemical industry counties; U.S. cancer mortality	"Chemical industry counties" were identified. White age-adjusted mortality rates for 35 cancers were compared for chemical industry counties vs. others. As control, counties were also compared based on urbanization, class. In chemical counties with high lung, liver and bladder cancer, looked at specific industry, time trends of cancer and other risk factors. 1950-1969.	Excess bladder, lung, liver, other cancers in these 139 counties. Could not be explained by urbanization, socioeconomic factors. Particular sub-industries of concern identified.	N	Y – cancer mortality in these counties is associated with industrial chemical use
Brinton LA, Stone BJ, Blot WJ & Fraumeni JF. **Nasal cancer in U.S. furniture industry counties.** *Science* (1976).	U.S. furniture industry counties vs. national cancer rates	Presented ratios of age-adjusted cancer mortality for white males between "furniture industry" counties and controls. 30 cancers. 1950-1969.	Furniture industry counties had low rates of nearly all cancers, but statistically significant excesses for nasal cancer, melanoma, multiple myeloma	N	N
Blot WJ, Brinton LA, Fraumeni JF & Stone BJ. **Cancer mortality in U.S. counties with petroleum**	U.S. petroleum industry counties vs. control counties	Computed ratios of age-adjusted mortality from 23 cancers between white male residents of petroleum industry counties vs. those in control	Elevations for all but 4 cancers; largest elevations: nasal cavity/ sinuses, lung cancer. Also significant:	N	Y – petroleum industry may be a risk factor for cancer

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
industries. *Science* **198**, 51-53 (1977).		counties. 1950-1969.	skin, testis, stomach, rectum. Low for brain cancer)		
Frazier TM, Lalich NR, Pedersen DH. 1983. **Uses of computer-generated maps in occupational hazard and mortality surveillance.** Scand J Work Environ Health 9(2 Spec No):148-54.	U.S., by county	As a way to demonstrate different uses of maps in occupational injury/illness, mapped distributions of: formaldehyde-using facilities; workers potentially exposed to formaldehyde using different selection criteria; distribution of Indiana workers potentially exposed; and nasal cancer deaths.	Maps and methods were discussed but no conclusions drawn.	N	N
Watson AP & White CL. **Workplace injury experience of female coal miners in the United States.** *Archives of Environmental Health* **39**, 284-293 (1984).	U.S., by region	Compared workplace injury in female coal miners by U.S. region and other characteristics using percentages and significance tests. 1978-1980.	More than 70% of "accidents" occurred in Appalachia	N	Y – Geography is partly a surrogate for type of mine and entry-level position
Frazier TM, Sundin DS. 1986. **Industrial demographics and population at risk for silica exposures.** In: Goldsmith DF, Winn DM, Shy CM, editors. *Silica, silicosis and cancer.* Philadelphia: Praeger Scientific. p 3-9.	U.S., by county	Demonstrated multiple uses of National Occupational Hazard Survey data to explore silica exposures. Mapped worksites by county with potential free silica exposures; county dist of # workers with exposure; county dist of % of workers exposed; county cause specific mortality from silicosis and lung cancer. 1972-1974.	Maps show distributions	N	N
Froines JR, Dellenbaugh CA, Wegman DH. **Occupational health surveillance: A means to identify work-related risks.**	Los Angeles County	Develops a method for identifying work-related risks by ranking industry hazards according to several systems, then comparing county employment patterns with industry rankings.	Discusses strengths and limitations of different ranking systems, and how the method could be applied.	N	N

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
Am J Pub Health 76(9) 1089-1096 (1986).					
Cattledge GH, Schneiderman A, Stanevich R,Hendricks S, & Greenwood J. **Nonfatal occupational fall injuries in the West Virginia construction industry.** *Accident Analysis and Prevention* **28** (1996).	U.S., by state	Reported fatality rates for falls in construction by state, region, sub-region. 1980-1989.	Falls most commonly from TX, CA, FL, IL, PA, NY. 8.3% died in state other than residence. By region, south (esp west south central) had highest rate; Northcentral had lowest.	N	N
Waehrer G, Leigh JP, Cassady D & Miller TR. **Costs of occupational injury and illness across states.** *J Occup Environ Med.* **46**, 1084-1095 (2004).	50 states	Estimated costs per worker from fatal and nonfatal occupational injury/illness (including pain and suffering) by state. Performed regression to identify predictors of high/low state costs based on industry alone. 1993.	Industry alone accounted for 73% of variation in state costs.	N	Y – industry is an important predictor of costs.

C. RESEARCH CONTROLLING FOR GEOGRAPHY OR OCCUPATIONAL INJURY/ILLNESS

This category includes both area-level studies of general injuries and illnesses that look at work-related factors; and studies of occupational injury/illness that include geographic units as control variables or otherwise not the main research question. This body of research is somewhat ancillary, and no attempt is made to review it comprehensively. It is presented here and in Table 4 because insights about geographic risk factors can be derived from review of results sections of these studies.

One *occupational injury/illness study* that looks at geographic units but does not focus on them is Dembe et al.'s (2004) analysis of National Longitudinal Survey of Youth data. In seeking to identify predictors of occupational injuries and illnesses among respondents, the study controlled for U.S. region. They found that residents of the North Central and Western regions comprised higher portions of those with work-related injuries and illnesses than of those without. By contrast, living in the Northeast or South was protective. The finding was not addressed in the introduction or discussion. Other studies in this category include: (Miller & Levy, 1997; Miller & Levy, 1999; Ohsfeldt & Morrisey, 1997; Smitha et al., 2001).

Studies like these address an important confounder and thus significantly improve upon studies assuming geographic homogeneity. However, researchers with access to such geographically coded data should consider examining it more directly. Ideally, they should also incorporate appropriate error terms in statistical models to account for geographic autocorrelation. Further, their research would be enhanced by discussion of reasons for observed geographic variation (or lack thereof).

The second subcategory is articles focused on *geographic variation* that include but do not emphasize occupational injury/illness. Wigglesworth (2005) examined the consistency with which states having the highest and lowest motor vehicle traffic mortality also had the highest and lowest death rates from other types of injury. Of all types examined, occupational injuries showed the strongest relationship, demonstrating exact concurrence with the hypotheses. Wigglesworth suggested a number of risk factors that could account for the relationship between occupational and motor vehicle fatalities, including policy, socioeconomic status, engineering or educational interventions, and the concept of area safety culture, but the study did not enable examination of those factors. Muntaner, et al. (2002) find strong international associations between unintentional injury mortality and working class power, and suggest that this might reflect the impact working class power has on occupational injury. Joines, et al. (2003) conducted the only identified study in this category to make use of spatial lag regression. They analyzed county-level predictors of low back pain hospitalization, finding among other things that surgery was related to employment in the heavy lifting/transportation industries.

This brief review demonstrates that articles using geography or occupational injury/illness as control variables can be used to glean insights about these phenomena. However, identifying and accessing these articles depends on keyword coding and is even more challenging than identifying articles directly focusing on this theme. The fact that researchers frequently feel the need to control for geography supports this review’s contention that geography represents an important determinant of outcomes. However, in these papers the reader is generally left to speculate as to the hypothesis driving this inclusion. Further comment about hypotheses would importantly enrich the articles.

Table 4: Geography of Occupational Injury/Illness as Control Variable

Citation	Geography, aggregation	Main Focus	Methods	Relevant findings	Spatial or multilevel research methods?	Theory-based?
Ohsfeldt RL & Morrisey MA. **Beer taxes, workers' compensation, and industrial injury**. *The Review of Economics and Statistics* **79**, 155-160 (1997).	U.S., by state and 2-digit SIC code	State policy effect on occupational injury/illness	Regression analysis examined the state beer tax as a predictor of lost workdays due to nonfatal injury. Control variables included WC factors, unemployment, education, firm size, age, gender, dry population, religion, year, 2-digit SIC. 1975-1985.	State beer taxes significantly associated with injury (a 25-cent increase in 1992 beer tax would save about 4.6 million lost workdays)	Y - adjusted state beer tax for neighboring state tax	Y – Alcohol is a risk factor for work-related injury.
Miller TR & Levy DT. **Geographic variation in expenditures for workers' compensation in physician claims.** *American Journal of Industrial Medicine* **32**, 27-34 (1997).	17 states, by state	Predictors of state occ. injury/illness *treatment costs*	Regression of state risk factors on workers' compensation payment per physician claim for 3 injury groups. Predictors include personal, injury, state characteristics, other. 1979-1988.	Much variation in per-episode costs across states remains after controlling for urbanicity and health services risk factors	N	Y– Health services risk factors predict treatment costs.
Miller TR & Levy DT. **Geographic variation in expenditures for workers' compensation hospitalized claims.** *American Journal of Industrial Medicine*	17 states, by state	Predictors of state occ. injury/illness *treatment costs*	Regression of state risk factors on workers' compensation payment per hospitalization for 3 injury groups. Predictors include: case mix, severity, personal characteristics, urban, health services, other.	Much variation in costs across states beyond case mix and state risk factors. State rate regulations have an important impact on lowering costs	N	Y– Health services risk factors predict treatment costs.

Citation	Geography, aggregation	Main Focus	Methods	Relevant findings	Spatial or multilevel research methods?	Theory-based?
35, 103-111 (1999).			1979-88.			
Smitha MW, Kirk KA, Oestenstad KR, Brown KC & Lee SD. **Effect of state workplace safety laws on occupational injury rates.** *Journal of Occupational and Environmental Medicine* **43** (2001).	42 states, by state	State policy effect on occupational injury/illness	Regression examining state-level risk factors for nonfatal occupational injury: Workers' Compensation requirements: safety committee, safety program, loss control, 'extra-hazardous employer' laws. Control vars: OSHA state plan, OSHA inspections, OSHA consultations, OSHA fines, employer size, unionization, unemployment, workers' comp maximum payment and waiting period, % high school graduation, age distribution. 1992-1997.	Injuries reduced with workers' compensation requirement of safety committee, safety program, and insurance carrier loss control regulations	N	Y – workplace injury is "a function of labor-force size, unemployment rates, age of the workforce, employment-type distribution, government involvement in the labor market," economic cycle position, and policy (OSHA and workers' compensation.)
Muntaner C, Lynch JW, Hillemeier M, Lee JH,David R, & Benach J, B. C.,. **Economic inequality, working-class power, social capital, and cause-specific mortality in**	16 "wealthy countries", by country	Effect of social factors on injury in general	Examined how social capital, economic inequality, and working class power are differentially correlated with health outcomes including unintentional injury. 1989-1992.	Injury strongly related to working class power factors.	N	Y – Working class power is a stronger predictor of health indicators than social capital

Citation	Geography, aggregation	Main Focus	Methods	Relevant findings	Spatial or multilevel research methods?	Theory-based?
wealthy countries. *International Journal of Health Services* **32**, 629-56 (2002).						
Joines JD, Hertz-Picciotto I, Carey TS, Gesler W & Suchindran C. **A spatial analysis of county-level variation in hospitalization rates for low back problems in North Carolina.** *Social Science and Medicine* **56**, 2541-2553 (2003).	North Carolina by county	Predictors of all low back pain hospitalizations, not just occupational	Used spatial lag regression to examine county-level predictors of low back pain hospitalization. Predictors included county percent employment by industry. 1990-1992.	Surgery was non-linearly related to county employment in heavy lifting/transportation industries, among other factors.	Y – spatial lag regression	Y – Socioeconomic and health resources factors play a role in hospitalization for low back pain.
Dembe AE, , Erickson JB, & Delbos R, . **Predictors of work-related injuries and illnesses: National survey findings.** *Journal of Occupational and Environmental Hygiene*, 542-550 (2004).	U.S., by region	General predictors of occupational injury/illness, not just geographic	Analyzed National Longitudinal Survey of Youth data for predictors of occupational injury/illness. Relevant analysis compared % reporting events by region and urban/rural status, weighted data. 1998.	Residents of the North Central and Western regions comprised higher portions of those with work-related injuries and illnesses than of those without, while the Northeast and South were protective.	N	N
Wigglesworth, E.,.	U.S. states	Identifying states	Evaluated number of	States that had high road	N	N

Citation	Geography, aggregation	Main Focus	Methods	Relevant findings	Spatial or multilevel research methods?	Theory-based?
Do some U.S. states have higher/lower injury mortality rates than others? *Journal of Trauma: Injury Infection & Critical Care* **58**, 1144-1149 (2005).		with consistent high and low injury rates across injury categories	states with highest and lowest road traffic fatalities that also had highest/lowest rates of other injury mortality types. Both all state occupational fatalities were examined and those from construction and retail trade. 1983-1995.	traffic injuries had high occupational fatalities across all occupations and in the selected construction and retail trade industries, and often high injury rates for other outcomes also; similarly for low injury states.		

D. HYPOTHESIS-TESTING ANALYTIC RESEARCH

This final category, shown in Table 5, includes research studies that have as their purpose to examine geographic risk factors for occupational injury and illness, and which are built on hypotheses about the expected findings rather than having a goal of description. Only a small number of studies fit into this category. Only one used spatial statistics and none used spatial or multilevel regression methods. Two sub-types are identified.

The first subtype looks at particular industries/occupations, risk factors, and health outcomes, often examining small spatial aggregations near to events to examine particular theories. For example, Grabowski et al. (2002) used GIS tools to identify geographic risk factors for pilot non-survival in the event of a plane crash, finding fatalities elevated in mountainous areas and those with poor weather. Jemal et al. (2000) followed an earlier NCI study on risk in shipbuilding counties, examining risk of lung cancer and mesothelioma. They found that rates were higher in shipbuilding counties, but interestingly, as female smoking has increased over time, rates for *women* in these counties were especially high.

The second type of research examines theories based on geographically-varying social factors. Loomis, Richardson, and colleagues at the University of North Carolina (UNC) have conducted much of the research in this area. Loomis et al. (2003) examined reasons for the U.S. decline in occupational fatality rates from 1980-1996. In terms of geography, they found that the South and West had the highest rates of fatal workplace injury in 1980 and the steepest declines from 1980-1996. Levels in the Northeast were lower and more stable. They suggested the possibility of change in industrial makeup by region as a reason. In a follow-up study, they found that adjusting for employment structure changed fatality trends by 10-15 percent, varying by region(Loomis et al., 2004). Richardson et al. (2004) looked at fatalities in the U.S. South and elsewhere, finding that African American men in Southern states had the highest rates, but that Southern Hispanic men were catching up. They attributed the findings both to the “racial and ethnic division of labor” and to the changing industrial structure of the South.

Other studies have examined different sets of social risk factors. Volinn et al. (1998) examined county risk factors for work related low-back pain disability. They found that county unemployment rate, percent receiving food stamps, and per capita income affected rates. They interpreted their findings to suggest that job insecurity stress affects outcomes. In another type of research question, McGlashan et al (2003) looked at cancers in African miners, finding evidence that common cancers were likely more caused by socioeconomic and other risk factors in miners’ areas of origin than by their occupation. Finally, Ussif (2004) looked internationally, finding support for the idea that economic expansion led to increased occupational injury rates, while safety measures led to reductions.

As can be seen from the atomized findings presented in this section and in Table 5, this is not a cohesive “literature,” but rather, a collection of studies around a similar theme. With the exception of the UNC studies, these articles generally do not cite or build upon one another, and there is no real temporal progression of knowledge or insight. Topics explored are also wide-ranging. But the glass is also half-full. These studies are diverse, interesting, and yield important insights. It is hoped that this review can contribute to increased cohesion of the literature in the future.

Table 5: Theory-Based, Analytic Studies

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
Blot WJ & Fraumeni JF. **Geographic patterns of lung cancer: Industrial correlations.** *American Journal of Epidemiology* **103**, 539-550 (1976).	U.S., by county (White residents); counties with at least 2,000 non-whites (non-whites). U.S. by region.	Regression examined associations between county age-adjusted lung cancer mortality rates by sex, race and possible predictors (Population density, % urban, % rural farm, % minority, % foreign parentage, median schooling, region, 18 manufacturing industries.) 1950-1969.	Risk factors included living/working in paper, chemical, petroleum, or transportation industry county; urban; others.	N	N
Blot WJ & Fraumeni JF. **Geographic patterns of oral cancer in the United States: Etiologic implications.** *J Chron Dis* **30**, 757 (1977)	U.S., by county (Whites by race and sex; nonwhites in counties with at least 1000 non-whites). U.S. geographic region	Regression examined associations between county age-adjusted oral cancer mortality rates, demography, 18 manufacturing industries, rural/urban, alcohol sales in 6 states, and region. 1950-1969.	Significant increase in white male oral cancer death rate in counties with high concentration of leather, paper, chemical manufacturing after controlling for other risk factors. White females: apparel, textile. Non-white: nonsigificant elevations.	N	N
Volinn E, Lai D, McKinney S & Loeser JD. **When back pain becomes disabling: a regional analysis**. *Pain* **33**, 33-39 (1998).	Washington State, by county	Regression examined predictors of county workers' compensation claim rates for back sprain. Socioeconomic risk factors: unemployment rate, % receiving food stamps, per capita income. Controls: labor force size, % in occupations with back sprain risk. 1979, 1983, 1985.	Socioeconomic risk factors predicted 1/3 of variation in county claims in 2 of 3 tested years; unemployment rate was a significant predictor in all 3 years	N	Y – "disability is a symptom of distress. Where there is a rise in job insecurity and an attendant rise in economic insecurity, there is a greater likelihood that back pain will become disabling."

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
Jemal A, Grauman D & Devesa S. **Recent geographic patterns of lung cancer and mesothelioma mortality rates in 49 shipyard counties in the United States, 1970-1994.** *AJIM* **37**, 512-521 (2000).	49 shipbuilding counties vs. coastal non-shipyard and all non-shipyard counties	Calculated age-adjusted rates (by race and gender) for lung cancer, mesothelioma between "shipbuilding counties" and coastal non-shipyard and all non-shipyard counties. Mapped national county variation in outcomes. Compared rates by U.S. region, rural/urban, year. 1970-1994.	Rates higher in shipyard counties, especially for women	N	Y– Asbestos exposure causes shipyard counties to have elevated lung cancer rates. High rates in women compared to earlier studies are probably due to asbestos combined with rising smoking rates.
Lange JL, Schwartz DA, Doebbeling BN, Heller JM & Thorne PS. **Exposures to the Kuwait oil fires and their association with asthma and bronchitis among Gulf War veterans.** *Environ Health Perspect* **110**, 1141-6 (2002).	First Persian Gulf War theater	Computed correlations between modeled and self-reported exposures to oil well fires during war. Conducted logistic regression on associations between exposure and asthma/bronchitis, control outcomes (depression, injury). 1990-1991 exposures.	Respiratory outcomes and control outcomes associated with self-reported exposure but not modeled exposure	N	Y– Recall bias may explain reports that Gulf War syndrome exists
Grabowski JG, , Curriero FC, , Baker SP, & Li G, . **Exploratory spatial analysis of pilot fatality rates in general aviation crashes using Geographic Information Systems.** *American Journal of*	United States, grid with intersections 50 miles apart	Used GIS and spatial statistics to examine predictors of pilot fatality vs. survival in general aviation crashes. Relevant risk factors were crash site elevation and bad weather. 1983-1998.	Mountainous areas and poor weather, among other factors, predicted fatality	Y –spatial statistics.	Y – Use of GIS tools will improve upon prior studies and may identify risk factors that do not conform to state boundaries

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
Epidemiology **155**, 398-405 (2002).					
McGlashan ND, , Harington JS, & Chelkowska E. **Changes in the geographical and temporal patterns of cancer incidence among Black gold miners working in South Africa, 1964-1996.** *British Journal of Cancer* **88**, 1361-1369 (2003).	Home locations of Black South African mining workers by province and country (South African provinces, Botswana, Swazi and Kangwane, Lesotho, Mozambique, Malawi and Northern Territories.	Descriptive statistics and crude/age-standardized incidence rates for numerous cancers experienced by miners, by miner home location. 1989-1996.	Cancer findings differed based on miner home areas.	N	Y – The two major miner cancers arise from socio-environmental causes related to home area rather than work.
D Loomis, JF Bena & AJ Bailer. **Diversity of trends in occupational injury mortality in the United States, 1980-96.** *Injury Prevention* **9**, 9-14 (2003).	U.S., by region	Among other analyses, Poisson regression estimated time trends in fatal occupational injury rate (unintentional, homicide), by U.S. region. 1980-1996.	South and West had highest rates in 1980s and fastest reductions in rates. Northeast had low 1980 rate and slower reductions.	N	Y – explained based on temporal change in distribution of dominant industry by region.
Ussif AA. **An international analysis of workplace injuries.** *Monthly Labor Review*, 41-51 (2004).	Canada, Finland, France, U.S., Sweden, by country	Examined time trends in occupational injuries by country. Conducted regression of effects of “injury-reducing variables” and economic expansion factors on injury by country. 1970-1999.	Countries had similarities in patterns. # workers, injury reducing factors, and procyclical factors all were significant	N	Y–Economic expansion leads to rises in occupational injury rates, while safety measures lead to reductions. The latter effect may be greater.

Citation	Geography, aggregation	Methods	Relevant findings	Accounted for auto-correlation?	Theory-based?
Richardson D, , Loomis D, , Bena J, & Bailer J, . **Fatal occupational injury rates in Southern and non-Southern states, by race and Hispanic ethnicity.** *American Journal of Public Health* **94**, 1756-1761 (2004).	U.S., by Southern vs. non-Southern states	Among other analyses, compared annual change in occupational fatality rates by region and race/ethnicity. 1990-1996.	Black men in Southern states had the highest fatality rate during the period, but Southern Hispanic men were taking over the lead.	N	Y – The “racial and ethnic division of labor” and changing industrial structure of the South drive disparities
Loomis D, Richardson DB, Bena JF, Bailer AJ. **Deindustrialization and the long term decline in fatal occupational injuries.** *Occup Environ Med* **61**, 616-621 (2004).	U.S., by region	Computed rates of fatal unintentional occupational injury based on death certificates and census data. Regional rates computed by all industry, and by “shrinking” and “growing” industry. 1980-1996.	There was geographic variation in fatality trends, and adjusting for employment structure changed these trends by 10-15%, varying by region.	N	Y – explained based on temporal change in distribution of dominant industry by region.
Antao, V.C. dos S., Petsonk EL, Sokolow LZ, Wolfe AL, Pinheiro GA, Hale JM, Attfield MD. **Rapidly progressive coal workers' pneumoconiosis in the United States: Geographic clustering and other factors.** *Occup Environ Med* **62,** 670-674 (2005).	U.S., by county	Looked at determinants of rapidly progressive coal workers’ pneumoconiosis – an indicator of inadequate prevention of high exposure to respirable dust. County proportions of rapid progression versus others were mapped.	There was a higher proportion of rapidly progressive cases in eastern Kentucky and western Virigina. Cases tended to be younger, to work in smaller mines, and to have worked longer at the face of the mine than those without rapid progression.	N	Y – geographic clustering may be explained by coal rank (coal age/hardness) and variations in mining methods, safety methods, regulatory enforcement, other geological factors, and worker age.

IV. CONCLUSION

While this paper demonstrates the utility and benefit of geographic examination, the occupational injury/illness field lags behind other fields in taking advantage of the emerging capacity to examine data geographically. Extensive searching was required to construct this list of references, and only seven of the articles were identified through Medline. Despite the consistent finding in geographic research that smaller aggregations improve the ability to detect associations, few publications in the past decade looked at aggregations below the state level (granted, data limitations are likely the main contributor to this gap). The top occupational injury/illness surveillance databases (Bureau of Labor Statistics Survey of Occupational Injuries and Illnesses and Census of Fatal Occupational Injuries) provide only state level data. Very few articles used spatial statistics, and few evaluated more than the most basic geographic risk factors. No identified article articulated a broad theory or conceptual framework for why occupational injury/illness risk would vary geographically. The literature does not build upon itself over time other than within institutions. And, organized efforts to promote occupational safety and health surveillance have not viewed geography as a priority.

The field is wide open for additional research and surveillance work, including developing indicators, using spatial statistics tools, developing theory, examining and displaying data at aggregations below the state level, and addressing both questions specific to industries and occupations, and those relevant to broader groups.

Tools for Research:

To stimulate additional research in this area, companion papers in this dissertation present several tools for researchers. Based on this literature review and social science literature, a ***conceptual model*** is developed to describe why occupational injury and illness would vary geographically (Neff, Curriero, & Burke, 2006). In summary, the model posits four main geographically-varying factors that affect risk: the economy; demographics; policy/culture/values; and industry mix. These risk factors affect business and worker incentives and capacity for safety through their effect on an area's balance of power between business and workers and other mechanisms. In turn, workplace-level behavior is affected, and is the proximal cause of injury/illnesses.

The papers in this dissertation also present ***examples of methods*** for analyzing one database for surveillance, hypothesis-generating and hypothesis-testing purposes. In addition, below I suggest just a few of the ***research issues*** worthy of additional exploration, in addition to replicating my thesis analysis in other databases.

- **Risk factors related to workplace location –** transportation routes, urbanization, weather, community characteristics, local policy or government characteristics, differences in local raw materials used in production, international geographic variation, natural features, elevation, proximity to health services, urbanization, walkability, business headquarters location, area industrial composition, change in industrial composition
- **Worker geographic risk factors –** area demographics, worker residence area, worker area of origin, commuting issues
- **Tools for practitioners –** indicator development, temporal tracking of rates by area, area risk factors that could drive program planning, evaluation of interventions in context of area risk factors.
- **Other –** documenting the geographic spread of "safety incentive" programs and

other work organization conditions with safety and health implications, geographic diffusion of safety innovation, determinants of where in a community or workplace events happen, production distance from suppliers or distributors, original location of inputs used in production.

Lastly, I list ***existing databases*** that might be explored geographically for this research and surveillance. (Table 6, next page)

Following are several possible explanations I theorize for the current underdeveloped state of the literature. Most importantly, it is difficult to find geo-referenced data. In addition, some have felt that given the current state of surveillance, there are other pressing needs. Further, the organization of services in occupational injury/illness has meant that OSHA is the nation's primary intervention provider. Since its established targeting methods do not make use of geographic analysis, the apparent benefit of geographic tools for occupational injury/illness intervention planning is diminished. Researchers using geography in other domains of public health have mostly not sought out occupational injury/illness work, perhaps because of the field's relative marginalization within public health. Within the occupational safety and health field, my informal conversations suggest that some are skeptical about the potential benefits of geographic analyses, or believe that mapping would not show much beyond the expected industrial and population distribution (and they do not yet recognize the benefits of visual tools that show this information). Lastly, to the extent many types of geographic studies take a big-picture approach to understanding problems and their determinants, this literature gap is reflective of a more generalized need for increased cross-cutting and contextual research in occupational safety and health.

Table 6: U.S. occupational injury/illness-related databases that might be used for geographic analysis

This is a list of *possibilities* for further exploration. Table partly based on databases described in NIOSH Worker Health Chartbook, Appendix A, 2004(NIOSH, 2004)[2]

Database	Main Relevant Information	Smallest Aggregation Known to be Available*	Comments on Possibility of Smaller Aggregation Availability	Source
ABLES (Adult Blood Lead Epidemiology and Surveillance) Program	Blood lead levels, demographics	State	States report data, mainly based on laboratory/physician findings. Smaller aggregation might be feasible, at least in some states.	http://www.cdc.gov/niosh/topics/ABLES/ables.html
Behavioral Risk Factor Surveillance System (BRFSS)	Data on multiple health-related behaviors and health outcomes	State, Metropolitan micropolitan statistical areas (MMSA)		http://www.cdc.gov/brfss/about.htm
Cancer Mortality Data	Cancer mortality from multiple causes. Can be linked with Census demographic information by county.	County		http://www3.cancer.gov/atlasplus/
U.S. Census	Numerous workforce, social, demographic, economic, and other variables	Census tract		http://www.census.gov
Census of Fatal Occupational Injuries (CFOI)	Occupational fatalities and information about the work, worker, event	State	States report data. Smaller aggregation might be permitted, at least in some states.	http://www.bls.gov/iif/oshcfoi1.htm
1998 Childhood Agricultural Injury Survey (CAIS)	Childhood agricultural lost-time injuries, details on 4 most recent injury events	Region	Phone survey with all agricultural production operations in area sampling frame. Thus there is not reason to think smaller aggregation would be biased,	NIOSH, Surveillance and Field Investigations Branch, Division of Safety Research. Tel. (304) 285-

[2] Some of these databases are not routinely analyzed in geographic fashion and others are usually only analyzed at the state or regional level due to small sample sizes at smaller aggregations. In these, it may be possible to negotiate release of data at smaller aggregations based on the following restrictions: de-identifying data before providing it; combining multiple years; or only using data from limited parts of the country where coverage is greater.

Database	Main Relevant Information	Smallest Aggregation Known to be Available*	Comments on Possibility of Smaller Aggregation Availability	Source
			though numbers might be small in some areas.	5916
Coal Workers' X-Ray Surveillance Program (CWXSP)	Chest x-ray, information about miner and work history	County	Has been analyzed at the county level (Antao et al., 2005)	NIOSH, Coal Workers' X-Ray Surveillance Program Activity, Division of Respiratory Disease Studies, Tel. (304)285-5724
Compressed Mortality File - NIOSH	92 causes of death with potential occupational causation, by sex, age and race from 1960-1989.	County of residence at time of death		http://www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm
Current Population Survey (CPS)	Population estimates for employment, unemployment, unionization, etc., including by demographic and work subgroup. Sometimes includes relevant supplemental questions	State, "other geographic areas"		www.bls.gov/cps
Economic Census	Profile of the U.S. economy. Includes sector-specific reports.	Zip code for some variables		http://www.census.gov/econ/census02/
Hazardous Substances Emergency Events System (HSEES)	Information about hazardous substances emergency events, some data on worker exposure; these are also an environmental parallel to workplace unintentional events ("accidents." _	Longitude and latitude of event (15 states)		http://www.atsdr.cdc.gov/HS/HSEES/hsees.html
Local Employment Dynamics (run by Census Bureau)	Data on employment, job creation and turnover, and earnings, by industry, age and sex	County, sub-county (not all states yet)		http://lehd.dsd.census.gov/led/led/led.html
Mine Safety and Health Administration Accident/Injury/ Illness Database	Data about injured/ill worker, work, the injury/illness (including narrative)	County		http://www.cdc.gov/niosh/mining/data/

Database	Main Relevant Information	Smallest Aggregation Known to be Available*	Comments on Possibility of Smaller Aggregation Availability	Source
Mine Safety and Health Administration Employment database	By mine: injury count and data on production and employment by type of work	Street address		http://www.cdc.gov/niosh/mining/data/
2000 Minority Farm Operator Childhood Agricultural Injury Survey (M-CAIS)	Childhood agricultural lost-time injuries, details on 4 most recent injury events	Region	Voluntary census survey of all "racial minority and Hispanic" farms. To account for refusal to participate, adjusted only to region level, so smaller aggregation not possible.	NIOSH
Multiple Cause of Death Data	Demographic and medical information.	County of residence at time of death		http://www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm
National Agricultural Workers' Survey (NAWS)	Demographics, legal status,	12 agricultural regions	288 counties in 25 states are sampled to reflect these 12 regions	http://www.doleta.gov/agworker/naws.cfm
National Notifiable Diseases Surveillance System (NNDSS)	Notifiable disease incidence	State	Reported by states; smaller aggregations might be permitted, especially in some states	http://www.cdc.gov/epo/dphsi/nndsshis.htm
National Occupational Respiratory Mortality System (NORMS) *(includes the National Surveillance System for Pneumoconiosis Mortality (NSSPM))*	NCHS multiple cause of death data for selected respiratory conditions; demographic and medical info from death certificates	County	Due perhaps to small numbers, most NIOSH analyses and the online data query present data at the state rather than county level.	http://webappa.cdc.gov/ords/norms.html
National Traumatic Occupational Fatality (NTOF) System	30 risk factors describe decedents, work, injury including narrative text. (NIOSH chartbook uses CPS data to	State	Death certificate data are available for counties and cities with (>250,000 residents) – not clear whether this is	NIOSH

Database	Main Relevant Information	Smallest Aggregation Known to be Available*	Comments on Possibility of Smaller Aggregation Availability	Source
	estimate employed group denominator for NTOF data.)		available for NTOF also.	
OSHA Data Initiative	Injury/illness rate for establishments in high-hazard industries	Street address		Some data at http://www.thememoryhole.org/osha/lwdii.htm. Additional data from OSHA.
Sentinel Event Notification System for Occupational Risk (SENSOR)	Topics include: asthma, silicosis, carpal tunnel, noise-induced hearing loss, pesticide poisoning; working teens. Multiple programs in different states.	Some states probably collect data on location		Multiple. See NIOSH Worker Health Chartbook pp.306-7 for contacts.
Survey of Occupational Injuries and Illnesses (SOII)	Nonfatal occupational injuries/illness rates, counts, characteristics, severity - by employer	State	Survey uses probability sampling by state. Researchers could try requesting data for analyses at smaller aggregations. Analyses would need to be performed at BLS office.	www.bls.gov/iif
Toxics Release Inventory	Toxic chemical releases to the environment; may reflect worker exposures as well.	Zip code		http://www.epa.gov/triexplorer/
Traumatic Injury Surveillance of Farmers (TISF) survey	Agricultural injuries – details about the person, work, injury, and causes.	State	Sampling performed by state – not clear if smaller aggregations possible.	www.cdc.gov/niosh/injury/traumaagric.html
Workers' Compensation	Injuries, illnesses, sequelae, costs, probably demographics, industry/occupation.	County might be possible in some states		Each state runs its own system and would need to be contacted separately.

For additional suggestions of area level data sources, see (Hillemeier et al., 2003).

There are true limitations to geographic approaches. Mapping is a limited tool, with refinement declining as aggregation increases. Further, because workers commute to workplaces from some distance away, very small geographic aggregations such as those surrounding only workers or only workplaces may miss important effects, so larger aggregations are often needed. In addition, maps can be deceptive, for example by visually equating areas with different background populations, and by failing to account for the widely differing areas of Western versus Eastern U.S. counties. There is a need to avoid the ecological fallacy of assuming that a finding about an area applies to all individuals within that area; risk is typically inequitably distributed. Finally, the same social risk factors that may affect injuries and illnesses differentially by geography can also contribute to differential underreporting (Azaroff et al., 2002; Azaroff et al., 2004).

In response to all these concerns, it is important to note the benefits of geographic research for this field. Findings from papers cited in this review have led to critical new insights in cancer and respiratory disease causation, sparked effective intervention including improved targeting, clarified and visually dramatized information for policy makers and the public, supported the development of theory, and built understanding of social and other contributions to multiple types of occupational injury and illness.

Every week, 105 people in the U.S. die from work-related injury and 938 from illness, while work injuries and illnesses send 77,000 workers to emergency rooms on average (based on Schulte, 2005). Today the literature on geographic variation in occupational injury and illness is "all over the map." A better understanding of the geographic spread and determinants of these events can help provide "driving directions" to streamline the path to prevention.

3

"In the Wrong Place…?": Spatial Tools for Occupational Injury/Illness Surveillance

Roni A. Neff, ScM, Thomas A. Burke, PhD, MPH, and Frank Curriero, PhD

ABSTRACT

BACKGROUND: The more effectively public health programs can characterize the extent, determinants of and variation in health conditions, the better they will be at shaping, targeting and evaluating interventions and at communicating with the public and policymakers about needs and priorities. Spatial analysis has become integral to mainstream public health surveillance. However, it has been underused in the occupational injury/illness field.

INTRODUCTION: This paper demonstrates methods for examining geographic data on occupational injury/illness through an analysis of county-level rates in the Occupational Safety and Health Administration (OSHA) Data Initiative (ODI), 1997-2001. Methods including mapping, spatial statistics, and ranking are presented to address the following four questions: Does occupational injury/illness vary spatially at the county level?; Does variation remain after accounting for industry hazard?; Where are rates higher or lower than expected?; and What social risk factors (e.g., economy, demography, policy/culture/values) seem to covary with occupational injury/illness? Finally, we ask whether it is appropriate to use the ODI for surveillance.

RESULTS: There is evidence that reported nonfatal occupational injury/illness rates vary by geography in the ODI, including after adjusting for industry hazard. Key areas of high rates are in West Virginia-Kentucky, Michigan-Wisconsin, Northern New England, and Northern California. The South has low reported rates. Social risk factors covary somewhat with occupational injury/illness in this sample. These findings are similar to other analyses of nonfatal occupational injury/illness distribution.

CONCLUSION: These analyses can provide data to improve intervention targeting, suggest risk factors for further investigation, and make the case for targeting resources to prevention in hard-hit areas so that one day, “the wrong place” can be transformed into just “a place.”

***   *   *   *   ***

I. INTRODUCTION

When a worker’s injury or illness is attributed to bad luck, it is often said that he or she was just “in the wrong place.” But what if a place is “wrong” for a lot of workers? This paper builds from the premise that most work-related injury and illness can be prevented, and that the more effectively public health programs can characterize the determinants, variation, and extent of health conditions, the better they will be at shaping, targeting and evaluating interventions and at communicating with the public and policymakers about needs and priorities. Spatial analysis has become integral to mainstream public health surveillance (Cromley & McLafferty, 2002; Hillemeier et al., 2003; Krieger et al., 2002; Krieger, et al., 2003; Waller & Gotway, 2004). However, it has been underused in the occupational injury/illness field. In particular, few reports look below the state level (Neff, 2006). This paper demonstrates a set of geographic methods for examining data on occupational injury/illness through an analysis of county-level nonfatal injury/illness rates in the Occupational Safety and Health Administration (OSHA) Data Initiative (ODI), 1997-2001.

The Centers for Disease Control and Prevention (CDC) defines public health surveillance as “the ongoing systematic collection, analysis and interpretation of outcome-specific data for use in the planning, implementation and evaluation of public health practice.” (Thacker, 2000). Thacker defined three main types of public health surveillance: hazard, exposure, and health outcome (Thacker et al., 1996). Each can be examined in spatial context.

Spatial analysis (using maps and/or spatial statistics to examine distributions) improves surveillance by: increasing the ability to target programs to areas of need and to design and evaluate programs relevant to local risk factors and issues; helping policymakers and program managers understand and contextualize issues relevant to their jurisdictions; and contributing to hypothesis generation about causative factors. The U.S. government has set the Healthy People 2010 objective of “Increas[ing] the proportion of all major national, state and local health data systems that use geocoding [geographic identifiers that enable mapping] to promote nationwide use of geographic information systems (GIS) at all levels." The target is 90 percent (Objective 23-3)(Krieger et al., 2002; U.S. Department of Health and Human Services, 2000).

Why would occupational injury/illness rates vary geographically? Our conceptual framework, developed in a companion paper (Neff, Curriero, & Burke, 2006), posits four main geographically-varying risk factors: the economy; demographics; policy/culture/values; and industry mix. These affect business and worker incentives and capacity for safety through their effect on an area's balance of power between business and workers, among other mechanisms. In turn, workplace-level behavior is affected, and is the proximal cause of injury/illness. "Behavior" includes actions directly related to safety and health and those that influence them unintentionally, such as promoting unhealthy work organization.

In this paper, the terms, "geographic" and "spatial" are used interchangeably to refer to the variation across space in risk factors or health outcomes. "Area-level" variation refers to variations between discrete geographic areas such as counties. Area-level characteristics can reflect the aggregated effect of individual characteristics (such as the summed experience of workers in a high hazard industry that dominates an area) or the holistic effect of being in an area with particular conditions (such as a cultural view that occupational injury is normal). Some characteristics such as inequality and segregation only make sense when examined at the area level. While most exposures considered here are not primarily shaped by political boundaries such as counties, such outlines are used for convenience as well as policy relevance of findings.

Four key spatial analysis concepts are relevant for this paper. *Global spatial* *dependence* or *autocorrelation* means that across a whole map, nearby areas are more likely to be similar to each other than to those farther away, that is, that the variation in the map has a clustered pattern rather than just being a random array. *Spatial variation in risk* refers to the extent to which risk or odds varies spatially, often after controlling for key risk factors through regression. An associated paper examines this (Neff, Curriero, & Burke, 2006). By contrast with global properties, local properties of a map focus on particular areas. *Cluster detection* involves identifying local "hotspots" on a map. Lastly, the viewer's own interaction with the map by comparing findings in areas of interest with those elsewhere (the so-called *"'where's my house?'* phenomenon")(Waller & Gotway, 2004) contributes to a map's policy implications.

II. BACKGROUND

In the 1970's and 1980's, occupational safety and health was in the vanguard of using geographic analysis for public health surveillance and hypothesis generation. National Cancer Institute (NCI) and National Institute for Occupational Safety and Health (NIOSH) staff were active in finding ways to present information in geographic context, generally at the county level. For example, they combined data from the National Occupational Hazard Survey (NOHS, 1974) with data on industry location and chemical toxicity to map occupational health hazards and exposures by U.S. county (Frazier & Sundin, 1986; Stone et al., 1978; Sundin & Frazier, 1989). The surveillance circle was completed with county-level health outcome data from the 1975 and subsequent Atlases of Cancer Mortality and from the National Occupational Respiratory Mortality System (NORMS) (Devesa, et al., 1999; Kim, 1998; Mason et al., 1975; Mason et al., 1976; NIOSH, 2006a; NIOSH; Pickle et al., 1987; Pickle et al., 1990).

These surveillance tools were used widely in public health practice (Burke, T. personal communication, May 2006). Further, they spawned numerous research studies aimed at hypothesis generation and later, hypothesis testing (Blot & Fraumeni, 1976; Blot & Fraumeni, 1977; Blot et al., 1977; Brinton et al., 1976; Frazier & Sundin, 1986; Frazier et al., 1983; Hoover & Fraumeni, 1975; Hoover et al., 1975; Stone et al., 1978). Insights include the associations between: smelter worker arsenic exposure and lung cancer; shipyard worker asbestos exposure and lung cancer; furniture workers and nasal cavity cancers; and truck drivers and bladder cancer (Devesa, et al., 1999).

Since the mid-1980's, geographic analysis has played a much less prominent role in occupational injury/illness surveillance, even as it has become mainstream in other areas of public health. Expert recommendations for surveillance in the field have not prioritized geographic analysis. For example, NIOSH's 2001 report on surveillance priorities did not mention mapping or gathering geo-referenced data, although it did include some priorities *relevant* to geographic surveillance, such as creating data useful for localities and examining high risk populations (NIOSH, 2001).

In the past few years, the Council on State and Territorial Epidemiologists (CSTE) designed and piloted occupational injury/illness indicators for use by states. No geographic indicators were included, and no recommendation was made for collecting or using data that would enable geographic comparison within states (CSTE in collaboration with NIOSH, 2004; CSTE, 1999). John P. Sestito, J.D., M.S., NIOSH Surveillance Program Coordinator, recalled that the possibility of geographic indicators was not formally discussed during planning of either project. He theorizes that this was because data at relevant aggregations are not available in many states, and because geographic display and analysis are often perceived as a second-level issue that might be addressed after covering basic priorities (Sestito, J., personal communication, April, 2006).

While not treated as a priority, some basic geographic tools, particularly state level counts and rates, continue to be used in occupational safety and health surveillance today. These tools inform policy and practice. The most comprehensive and ambitious national surveillance report of occupational injuries and illnesses is NIOSH's *Worker Health Chartbook*. The *Chartbook* presents state-level maps of numerous health outcomes, including subsets by industry, drawn from a variety of surveillance data sources (NIOSH, 2004). In a valuable innovation, NIOSH provided both graphics and underlying data in multiple electronic formats. There are other state-aggregated reports, including: (Adekoya & Pratt, 2001; AFL-CIO, 2006; NIOSH, 1991; NIOSH, 1993; U.S. Dept. of Labor, Bureau of Labor Statistics, 2003; Windau et al., Summer 1998). An example of the policy impact of these reports comes from NIOSH's 1993 chartbook on National Traumatic Occupational Fatality data. Its maps showing Alaska's fatality rate to be far above the U.S. rate led to a multiagency collaborative on the issue – today considered a major success story for the field (NIOSH, 1993; Smith, 2001).

Some *county-level* surveillance tools also remain current; advances in Internet technology have enabled provision of online interfaces allowing users to generate county-level maps to their own specifications for the Cancer Atlas, respiratory mortality, and mortality from causes of death that have been associated with occupation. At least two states have also presented county-level maps in recent years (New Jersey Department of Health and Senior Services, 2005; Stanbury et al., 2004).

By contrast with the large literature on environmental injustice, only the AFL-

CIO's annual "Death on the Job" series of reports was identified as examining injustices and disparities in occupational safety and health from a geographic standpoint. They find large differences in the level of intervention by state (AFL-CIO, 2006).

To help promote additional geographic analysis of occupational injury/illness, this paper seeks to demonstrate the use of spatial tools. Methods are presented to address the following questions:

1. Does occupational injury/illness vary spatially at the county level?
2. Does variation remain after accounting for industry hazard?
3. Where are the "hotspots"? Where are rates lower than expected?
4. What social risk factors seem to covary with occupational injury/illness?

Finally, we ask whether it is appropriate to use the ODI for surveillance.

III. METHODS

A. METHODS - DATA

1) Injury/Illness Data: The injury/illness database is the Occupational Safety and Health Administration (OSHA) Data Initiative (ODI), a national survey of establishment injury/illness rates in high rate industries, conducted to improve OSHA's enforcement targeting (OSHA Directorate of Compliance Programs, 1997, 1998, 1999, 2000, 2001). We are unaware of any prior description of the ODI database in the peer-reviewed literature. From 1997-2001, OSHA annually surveyed roughly 80,000 establishments with over 40-60 employees in industries with the highest injury/illness rates in the Bureau of Labor Statistics' Survey of Occupational Injuries and Illnesses, plus all manufacturing industries. The sample excludes industries not regulated by OSHA including mining and most government workers, and the construction industry. The ODI is structured so that each eligible establishment should be surveyed at least once every three years, however, due to large numbers, nursing homes and department stores are sampled less frequently. Still, nursing homes dominate the database (15.34%). OSHA annually audits 250 establishments. The 1997 and 1998 audits found that 19 and 21 percent, respectively, of cases had “major recording errors.” Many more errors were in the direction of underreporting than overreporting (OSHA, 2001).

In the ODI, establishments report the number of injuries and illnesses in several categories and the number of full time equivalent workers (excluding contractors). From these, establishment lost workday injury/illness rates (LWDII) are calculated, reflecting the number of injuries and illnesses requiring time away from work or reassigned duties per 100 full time equivalent workers. Other basic data from logs are collected as well. Appendices 1 and 2 present the top ODI industries by number and LWDII rate. Table 7 describes the ODI database for the included years, 1997-2001.

Table 7: OSHA Data Initiative database, 1997-2001: Establishment-level lost workday injury/illness (LWDII) rates in most high-rate industries

YEAR	N (after exclusions*)	LWDII mean (SD)	# not Geocoded (%)	Response rate	National Average LWDII**
1997	44,199	7.41 (6.68)	544 (1.2)	NA	3.3
1998	36,043	8.14 (6.77)	332 (0.9)	91%	3.1
1999	42,270	7.66 (6.45)	369 (0.9)	95%	3.0
2000	45,026	7.30 (6.21)	232 (0.5)	96%	3.0
2001	49,308	5.93 (5.65)	434 (0.9)	94%	2.8
TOTAL	216,846	7.22 (6.38)	1,941 (0.9)	NA	

* Exclusions: HI, MN, OR, SC, WA, WY; establishments with <60 employees; establishments reporting LWDII>50
** National average from Bureau of Labor Statistics, covering all private industry

We combined five years of data for analyses, based on OSHA recommendation (J. Dubois, personal communication, February 2006) and the benefits of increased sample size. As will be described, some ranking and sensitivity analyses were segmented by year. Results are reported acknowledging that some establishments were counted more than once because there are no unique identifiers. A further limitation is that the sampling strategy varied across years, and sensitivity analyses did show possible differences in the sample across years. To improve our sample consistency, we limited it to establishments with at least 60 employees and excluded six states due to low numbers: Hawaii, Minnesota, Oregon, South Carolina, Washington, and Wyoming. (Participation in the ODI is optional each year for these and other states that run their own OSHA programs.)

A remaining concern is that in some years OSHA excluded establishments reporting LWDII under 7.0 in the prior two surveys, so in areas with many such establishments, there would be differential misclassification bias and/or different numbers of establishments sampled by year. Finally, OSHA recommended excluding the 0.1% of establishments reporting LWDII above 50 per 100 employees because of the likelihood of error. (J. Dubois, personal communication, February 2006). Appendix 3 presents maps of ODI data by year, and Appendix 4 presents sampling differences by year and an evaluation of the impact of excluding low rate establishments.

Establishment locations were geocoded (linked to maps) to zip codes and assigned to counties in which their zip code centroids were located. Establishments with data that could not be geocoded were not used in the analysis. (Appendix 5 compares them with geocoded establishments.)

A year- and county-aggregated file was created and merged with the database of risk factors (see below). The outcome variable, "mean LWDII," was calculated to reflect the mean reported LWDII of all responding establishments in a county.

2) County and State Risk Factors: A search was conducted for county and state economic, demographic, policy/culture/social, and industry risk factor variables related to the conceptual framework categories, and 90 were gathered. Selected risk factors included in this paper are in Table 8 below.

Table 8: Risk factors used in mapping and their distribution among ODI counties
(N=2,657 counties of 3,141 total).

RISK FACTOR	MEAN (range)	SOURCE
Demographics		
% African American	9.29 % (0, 86.1)	Census 2000. Calculated
% White	84.45 % (12.2, 99.8)	Census 2000. Calculated
Local Economy/Socioeconomic Status		
% Less than poverty	14.07 % (2.1, 50.1)	Census 2000. Calculated
Index of Area Industry Hazard		
"Expected" mean LWDII if every ODI estab. had its industry mean	5.94 (0.7, 12.65)	ODI and Bureau of Labor Statistics, calculated.

Because of the importance of industry in determining hazard levels, an "Index of Area Industry Hazard" was developed to reflect the rates that would be expected based on county industry mix alone. The Index is calculated by assigning each establishment its industry mean (or "expected" rate based on industry), then calculating a county average of these rates. (Based on methods suggested by Lynn Goldman and Gordon Smith in personal communications (November 2004.)) For "industry mean," the Index uses rates reported by the Bureau of Labor Statistics (BLS), Survey of Occupational Injury and Illness (SOII) in the corresponding year. The SOII has more consistent sampling procedures and less incentive for underreporting than the ODI, so its findings might be expected to reflect a rate closer to the true industry mean.(See Appendix 6 for R code.) For comparison, a second Index of Area Industry Hazard was calculated using ODI data instead of SOII data for industry mean, as a purer reflection of within-database industry hazard without the issue of database differences.(see Appendix 7)

Analyses were conducted at the county level (and state level where justified.) Counties are sizable enough to reflect social, political, economic and demographic risk factors in a meaningful way. In the mean U.S. county, 67.4% of workers 16 and over work in the county where they live (U.S. Bureau of the Census, 2000). In addition, most local health departments operate at the county level, giving this level practical relevance.

B. METHODS -- ANALYSES

1. Does occupational injury/illness vary spatially at the county level? A set of maps was developed to describe the database. Map 2 presents state mean occupational injury/illness rates. To demonstrate the increase in information from showing data at the county rather than state level, Map 3 presents the same information in county aggregation. These and other maps are presented in quintiles unless noted.

To examine whether there is a pattern to the observed variation or whether rates are just randomly distributed, testing for global spatial autocorrelation is performed. Moran's I, comparable to a correlation coefficient, is a widely used method (Waller & Gotway, 2004). It reflects the correlation between values in one location with those in other locations, generally taking the form:

$$\frac{\text{Sum for all location pairs: } \dfrac{(\text{proximity}) * (\text{similarity})}{\text{proximity}}}{\text{sample variance}}$$

Moran's I was computed for Map 3. To compute a pseudo-significance level, 9999 permutations were performed. A "queen" (as in chess) weight file was used, comparing a county's mean with the average of means in all counties that shared boundaries or vertices with it. Sensitivity analyses tested other inputs.

2. What can be learned from comparing observed and expected rates by industry?

Industry hazard is one of the strongest predictors of occupational injury and illness (Waehrer et al., 2004), so taking it out of the equation may help identify areas with other important factors contributing to area rates. Map 4 shows the SOII Index of Area Industry Hazard and Map 5 shows the diversity in "observed mean" divided by "expected mean," reflecting areas where problems are greater or lesser than might be expected based upon industry hazard. Appendix 8 shows the equivalent map using the ODI Index of Area Industry Hazard. A table is also presented to show the geographic variation in top industries. The top ten industries in each of the five highest and lowest rate states, and across the sample, are listed.

3. Where are the "hotspots"? Where are rates lower than expected?

Local Indicators of Spatial Autocorrelation (LISAs) involve assigning values to each county based on their similarity with other nearby counties. Counties are identified as "high-high" (red) – meaning a high rate county for which the mean rate in surrounding counties was also high, or "low-low" (blue) – low rate counties surrounded by the same. Because each of the red and blue counties is surrounded by other counties that also have high or low rates, the depicted cluster areas could be considered just the cluster cores, with the cluster boundaries extending outward to all counties sharing borders or vertices with them (Anselin, 2003). The mean of all LISAs in a map is proportional to its global Moran's I (Anselin & Koschinsky, 2006). LISAs were computed for Map 3 using a queen weight file, statistical significance criterion of $p<0.05$, and 9999 permutations. These inputs were varied to determine their impact on Moran's I, and because results varied little, the above default methods were used.

A related tool of interest for surveillance is ranking areas based on their injury/illness rates to aid in prioritization and to motivate intervention. We present ranks based on four methods. First, we show *states* with the highest and lowest mean LWDIIs, then top-ranked *counties* with the highest means (many counties had means of 0 so low ranks are not presented.)

The third method is motivated by the fact that in contrast to states, the highest means at the county level are likely to be based on small numbers of establishments and influenced by outliers. While this potential distortion affects all estimation, it is especially problematic when focusing on the most extreme counties rather than, for example, the top 1/5 of counties. Accordingly, Louis and Shen (1999) recommend using a Bayesian statistical method to "shrink" ranks towards the mean, with those counties having the largest variance in their means getting shrunk the most.[3]

An important concern is that county means may incorporate multiple observations on the same establishment. Since establishment rates tend to be consistent across years,(Hunt, 1993) variances in counties with many repeat observations will be inappropriately low, and thus the Bayesian shrinkage will yield too-extreme ranks. While this problem of repeated observations is a concern throughout this paper, it is especially important in Bayesian ranking because variance plays a central role. Accordingly, Bayesian ranking was performed separately by year. To select a year to report in this paper, top 10 lists for ">30 establishments" (see below) were compared across years. 1998 had the highest overlap with other years. The top ten Bayesian-ranked counties for 1998 are reported with their direct ranks juxtaposed for comparison.

The final method addresses the concerns that even using the Bayesian method,

[3] The Bayesian ranks are not always integers, so Shen and Louis recommend ranking the ranks themselves to produce an ordered set of integer ranks they call "R-hat" ("R" for rank, and "hat" meaning it is estimated).

most of the highest ranked counties have under five observations and further, that there was little consistency across years in the top ten list of Bayesian-ranked counties. Accordingly, we report top-ten direct ranks for counties having over 30 establishments in the sample and provide contrasting Bayesian ranks. This method produces the most stable ranks, with the same counties and states consistently in similar rank levels across years. It also focuses on counties with significant industry, where intervention may be especially valuable, although it does exclude almost 90 percent of the 2,657 counties.

4. What social risk factors seem to covary with occupational injury/illness? While several methods of comparison have been used to map risk factors in conjunction with outcome variables, most require complex symbology or looking back and forth at multiple maps. A simple exploratory approach is to show how LWDII rates vary among counties with high levels of selected risk factors (developed in collaboration with Sue Baker, 2006). We assign each county three dichotomous variables: “top 1/3 for risk factor x”(coded: 0/1); “middle third mean injury/illness rate” (coded: 0/3); and “top third mean injury/illness rate” (coded 0/5). These are summed to create new variables, as follows:

- High risk factor, LOW injury/illness rate (sum=1)
- High risk factor, MEDIUM injury/illness rate (sum=4)
- High risk factor, LOW injury/illness rate (sum=6)

To focus on the key information, a white background is used both for counties and states not in the database, and for those falling in the bottom 2/3 for the risk factor. We emphasize that these maps simply depict co-location of potential risk factor and outcome; they do not represent controlled analysis.

The number of “overlaps” between top tertile of the risk factor and of LWDII can be counted and compared with what might be expected if both categories were distributed geographically randomly, calculated as (n1/N * n2/N)*N, where n1 and n2 represent the number of observations in the top third of each variable and N is the total number of included counties.

Numerous risk factors were initially mapped. We present data on % living below the poverty line, % African American, and % Caucasian. These were selected based on the results of our associated multivariate regression analysis (Neff, Curriero, & Burke, 2006) and their correspondence with the other maps. For context, the tertiles of occupational injury/illness are also mapped.

Software: Mapping and some of the geographic analyses were performed using ArcGIS version 9.0 (ESRI, 2005). Stata Intercooled version 9.1 (StataCorp, 2005) and Microsoft Excel 2003 were used for calculations. The freeware program, Geoda, was used for calculation of Moran's I and LISAs.(Anselin, 2003) Finally, a code authored by Louis in the R statistical package was used to generate the Bayesian ranks.(Louis, 2006; R Development Core Team, 2005) and a code by Curriero in R was used to generate the Index of Area Industry Hazard. Appendix 9 presents steps for performing many of the analyses in this paper.

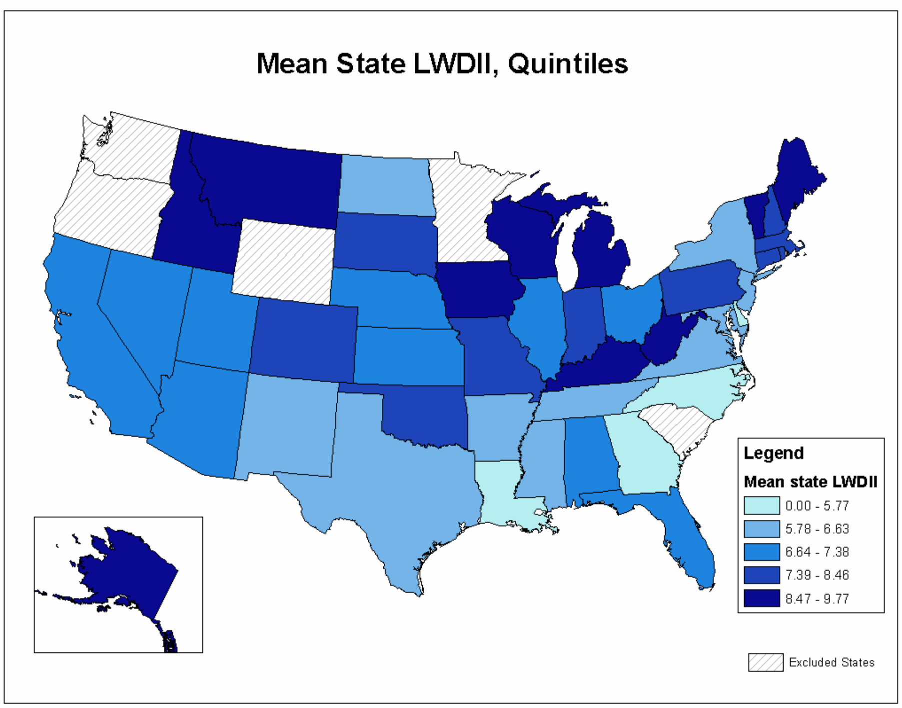
IV. RESULTS

The 216,846 establishments in the 1997-2001 ODI sample were located in 2,657 of the 3,141 U.S. counties. The number of establishments per county ranged as high as 5,685 (Los Angeles), although the median was 25 and the 90th percentile was 178. With establishments reporting LWDII over 50 excluded, county mean LWDII ranged from 0.00 to 25.196. The mean was 7.176 (Standard Deviation 2.93). Sixteen counties had mean LWDIIs of zero, all based on small numbers.

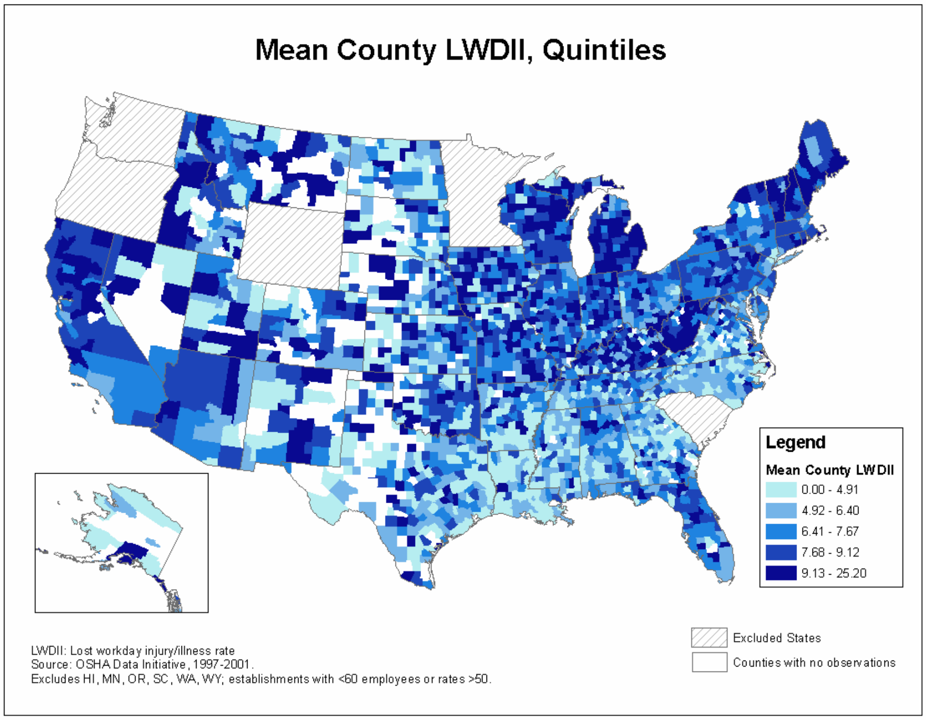
1. Does occupational injury/illness vary spatially at the county level? Map 2 depicts the national distribution of mean LWDII rates for states included in the ODI database, while Map 3 presents county mean rates. Both maps show rates to be especially high in Kentucky - West Virginia, Michigan - Wisconsin, New England, and California, and relatively low in the South.

Map 2's higher aggregation makes some patterns less clear, such as the widespread low rates in the South and elevations in Northern California. Further, of the ten states in the top quintile, four (Alaska, Idaho, Montana, and Iowa) are shown in Map 3 to have varied rates and multiple excluded counties. Map 3 also shows that, presumably due to industry distribution, sampling is somewhat sparse in the middle of the country.

Map 2: Mean state LWDII, ODI 1997-2001, quintiles

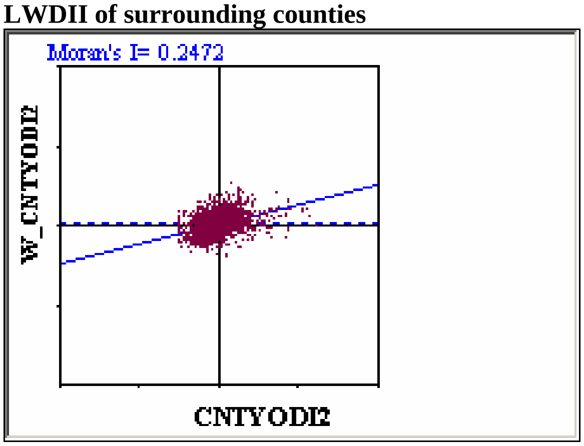


Map 3: Mean county LWDII, ODI 1997-2001, quintiles



The Moran's I for Map 3, mean county LWDII, was 0.25, showing positive spatial autocorrelation. This value was significant at the $p<0.001$ level. Figure 3 presents this statistic graphically, depicting a county's rate on the x axis versus the mean of the rates of all counties with shared borders or vertices on the y axis. Sensitivity analyses showed similar values when different weight matrices were used for the y axis.

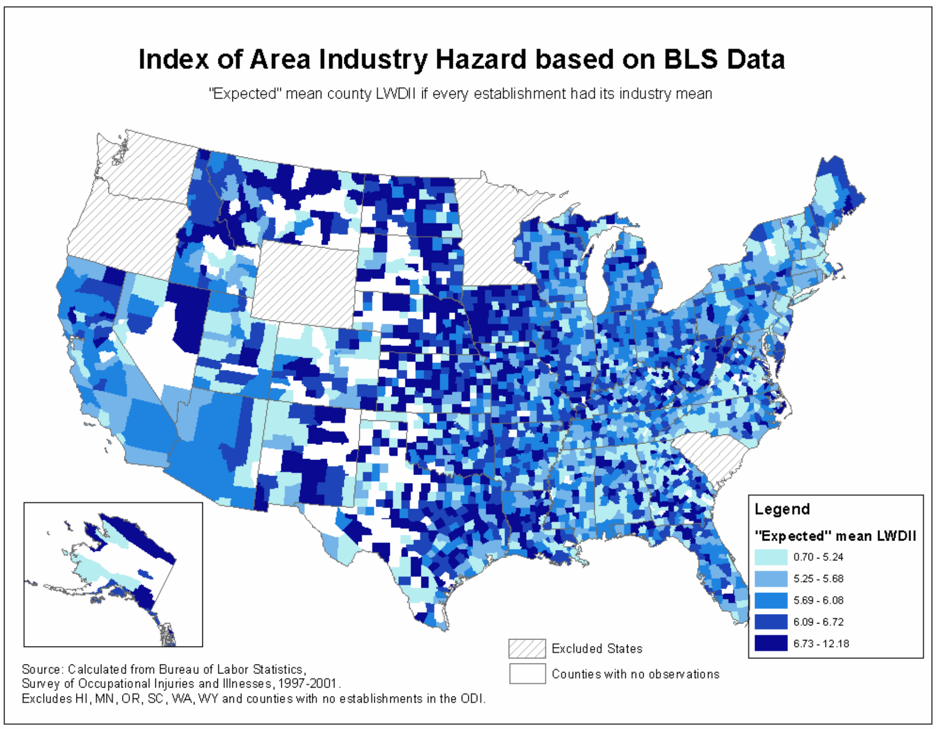
Figure 3: Moran's I test for global spatial dependence: Mean county LWDII compared to mean LWDII of surrounding counties



2. What can be learned from comparing observed and expected rates by industry?

Map 4 presents the Index of Area Industry Hazard. It shows the highest expected rates near the center of the country. This map does not strongly highlight the South as an area of expected low rates, nor New England, California or Michigan-Wisconsin as areas of expected high rates. When the Index version based on ODI industry mean (instead of SOII) was mapped, the distribution was highly similar (corr=0.86), although the ODI-based Index rates were higher on the coasts.(Appendix 7)

Map 4: Index of Area Industry Hazard based on BLS data



Map 5 contrasts the actual ODI rates ("observed") with "expected" rates based on industry hazard. The areas noted above as having especially high rates in Map 3 also had observed rates higher than expected based on industry hazard, whereas rates in the South and central areas tended to be lower than expected based on industry hazard. This difference was similar when using the ODI-based index, as shown in Appendix 8.

Map 5: "Observed" (ODI) vs. "Expected" (Industry Hazard) LWDII

(NOTE: Red/pink: Rates elevated even after accounting for area industry hazard; Blues: Rates lower than expected based on industry hazard.)

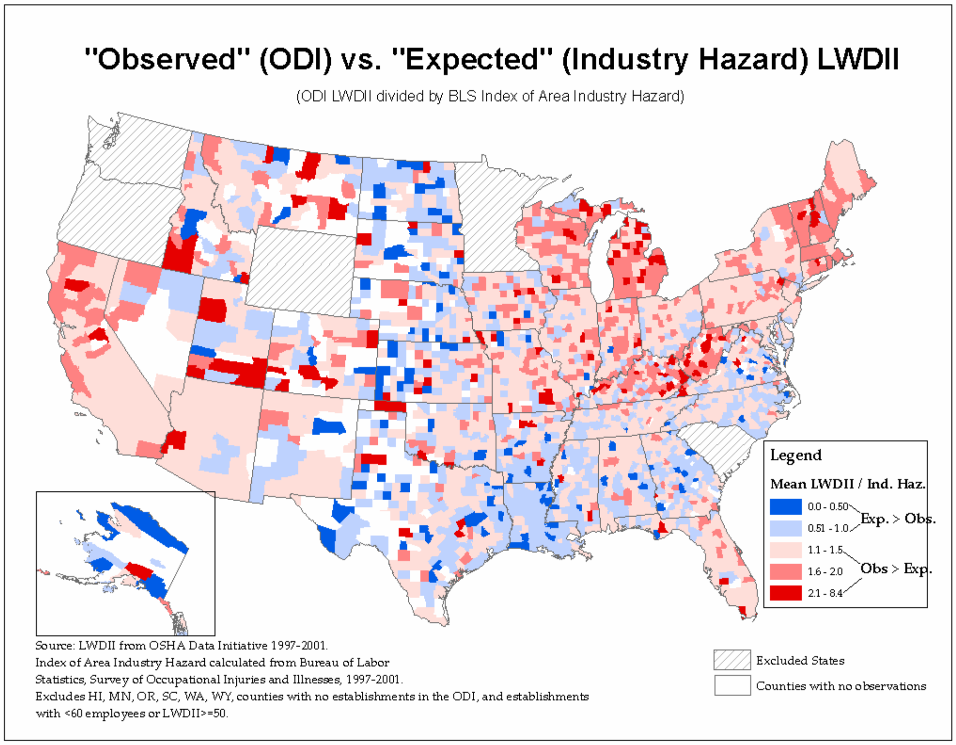


Table 9 shows the top ten industries with the most ODI establishments in the five highest rate and five lowest rate states, and across all states in the sample. It can be seen that there is substantial commonality in industry composition across the full sample; the top hazardous industries that fit the ODI sampling criteria are similar across states.

Table 9: Table #: Highest- and lowest-LWDII rate states: Variations in top 10 industries (with the most ODI establishments.)

INDUSTRY	HIGHEST RATE STATES					LOWEST RATE STATES						Number of states
	VT	WV	MI	ME	KY	LA	DC	DE	NC	GA	ALL STATES	
80: Health Services	x	x	x	x	x	x	x	x	x	x	x	11
20: Food And Kindred Products	x	x	x	x	x	x	x	x	x	x	x	11
34: Fabricated Metal Products, Except Machinery And Transportation Equipment	x	x	x		x	x		x	x	x	x	9
35: Industrial And Commercial Machinery And Computer Equipment	x	x	x		x	x		x	x	x	x	9
42: Motor Freight Transportation And Warehousing		x	x	x	x	x		x	x	x	x	9
24: Lumber And Wood Products, Except Furniture	x	x		x	x	x			x	x	x	8
30: Rubber And Miscellaneous Plastics Products	x	x	x		x			x	x	x	x	8
51: Wholesale Trade-non-durable Goods	x			x	x	x	x			x	x	7
33: Primary Metal Industries		x	x		x			x			x	5
36: Electronic And Other Electrical Equip And Components, Except Computer Equ	x		x	x					x		x	5
37: Transportation Equipment			x		x	x					x	4
25: Furniture And Fixtures	x		x						x			3
26: Paper And Allied Products				x		x				x		3
27: Printing, Publishing, And Allied Industries	x			x			x					3
22: Textile Mill Products				x					x	x		3
17: Construction Special Trade Contractors							x	x				2
28: Chemicals And Allied Products						x		x				2
15: Building Construction General Contractors And Operative							x					1
16: Heavy Construction Other Than Building Construction Contractors							x					1
31: Leather And Leather Products				x								1
32: Stone, Clay, Glass, And Concrete Products		x										1
39: Miscellaneous Manufacturing Industries							x					1
43: United States Postal Service							x					1
45: Transportation By Air							x					1
52: Building Materials, Hardware, Garden Supply, And Mobile Home Dealers								x				1
53: General Merchandise Stores		x										1

3. **Where are the "hotspots"? Where are rates lower than expected?** Map 6 (next page) depicts the statistically significant clusters identified in the LISA analysis. "High-high" counties (red), which have high rates and are surrounded by counties with high rates, are seen in Michigan-Wisconsin, Kentucky-West Virginia, Northern New England, Northern California and scattered elsewhere. Clusters of "low-low" counties (blue) are seen throughout the South, particularly in Louisiana.

Map 6: High and Low LWDII Clusters (Local Moran's I)

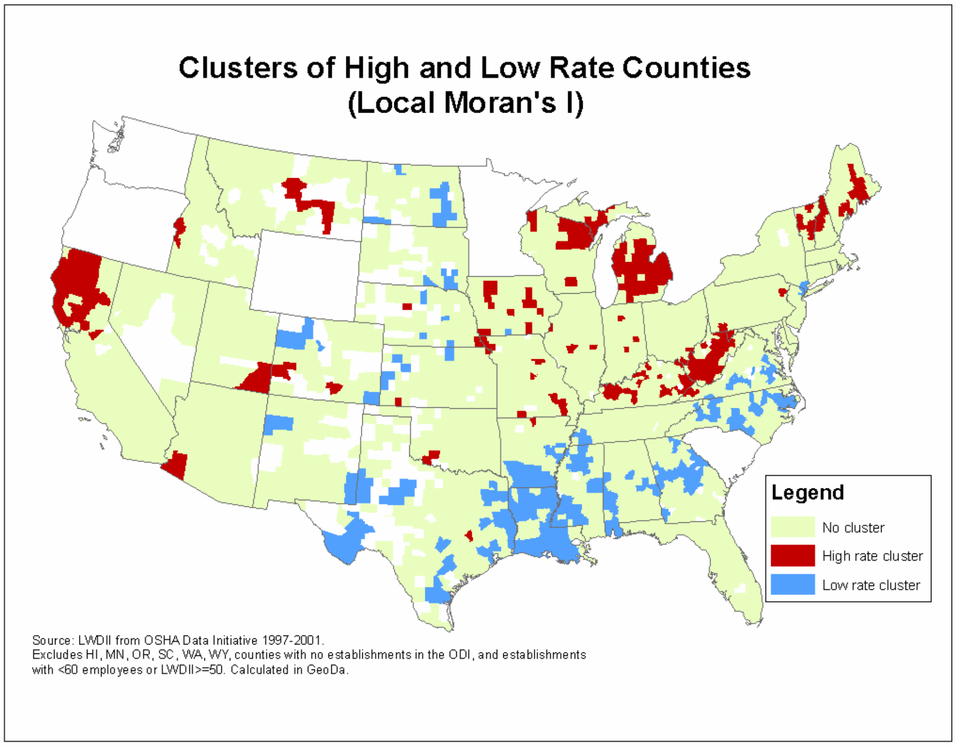


Table 10 presents states with highest and lowest mean rates. Vermont tops the list with a mean LWDII of 9.77 (N=773 establishments). West Virginia, Michigan, Maine, and Kentucky also ranked in the top five. By contrast, the lowest mean reported LWDII was in Louisiana (mean LWDII of 4.98, N=2836). Other states (and non-states) with low reported rates were Washington, DC, North Carolina, Delaware, and Georgia. (Table 11)

Table 10: States with the highest mean lost workday and injury rates in the ODI sample, 1997-2001.

State	Rank	LWDII	# Estabs in ODI
Vermont	1	9.77	773
West Virginia	2	9.76	1,481
Michigan	3	9.67	10,193
Maine	4	9.54	1,456
Kentucky	5	8.99	4,453
Idaho	6	8.92	1,062
Iowa	7	8.70	3,909
Alaska	8	8.59	271
Wisconsin	9	8.54	8,818
Montana	10	8.47	624

Table 11: States with the lowest mean lost workday and injury rates in the ODI sample, 1997-2001

State	Rank	LWDII	# Estabs in ODI
Louisiana	36	4.98	2,836
District of Columbia	37	5.16	139
North Carolina	38	5.45	9,290
Delaware	39	5.67	639
Georgia	40	5.77	7,012
New Jersey	41	5.81	6,642
Mississippi	42	5.98	2,708
Arkansas	43	6.26	3,075
New Mexico	44	6.34	354
Virginia	45	6.40	4,463

At the county level, Table 12 presents the ten counties with the highest LWDIIs based on direct means. Means in this group ranged from 20.02 to 25.2 per 100 workers.

Table 12: Top 10 counties with highest LWDII based on direct means, 1997-2001

County	State	Direct Rank	LWDII	# Estabs in ODI
Carter	Kentucky	1	25.20	5
Edmonson	Kentucky	2	23.87	5
Menominee	Wisconsin	3	23.76	5
Phillips	Montana	4	22.89	5
Martin	Kentucky	5	22.57	5
Lincoln	West Virginia	6	20.49	1
McLean	Kentucky	7	20.27	5
Robertson	Texas	8	20.18	9
San Miguel	Colorado	9	20.04	1
Gladwin	Michigan	10	20.02	14

This list is contrasted with Table 13, presenting the counties with the highest 1998 Bayesian ranks. These ranks are, in essence, adjusted from the originals based on variance in county mean LWDII. Direct ranks for these counties start at 38, and the distribution of included states is quite different from that in other lists. These counties have fewer ODI establishments than in the above lists because only one year is used and because even with low numbers, their means still had low variance.

Table 13: Top 10 counties with highest mean LWDII based on Bayesian method adjusting ranks for variance of mean

County	State	Bayesian Rank	Direct Rank	LWDII	# Estabs in ODI
Fountain	Indiana	1	53	17.53	2
Chaves	New Mexico	2	66	17.10	2
Carroll	Iowa	3	140	14.29	2
Plumas	California	4	212	13.06	3
Iron	Utah	5	136	14.37	2
Pickens	Georgia	6	669	9.83	2
Uvalde	Texas	7	514	10.51	2
Warren	Illinois	8	54	17.52	3
Shelby	Iowa	9	185	13.44	2
Iron	Michigan	10	38	18.84	2

Finally, Table 14 limits the sample to counties with at least 30 establishments in the 1998 sample (n=283 of 2,537), presenting the top direct ranked counties. Bayesian ranks are presented for comparison.

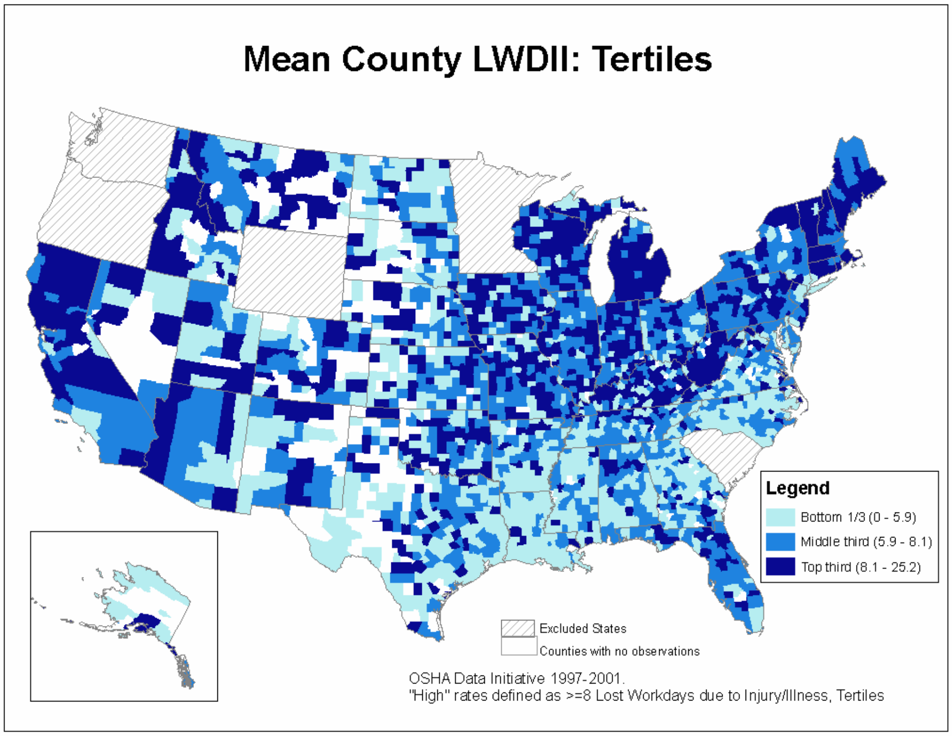
Table 14: Top 10 counties with highest mean LWDII among those with >30 establishments in ODI

County	State	Direct Rank	Bayesian Rank	LWDII	# Estabs in ODI
Genesee	Michigan	262	753	12.58	40
Sheboygan	Wisconsin	265	228	12.53	50
St. Clair	Michigan	307	636	12.02	37
New London	Connecticut	311	599	11.96	40
Wayne	Michigan	332	688	11.78	323
Rock	Wisconsin	368	656	11.46	37
St. Joseph	Indiana	387	440	11.31	64
Racine	Wisconsin	388	638	11.31	46
Muskegon	Michigan	402	412	11.22	42
Livingston	Michigan	412	513	11.16	31

5. What social risk factors seem to covary with occupational injury/illness?

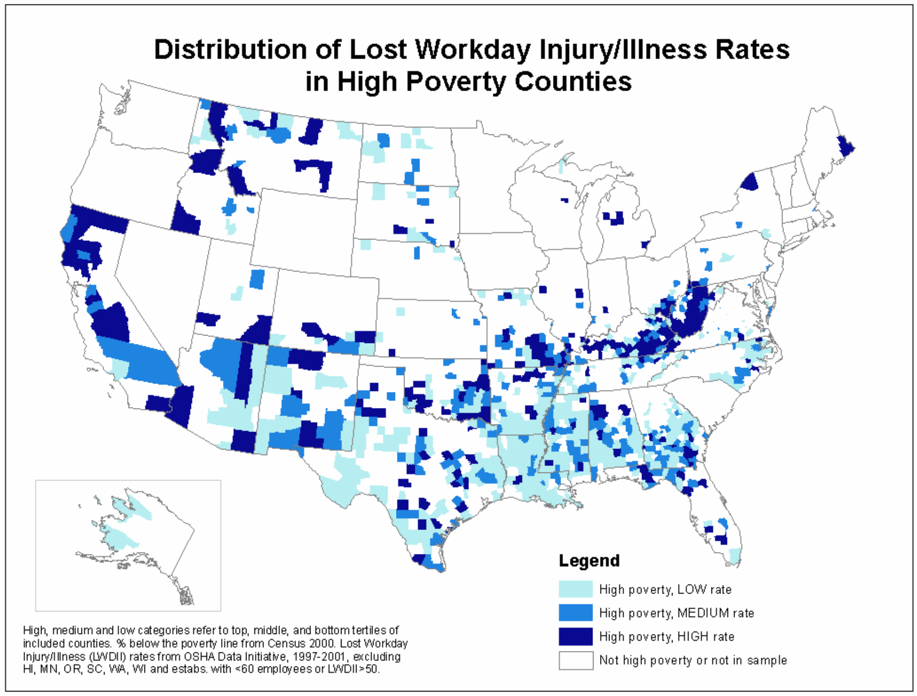
For context in viewing the following maps, Map 7 shows the basic tertile distribution of county mean injury/illness rates.

Map 7: Mean county LWDII, Tertiles



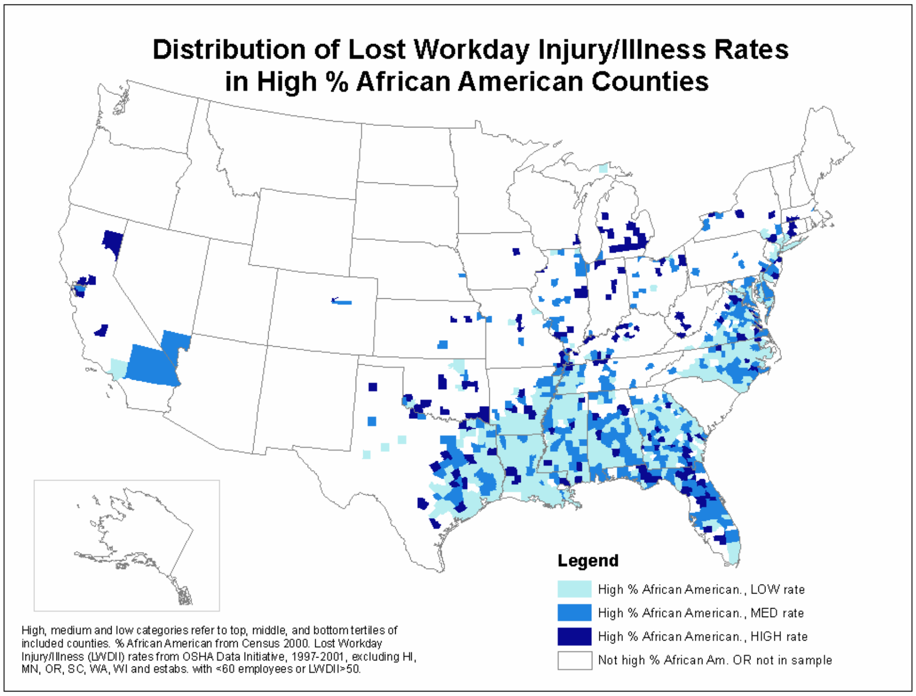
Map 8 shows the distribution of LWDII rates among the top 1/3 high poverty counties. A strong pattern of overlap between poverty and high rates is seen in the Kentucky-West Virginia area, and in general, middle and high LWDII rates coincide with poverty in the West and in Missouri, Arkansas, and Oklahoma (Ozark Mountains region). By contrast, while high poverty is seen throughout the South, there are few counties with high or even middle-level LWDII rates there. 264 counties were in both the top tertile of poverty and the top tertile of LWDII, versus 324 expected if the distribution was random.

Map 8: Distribution of LWDII in High Poverty Counties

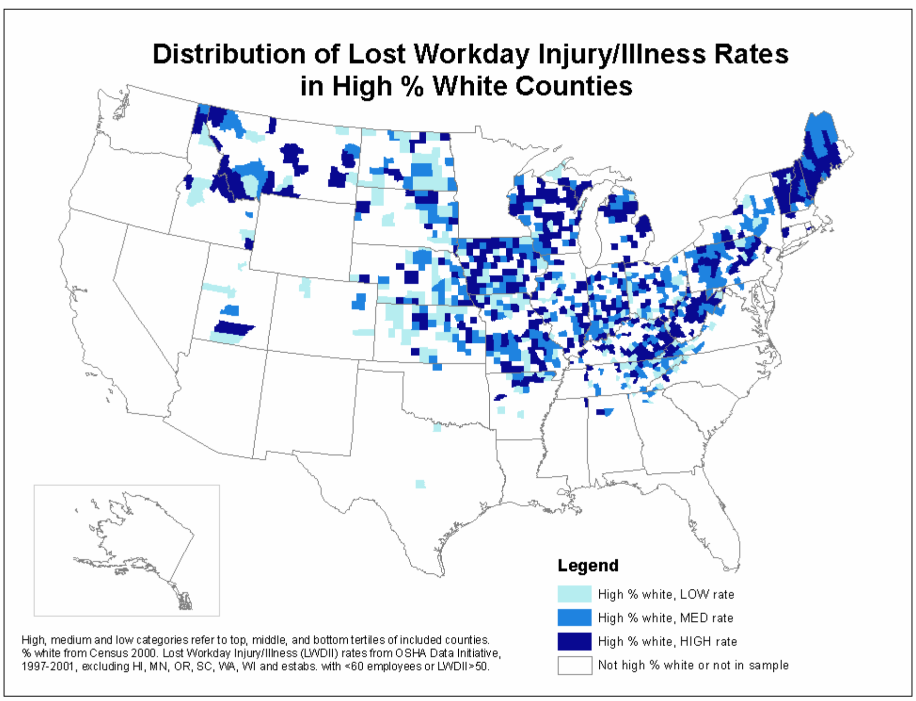


Map 9 shows the distribution of county mean LWDII among high percent African American counties. There were only 183 top tertile overlaps versus 331 expected – one of the lowest levels of congruence of any risk factor. Most of the top 1/3 % African American counties were located in the South, where the ODI found low reported injury/illness rates. The main areas where there were many African American counties with high LWDII rates were Michigan and Oklahoma. Map 10 looks at associations with % Caucasian. High percent white counties are almost exclusively in the northern half of the country. There were 429 top tertile overlaps with LWDII versus 324 expected – the highest congruence of any variable. The overlaps are widespread, with particular concentrations in New England, Wisconsin, and West-Virginia-Kentucky.

Map 9: Distribution of LWDII in High % African American Counties



Map 10: Distribution of LWDII in High % White Counties



These maps are a small selection of the risk factors examined. They reflect some of the more common general geographic patterns of association. Overall, there was a theme of maps showing more frequent association between areas of social privilege and high LWDII than the reverse.

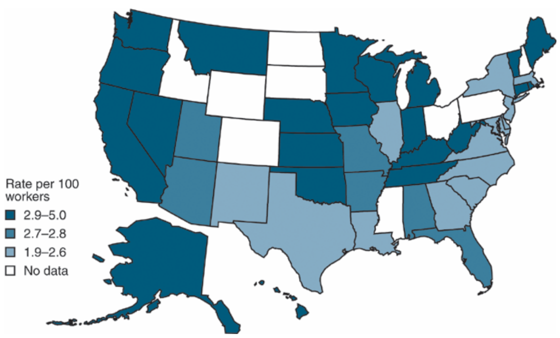
V. DISCUSSION

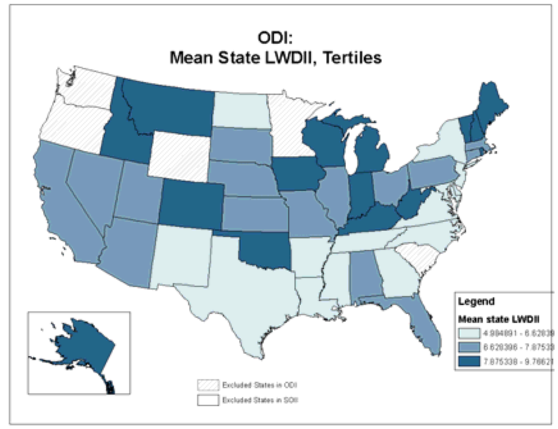
Overview of findings: These analyses demonstrate the utility of studying geographic variation in occupational injury and illness. The maps and Moran's I show that reported nonfatal occupational injury/illness rates do vary by geography in the ODI. Examinations of the Index of Area Industry Hazard describe locations of especially high-risk industry and support the idea that LWDII rates vary even after controlling for industry hazard.

The LISA cluster analysis suggested that key areas of high rates are in West Virginia-Kentucky, Michigan-Wisconsin, Northern New England, and Northern California and that the South has low reported rates. These areas were consistently highlighted in the "observed versus expected" maps and were often highlighted as having both high rates and social risk factors. The ranking analyses showed that the top five states with the highest mean LWDII rates were Vermont, West Virginia, Michigan, Maine and Kentucky. The states (and district) with the lowest reported rates were Louisiana, the District of Columbia, North Carolina, Delaware, and Georgia.

Comparison with other findings: These findings are comparable to other statewide analyses of nonfatal occupational injury/illness, such as the distribution of LWDII reported to the SOII, as mapped in NIOSH's *Chartbook* (2004, see Maps 11 and 12 below).(NIOSH, 2004) Indeed, after removing states missing in either database, only six states differ in tertile across the two maps (CA, NV, NE, KS, AR, CT).

Map 11: Bureau of Labor Statistics, Survey of Occupational Injuries and Illnesses 2001 state LWDII rates (Left)
Map 12: ODI 1997-2001 mean LWDII rates (Right, tertile version of Map 2)
(Note that white states on both maps have no data. The LWDII categories on the maps are not the same because the ODI focuses on high-LWDII industries while SOII is broad-based.)

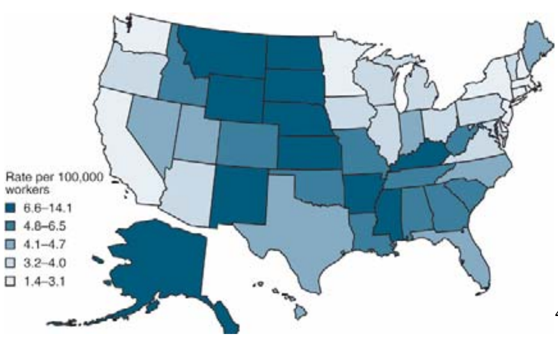




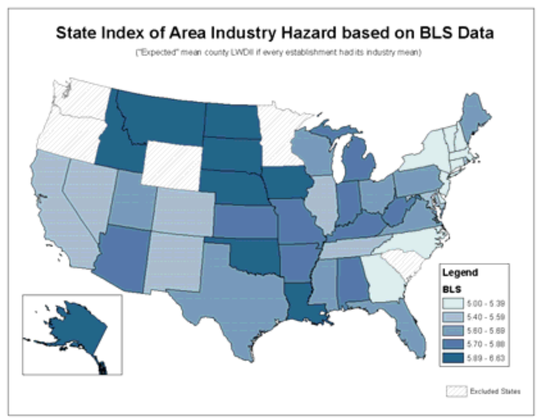
However, these nonfatal rates in both databases differ markedly from the state-level distribution of *fatal* occupational injury, where, for example, rates are high in the South and center of the country and particularly low in the Northeast and California (NIOSH, 2004). Indeed, our top-ranked state, Vermont, was in the bottom 5 *low* fatality states in 2004 (AFL-CIO, 2006). Fatal and non-fatal injuries have different drivers, including access to care (fatalities are higher where access is low), establishment size (small and self-owned have higher reported fatalities), industry type of inherent hazard, and event causes – for instance, motor vehicle crashes and violence are commonly implicated in fatalities, but play a much smaller role in non-fatal events (Herbert & Landrigan, 2000; Robinson, 1988). Further, the CFOI database of fatal injuries does not exclude establishments based on size, industry, or self-employed/contractor status, as nonfatal injury/illness databases commonly do. Nonetheless, to the extent there are risk factors common to both fatal and nonfatal events, underreporting may partly account for observed differences. It is much harder to underreport a death than a nonfatal event.

While recognizing that fatal and nonfatal events are partly “apples and oranges,” it is interesting to compare the distribution of fatal injuries with that of our Index of Area Industry Hazard. Like fatality data, the Index is expected to be less tainted by geographically-biased underreporting than the ODI. Maps 13 and 14 show that the Index and fatality distributions are highly similar. If both reflect more of an area's "true" risk, then the geographic difference between them and the ODI-reported rates may partly reflect area social risk factors and their geographically biased impact on underreporting. (The difference may also be influenced by the different databases and other factors including those discussed above.)

Map 13: Bureau of Labor Statistics 2001 state fatality rates (Left)
Map 14: Index of Area Industry Hazard [state version of Map 4] (Right)
Note that white states on Map 14 have no data.



[4]



Social Risk Factors: This project's conceptual framework suggests that economy, demographics, policy/culture/values, and industry mix are key area-level factors affecting occupational injury/illness, through their effects on business and worker incentives and capacity for safety. Maps are shown related to economy and demography.

Economy: In areas of high poverty, workers compete for scarce jobs, and thus have especially strong incentive to accept workplace hazards; this competition also reduces

[4] Chartbook Figure 1–10. Fatal occupational injury rates by State, 2002, Census of Fatal Occupational Injuries.*(Sources: BLS [2003a]; BLS [2003b].)* *http://www2a.cdc.gov/NIOSH-Chartbook/imagedetail.asp?imgid=10* (NIOSH, 2004)

business incentive to provide safety (Kahn, 1991; Robinson, 1986; Robinson, 1988). In such areas, workers may have extra life stress and reduced health status, thus potentially reducing their capacity to avoid injury/illness when exposed to risk. If businesses are also financially insecure, they may have low capacity to afford safety interventions and be more likely to cut corners (Robinson, 1988). Beyond the aggregated individual effects, low income areas may also be associated with deindustrialization. This is because remaining establishments may be older and/or hazardous, while new plants tend to be attracted to economically more vibrant areas (Loomis et al., 2004; Richardson et al., 2004; Roe et al., 2002). Evidence for an association between poverty and occupational injury/illness risk is found in several papers (Baker et al., 1992; Dembe et al., 2004; Kahn, 1991; Murray, 2003).

Even though Map 8 showed less match between high poverty and high LWDII than might be expected based on chance, the *distribution* of this association suggests that these areas are similar to some of the top hotspots identified elsewhere in this paper, especially in West Virginia-Kentucky and Northern California. These areas may reflect rural and farm poverty. This correspondence might not have been identified with quantitative examination alone, had not data been mapped.

A possible reason for the low correspondence between high rates and high poverty is the so called “compensating wage differential.” This economic theory suggests that workers are able to demand “hazard pay” in exchange for accepting hazardous jobs (Dorman, 2000; Robinson, 1986; Shrader-Frechette, 2002). By extension, we posit that an area level theory would suggest that counties with much hazardous industry would have, on average higher wages.

The hypothesis has generated much research and controversy due to the two corollaries that occupational exposures would be voluntary and thus more acceptable than other exposures; and that pay-related market incentives would be adequate to motivate employers to reduce hazards. Critiques have emphasized other factors contributing to wages including availability of alternative job opportunities and worker skills, education and physical strength. Further, workers frequently lack information about job hazards and do not always make decisions based on economic rationality (Robinson, 1986; Shrader-Frechette, 2002). Today there is some level of consensus that the wage differential exists to an extent for some groups and industries, but is not universal or complete. Shrader-Frechette cites studies suggesting that the hazard-based wage differential may only apply to white, male, college-educated and unionized workers; some studies suggest there may even be a negative differential for non-unionized workers.

We suggest that the compensating wage hypothesis may partly explain findings in the ODI sample. The ODI includes relatively large establishments, and size may be linked with relatively more available information about hazards (through formal and informal channels), and with employment of fewer of the most marginalized workers. Further, to the extent that underreporting is a concern, establishments reporting high rates could be said to be relatively transparent.

Demographics: High percent African American counties had been theorized to be at increased risk due to discrimination and to the many social challenges African Americans face in U.S. society, leading to a lopsided balance of power between workers and business (Murray, 2003; Robinson, 1988). The finding of low association between top LWDII counties and those with the highest percent African American (and the reverse) was not altogether unexpected, given the low LWDIIs observed throughout the South. Studies commonly find African Americans to be at elevated risk of occupational fatality (Loomis & Richardson, 1998; Loomis et al., 2003; Murray, 2003; Richardson et al., 2004; Stout et al., 1996), however, the literature on nonfatal occupational injury is mixed (Dembe et al., 2004; Murray, 2003; Oh & Shin, 2003; Robinson, 1989; Simpson & Severson, 2000; Smith et al., 2005; Smith et al., 2005). Oh and Shin (2003) attribute the diversity to use of differing data, methods, and model specifications, and it is also frequently suggested that the differential findings on race and fatal versus non-fatal injury may reflect the relative ease of underreporting nonfatal events. A companion paper explores quantitatively whether controlling for other risk factors including Southern location changes the direction of this association, finding that it does not (Neff, Curriero, & Burke, 2006). Accordingly, a set of possible explanations, particularly differential underreporting, are suggested. It is also possible that higher "hazard pay" wages combine with structural discrimination to move higher hazard industries with larger establishment sizes into counties that are more white.

Discussion of Methods: In comparing Map 2 (LWDII at the state level) with Map 3 (county level), the state level map is initially easier to "read" due to its reduced detail. However, much information is lost. Further detail could save money by avoiding targeting the entire state with interventions. Detail is also critical for making the case to local policymakers about needs for intervention.

If selecting one measure to report for surveillance purposes, the basic map of outcomes is recommended, preferably at the county level or other sub-state aggregation. Moran's I and LISA cluster analyses help with interpretation. The ranking analyses demonstrate that great care must be taken when reporting counties by name, since different methods produce very different ranks. Direct ranks are simplest and most comprehensible but should not be used if it is expected that the top counties have high variance in their outcomes. In those cases, the Bayesian method will be more appropriate. The direct analysis limited to counties with larger numbers of observations (such as 30) provides even more stable results and may be especially helpful for intervention targeting as it incorporates both industry density and risk. However, it can be confusing because it excludes so many counties. This mapping method provides a useful display tool for showing two variables at once and should be considered as an alternative to displaying multiple maps together.

Should the OSHA Data Initiative be used for surveillance? This paper may provide the first broad descriptive examination of the OSHA Data Initiative in the literature. The database has many limitations, including a differential incentive for underreporting, changes in sampling strategy by year, and data quality. The U.S. Government Accountability Office has questioned whether data limitations, particularly underreporting, make the ODI appropriate for OSHA's own purposes, never mind for surveillance (U.S. Government Accountability Office, 2005). At the same time, the ODI is the only publicly available database providing national establishment-level occupational injury/illness data. Further, the state-level distribution of its findings did not differ dramatically from those in the more scientifically valid SOII, even though the ODI focused on high injury/illness rate industries. Underreporting is not unique to the ODI (Azaroff et al., 2002; Azaroff et al., 2004; Conway & Svenson, 1998; Leigh & Robbins, 2004; Leigh et al., 2004; Pollack & Keimig, 1987; Pransky et al., 1999; Rosenman et al., 2006; Smith GS, 2003; Smith et al., 2005). OSHA determined that the database was adequate for a longitudinal analysis (ERG, 2004). We suggest that ongoing analyses of data in the ODI can be used to supplement other surveillance databases, so long as data are presented with appropriate caveats. At least as importantly, county-level analyses of ODI data can help *OSHA* target areas where reported rates are higher and lower than expected, to stimulate increased intervention and/or audits.

Limitations: The data limitations have been described above. The tools presented in this paper, even with supporting statistical testing, are observational and exploratory and cannot distinguish causality or directionality of relationships. They are useful for gaining insight into injury/illness distribution and for generating and exploring hypotheses, but no causal inferences may be made.

While these tools all have the benefit of making data visually appealing and interpretable, this can also be a drawback when appearances are both deceptive and convincing. Choices in map display, such as the number of categories and how breakpoints are determined, can significantly affect outcomes. Similarly, the ranking computations demonstrate that results are highly dependent on the chosen method. In interpreting the findings, the ecologic fallacy of inferring individual risk from area risk must be avoided.

Counties vary widely in size. Large Western counties have substantial visual impact compared to Eastern counties, which can imply clusters where they do not exist. The problem of differing county areas is addressed quantitatively when representing injury/illness as a rate per number of establishments in the county, but the visual distortion remains.[5] The county-level aggregation may also be too broad to detect true effects (Waller & Gotway, 2004). Counties are outlined based on administrative issues rather than representing cohesive or homogenous geographic areas, and boundaries often segment neighborhoods and industrial areas.

Strengths: This study is important in that it suggests an approach, a method, and a theory-based conceptual framework for bringing geographic analysis to the task of occupational injury/illness surveillance and ultimately to prevention activities. The various methods identified clear geographic variation in consistent geographic patterns across analyses. While the county level has drawbacks, counties are meaningful to the extent their policy, politics, and history have shaped the space across time, and they incorporate many commuting workers. Further, they provide data relevant for policy. The presented models of geographic analysis should be readily replicable using other databases. This paper may contribute the first data description of the ODI to the peer reviewed literature and suggests its utility for surveillance and intervention, as well as highlighting underreporting concerns for follow-up.

VI. CONCLUSION

Historically, the occupational injury/illness field was a leader in using geographic methods for surveillance. In the last two decades it has taken little advantage of the explosion of new tools, methods, and databases. Surveillance investments should be targeted to expand the georeferencing of occupational injury and illness databases as recommended in *Healthy People 2010* and towards addressing underreporting.

[5] Levinson and Haddon (1965) first proposed maps adjusting area to the proportion of a total population in an area as a way to visualize distributions that vary both as a function of attack rate and population density. (Levinson & Haddon, 1965) However, such maps can appear so distorted that it becomes difficult to focus on the content. Another way to handle the problem is to smooth the rates statistically, although this method makes assumptions that may not be accurate.

Budgets for occupational injury/illness prevention are miniscule and face constant political challenges. For example, the AFL-CIO reports that in 2005, only 2,117 federal and state OSHA inspectors were tasked with monitoring health and safety at the nation's 8 million workplaces (AFL-CIO, 2006). There is great need to identify efficient targeting schemes. Based on the findings in this paper, a pilot project should explore data-driven intervention targeting. Such a program could be a time and cost efficient way to reach many high-rate establishments at once, and could change area safety norms and expectations about the certainty of enforcement. Geographic targeting also offers the opportunity for local strategies with cross-industry impact, such as worker training in workplace rights and safety, management training related to safety and health programs, and media training.

If a geographic targeting scheme succeeded in bringing establishments in high rate areas closer to their industry means, it would impact not only worker health but also costs. Waehrer et al. (2004, calculated) document more than a three-fold variation in state occupational injury/illness costs per non-governmental employee. While 73 percent of this variation could be explained by industry, that means 27 percent could be explained by non-industrial factors, including, potentially, factors that can be addressed through the above-discussed interventions.

In the U.S., there are an estimated 55,000 annual occupational deaths, putting these events at the level of the country's eighth leading cause of death. Further, there were 4.26 million reported nonfatal occupational injuries/illnesses in 2004 (Steenland, et al, 2003, U.S. Dept. of Labor, Bureau of Labor Statistics, 2006). "The wrong place" is a real place where too many workers spend their days (and/or nights.)

Overall, this analysis has identified key areas of high rates in West Virginia-Kentucky, Michigan-Wisconsin, Northern New England, and Northern California, while the South had especially low rates. These findings are similar to other analyses of nonfatal occupational injury/illness distribution, but differ markedly from the distribution of fatal events. While there are many differences in drivers of fatal and non-fatal injury/illness, the fatalities distribution was actually fairly close to our “expected” nonfatal rates based on industry hazard. If fatality and industry hazard reflect more of a “true” area risk, then the geographic difference between them and the ODI-reported rates may partly reflect area social risk factors and their geographically biased impact on underreporting.

By helping to visualize the occupational injury/illness distribution, geographic tools can help in advocacy to increase the overall level of resources devoted to occupational injury/illness prevention. More directly, they can provide data to improve intervention targeting efforts, suggest risk factors for investigation, and make the case for targeting resources to prevention in hard-hit areas so that one day, “the wrong place” can be transformed into just “a place.”

4

Social Predictors of County-Level Occupational Injury and Illness

Neff RA, Curriero F, Burke TA.

ABSTRACT

BACKGROUND: Across U.S. counties, workers and businesses operate in a diverse landscape of demographics, economy, culture, policy, and industry. These local conditions in turn affect business and worker incentives and capacity for safe work. Geographic analysis has moved into the mainstream of public health and environmental tools, but few studies have examined geographic risk factors for occupational injury/illness. This paper analyzes predictors of county-level outcomes.

METHODS: Occupational injury/illness rates were drawn from the OSHA Data Initiative, 1997-2001, and social risk factors were from the U.S. Census 2000 and other sources. Data were mapped and described at the county level, and geographic and multilevel regression analytic tools were used.

RESULTS: High rates of occupational injury and illness were associated with poverty, Caucasian race, unionization, strong safety net, and industry hazard, as well as the control variables of non-Southern states and non-“rural farm” areas.

DISCUSSION: The conceptual framework categories and emphasis on geographically-varying social risk factors is supported. Findings counter to the hypothesized directions of relationships are discussed and interpreted. Underreporting may be a factor. Further research on geographic risk is needed, including replications in other injury/illness databases. Examining the broader context of occupational injury and illness points to methods for prevention through intervention targeting and design.

*   *   *   *   *

I. INTRODUCTION

Across U.S. counties, workers and businesses operate in a diverse landscape of demographics, economy, culture, policy, and industry. These local conditions play into business decisions about location and affect working conditions everywhere. Accordingly, they may affect business and worker incentives and capacity for safe work. In 2004, there were 4.26 million reported nonfatal occupational injuries and illnesses in United States private industry, often with serious repercussions for workers’ lives and finances (Bureau of Labor Statistics, 2005, Boden et al., 2001; Dembe, 2001; Dorman, 2000; Leigh & Robbins, 2004). These injuries and illnesses are not uniformly distributed geographically. A public health approach to prevention involves looking not only at proximal cause, but also stepping back to examine the context. This paper presents a multilevel regression analysis examining effects of county- and state-level risk factors on mean workplace occupational injury/illness rates by U.S. county.

A rich body of literature documents associations between geographic determinants and numerous health outcomes, providing insights into disease mechanisms, drivers of disparities, and ways to improve intervention targeting (Cromley & McLafferty, 2002; Hillemeier et al., 2003; Krieger et al., 2002; Krieger, et al., 2003; Waller & Gotway, 2004). By contrast, the literature on occupational injury and illness in geographic context is small. No identified studies were found based on broad conceptual frameworks about how geographic factors affect occupational injury and illness. There are only a few analytic studies, and fewer still that use methods to account for spatial features of the data (Neff, 2006).

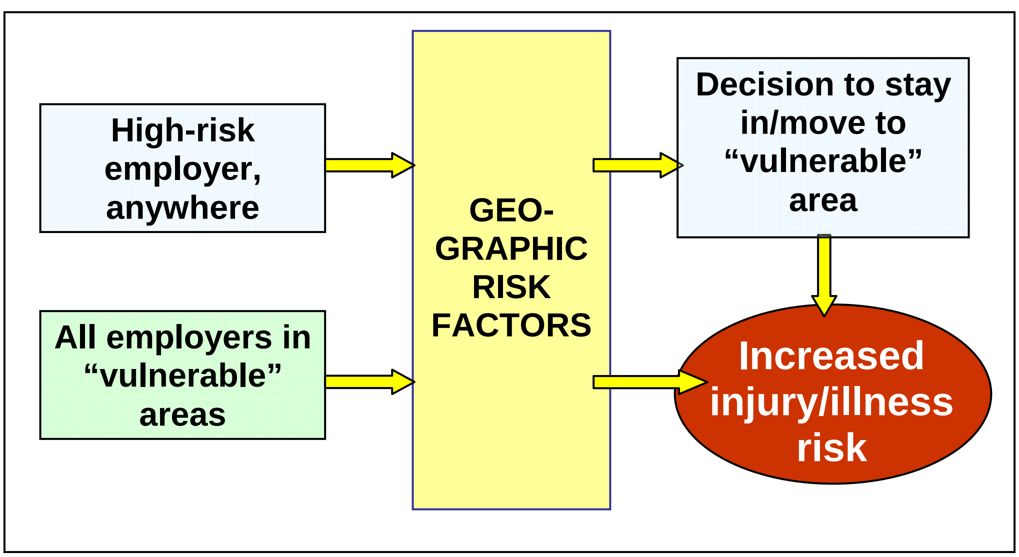
What does it mean to take account of those features? The key insight relevant to multivariate analysis is that risk factors often do not respect geographic boundaries, making nearby areas more similar to each other than to those further away. This phenomenon is known as positive spatial autocorrelation, and it violates the regression assumption of independence. If areas are not independent but are analyzed as if they are, the uncertainty of regression effects could be underestimated, leading to findings that overstate statistical significance. Geographic tools of regression include assessing whether autocorrelation exists, and if it does, adding an appropriate geographic weighting factor or random effect, typically using geographic software or multilevel regression methods. Further, it is useful to assess whether residual spatial variation remains after fitting a model, by mapping residuals and/or constructing a semivariogram – a graph reflecting the changing difference between residuals with increasing geographic distance.(Cressie, 1991; Anselin & Koschinsky, 2006)

II. BACKGROUND

Geographic risk factors are taken here to represent those measured at the area level (such as the county or state) rather than the individual or workplace level. They reflect both aggregated individual effects and effects derived from the overall conditions in an area. Geographic risk factors for occupational injury/illness include the economy, demographics, worker-friendliness of policy, values, and culture,[6] and industry mix.

As shown in Figure 4, we theorize that local conditions play into business decisions about location, making risky and "bad actor" employers more likely to locate in "vulnerable" areas where they are better able to get away with relatively low levels of worker protection. At the same time, geographic risk factors affect *all* employers wherever they are, often with implications for occupational injury/illness rates.

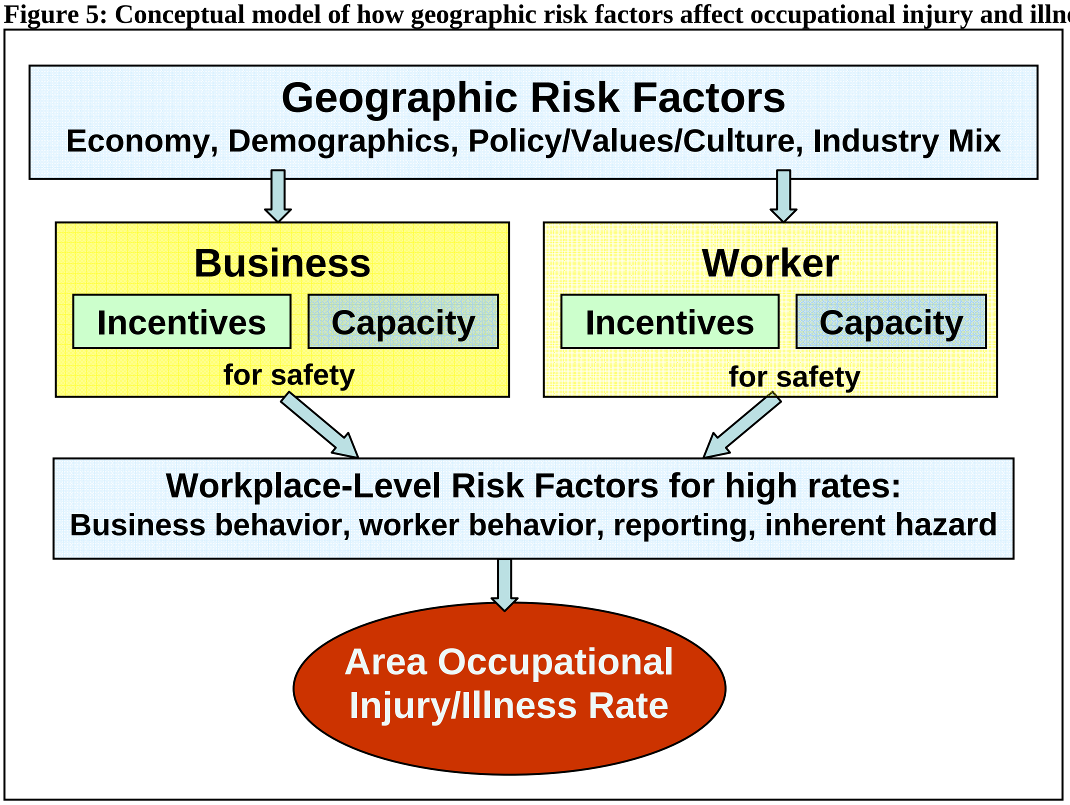
Figure 4: How geographic risk factors affect high risk employers and "all" employers.



[6] Policy, values and culture are grouped together because policy reflects the other two, representing a more formal way for them to affect health.

Figure 5 presents a conceptual model for how geographic risk factors affect occupational injury/illness rates. It shows that they drive business and worker incentives and capacity to avoid risk (by a variety of mechanisms including their impact on the balance of power between business and workers) (Robinson, 1988). For example, area poverty may make workers more desperate to get and keep jobs, so they have incentives to accept risky work and employers have less incentive to protect them. Area poverty may also mean businesses are operating at the margins and have low capacity for prevention, and that more workers suffer reduced physical or emotional health, limiting their capacity for prevention of or recovery from injury/illness. Business and worker incentives and capacity in turn drive workplace-level behavior, which is the proximal cause of injury/illness.

Figure 5: Conceptual model of how geographic risk factors affect occupational injury and illness



Workplace level "behaviors" include a wide range of actions directly and indirectly related to occupational injury and illness, including industrial hygiene controls, safety and health programs, management commitment to safety, safety attitudes and carefulness, decisions about establishment location and what the establishment will produce or do, decisions about reporting – and also behaviors that unintentionally affect injuries/illnesses, including those related to work organization or job stress.

Motivated firms can change their injury/illness rates; Shannon and Vidmar estimate that 42 percent of lost worktime injuries could be cut if all Ontario businesses matched the safety performance of the top 25th percentile for their industries (Shannon & Vidmar, 2004). At the same time, different establishment injury/illness rates may reflect different levels of inherent hazard, since industry categories are broad. Establishment rates may also be tarnished by underreporting. Leigh et al (2004) and Rosenman (2006) estimate that nationally about 2/3 of occupational injuries may be missed in surveillance systems. Azaroff et al (2004) documented powerful economic, demographic, political and social factors in occupational injury/illness underreporting, which may contribute to geographically differential underreporting. The database for our analysis is used in enforcement, increasing the differential incentive for high rate employers to underreport.

Based on the conceptual framework, we test the hypotheses that the variation in county occupational injury/illness rates will be associated with counties having:

1. weak local economies
2. minority demographics
3. policy/values/culture measures suggestive of an anti-worker (or pro-business) bias
4. high industry hazard.

The decision was made to conduct regression analysis at the county and state level rather than more locally, because the area-level exposures had most meaning (and were measurable) at these broader levels. Economic indicators are rarely reported below the county level, the studied policies are enacted at the state level, and workers may commute from a broad surrounding area. In the median U.S. county, 67.4 percent of workers live in the county where they work (U.S. Bureau of the Census, 2000).

III. METHODS:

A. DATA

1) Injury/Illness Database: The injury/illness database is the Occupational Safety and Health Administration (OSHA) Data Initiative (ODI), an annual national survey of establishment injury/illness rates in high rate industries, conducted to improve OSHA's enforcement targeting.(OSHA Directorate of Compliance Programs, 1997, 1998, 1999, 2000, 2001). Following Freedom of Information Act requests, most ODI data are compiled on the Internet (OSHA / The Memory Hole, 2005). Table 15 describes the database by year.

Table 15: OSHA Data Initiative database, 1997-2001: Establishment-level lost workday injury/illness (LWDII) rates in most high-rate industries

YEAR	N (after exclusions*)	LWDII mean (SD)	# not Geocoded (%)	Response rate	National Average LWDII**
1997	44,199	7.41 (6.68)	544 (1.2)	NA	3.3
1998	36,043	8.14 (6.77)	332 (0.9)	91%	3.1
1999	42,270	7.66 (6.45)	369 (0.9)	95%	3.0
2000	45,026	7.30 (6.21)	232 (0.5)	96%	3.0
2001	49,308	5.93 (5.65)	434 (0.9)	94%	2.8
TOTAL	216,846	7.22 (6.38)	1,941 (0.9)	NA	

* Exclusions: HI, MN, OR, SC, WA, WY; establishments with <60 employees; establishments reporting LWDII>50
** National average from Bureau of Labor Statistics, covering all private industry

From 1997-2001, OSHA annually surveyed roughly 80,000 establishments with over 40-60 employees in industries with the highest injury/illness rates in the Bureau of Labor Statistics' Survey of Occupational Injuries and Illnesses, plus all manufacturing industries. The sample excludes industries not regulated by OSHA, including mining and government workers, and the construction industry. Reflecting number of establishments among these high-hazard industries, 15.31% of the sample is comprised of nursing homes (SIC 8051, 8059, 8052). Other top industries are Plastic Products, Not Otherwise Classified (SIC 3089, 2.85%); Long Distance Trucking (SIC 4213, 2.73%); and Lumber and Other Building Materials Dealers (SIC5211, 1.89%). See Appendices 1 and 2 for top industries by number surveyed and by injury/illness rates. The ODI is structured so that each eligible establishment should be surveyed at least once every three years; however, due to large numbers, nursing homes and department stores are sampled less.

For this study OSHA provided a 1997-2001 database of responding facilities, including industry Standard Industrial Classifications (SICs), size (number of employees, grouped into categories), and establishment Lost Workday Injury/Illness Rates (LWDII). LWDII is the number of lost workday injuries/illnesses per (the equivalent of) 100 full time workers (excluding contractors). Compared to injuries, underreporting of illnesses is expected to be particularly substantial due to latency, challenges to establishing work-relatedness, and other barriers (Azaroff et al., 2002).

We combined five years of data for analyses based on OSHA recommendation (J. Dubois, personal communication, February 2006) and the benefits of increased sample size. We acknowledge that some establishments were counted more than once, due to lack of unique identifiers. A further limitation is that the sampling strategy varied across years, and sensitivity analyses did show possible differences. To improve sample consistency, we limited it to establishments with at least 60 employees and excluded six states due to low numbers: Hawaii, Minnesota, Oregon, South Carolina, Washington, and Wyoming. (Participation in the ODI is optional each year for states that run their own OSHA programs.)

A remaining concern is that in some years OSHA excluded establishments reporting LWDII under 7.0 in the prior two surveys, so in areas with many such establishments, there would be differential misclassification bias and/or different numbers of establishments sampled by year. Appendix 3 presents maps of LWDII by year, and Appendix 4 describes sampling differences by year and considers the impact of this issue. Finally, OSHA recommended excluding the 0.1 percent of establishments reporting LWDII above 50 because of the likelihood of error (Dubois, J., Personal communication, February 2006).

Establishment locations were geocoded (linked to points on the map) based on zip code and then assigned to the counties in which their zip code centroids fell. Establishments that could not be geocoded were excluded from the sample. Appendix 5 compares geocoded and non-geocoded establishments. A year- and county-aggregated file was created and merged with the below-described risk factor database. The outcome variable, "mean LWDII," was calculated to reflect the mean reported LWDII of all responding establishments in a county.

2) Risk Factor Database:

Following a search for variables reflecting conceptual framework categories, a database of 90 county[C]- and state[S]-level risk factors was compiled. Data came from multiple sources, primarily the U.S. Census 2000, Census Counties 1998, and Economic Census 1997.(Appendix 10) Table 16 presents risk factors selected for the final model based on statistical significance. The demographic measure reflects racial composition: percent African American [C]. One measure of local economy/ socioeconomic status (SES) is included: percent of population living below the poverty line[C]. Two state-level measures of culture and policy relevant to workers were in the model: percent unionized 1999[S], and rank in the ratio of unemployment benefits to average weekly wage [S] (low rank reflects strong safety net.) Control variables were: percent rural, farm[C]; and whether a state was in the South[S].

Because of the importance of industry in determining hazard levels, an “Index of Area Industry Hazard” was developed to reflect the rates that would be expected based on county industry mix alone. Specifically, the Index measures the “expected” mean county (or state) LWDII if every facility had its industry mean (Goldman, L. Personal communication, November 2004; Smith G. Personal communication, November 2004). See (Neff, Burke, & Curriero, 2006) for further details.

Table 16: Risk factors used in regression analysis and distribution among counties or states with observations in data set

(N=2,657 counties of 3,141 total; and 45 of 50 states plus the District of Columbia).

RISK FACTOR	COUNTY MEAN (RANGE)	THEORIZED RELATIONSHIP TO INJURY/ILLNESS
Demographics		
% African American [C]	9.29% (0, 86.1)	Related to workplace and institutional discrimination, possible low worker knowledge of rights, and possible low political empowerment.
Local Economy/ Socioeconomic Status		
% Less than poverty [C]	14.07 % (2.1, 50.1)	Related to economic insecurity and willingness to accept hazardous work, social stress and related vulnerability; pre-existing medical conditions, social infrastructure, area political empowerment, and worker knowledge of rights.
Policy, Culture		
% unionized 1999 [S]	11.46 (3.2, 25.3)	Initially included as a measure of worker power. Can also reflect industry hazard and deindustrializing area.
Rank: ratio of unemployment benefits to average weekly wage [S]	26.70 (2, 51)	Measures the strength of the safety net, and may relate to how workers balance the need to preserve job security against other pressing needs like safety.
Industry hazard		
Index of Area Industry hazard: “Expected” county lost workday injury/illness rate if every ODI establishment had its industry mean [C]	5.94 (0.7, 12.65)	Perhaps the top determinant of establishment hazard is its industry.
Control		
% Rural, Farm [C]	4.21 % (0, 31.7)	Urbanicity risk factors include declining jobs and resultant reduced worker bargaining power.
South [S]	35.7% South; 64.7% non-South	Used in model due to visible difference in LWDII between South and other states. May reflect racial composition, culture, attitudes towards unions, political conservatism, or other factors.

B) DATA ANALYSES

1. Exploratory Analyses: The data were explored by examining the distribution of each risk factor and correlations between risk factors (Appendix 11.) To look at data geographically, risk factors were mapped individually and the outcome (Neff, Burke, & Curriero, 2006). Bivariate regressions were performed to examine the direct effect of each risk factor on the outcome (Appendix 11).

2. Statistical Modeling:

The primary outcome of interest in our analysis was injury/illness rates at the county level, but some state level risk factors were of strong theoretical interest. This motivated a multilevel modeling approach that could properly incorporate both state and county level risk factors. This type of model accounts for the fact that all counties within a state have identical values for the state level risk factors, and for the fact that the landscape of injury/illness risk factors may vary more gradually than can be explained at the county level. Accordingly, there is less opportunity for variation than the linear regression model assumes, so coefficient statistical significance needs to be reduced.

A multilevel regression model was built in several steps. First, a multivariate linear regression model was built at the county level, incorporating county-level risk factors based on theory and statistical significance ($p<0.05$). State-level risk factors were then added as a second step so that the final model form included both state and county fixed effects, the county error term, and an additional state level error term to account for possible residual state level variation (Snijders & Bosker, 1999). All error terms were assumed Gaussian independent with constant variance. Interactions between risk factors were tested.

The final multilevel model took the form:

$$\textit{Mean county LWDII} = \textit{county-level risk factors} + \textit{Error}_{county} + \textit{state-level risk factors} + \textit{Error}_{state}$$

Model selection was guided by a combination of: a) comparing deviance across models (see below); b) seeking to minimize the model condition number, which became quite elevated in many models (see below); c) statistical significance of risk factors; d) theoretical significance.

The deviance (-2 * log likelihood) reflects the level of lack of fit between the model and data, so models with lower deviances (higher log likelihoods) are preferable. Nested models are compared by subtracting the deviance of one from that of the other and performing a chi-squared test on the result. Statistically, this difference in model deviance can be interpreted as relating to the significance of factor effects for nested models (Snijders & Bosker, 1999).

The condition number is related to the level of precision when solving linear matrix systems of equations for iteratively estimating parameters, as is done here. It represents the square root of the ratio between the Hessian matrix's top and bottom eigenvalues. Very high condition numbers could be due to high correlation with other risk factors, lack of variation in the risk factors themselves, or close relationships with the outcome.

3. Model Diagnostics: It is possible that after accounting for county and state level risk factors, there could be remaining spatial dependence at either geographic level. Spatial dependence describes the situation where data closer together – here, neighboring counties and/or states – are more similar than those further away. Although spatial dependence may be interpreted as providing clues to missing covariates that themselves vary geographically, it is a clear violation in traditional regression analysis. Final regression models were examined for residual spatial variation in two ways.

First, county and state level residuals (the difference between actual rates and those predicted by the model) were standardized and mapped. If there was little spatial dependence, there would be no obvious patterning of high and low values. Second, a semivariogram plot (Cressie, 1991) was computed. In this plot, pairs of counties at varying geographic distances are compared, and the average difference between their squared residuals plotted. The plot's x-axis shows distances between pairs of county centroids and the y-axis shows the semivariogram at that distance. If the plot suggests an increasing pattern with distance, that suggests that nearby pairs of counties are more similar than those further away and thus that spatial dependence remains.

Model diagnostics were also performed to assess nonspatial issues such as multicollinearity.

4. Sensitivity Analyses: Establishment-level linear regression analysis was performed by year to evaluate the effects of repeated sampling of establishments, changes in sampling, and changes in economic or other factors across time. These analyses represent replications and enable us to assess stability of the regression models selected. Establishment-level regressions were also performed using all years combined, including controlling for establishment size and comparing results in the top ¼ of high rate industries with those in other industries. An alternate outcome measure was tested: "average county difference between observed and expected mean LWDII" (with expected mean LWDII defined as the Index of Area Industry Hazard, described above.) Finally, regression models were run separately for Southern and non-Southern states.

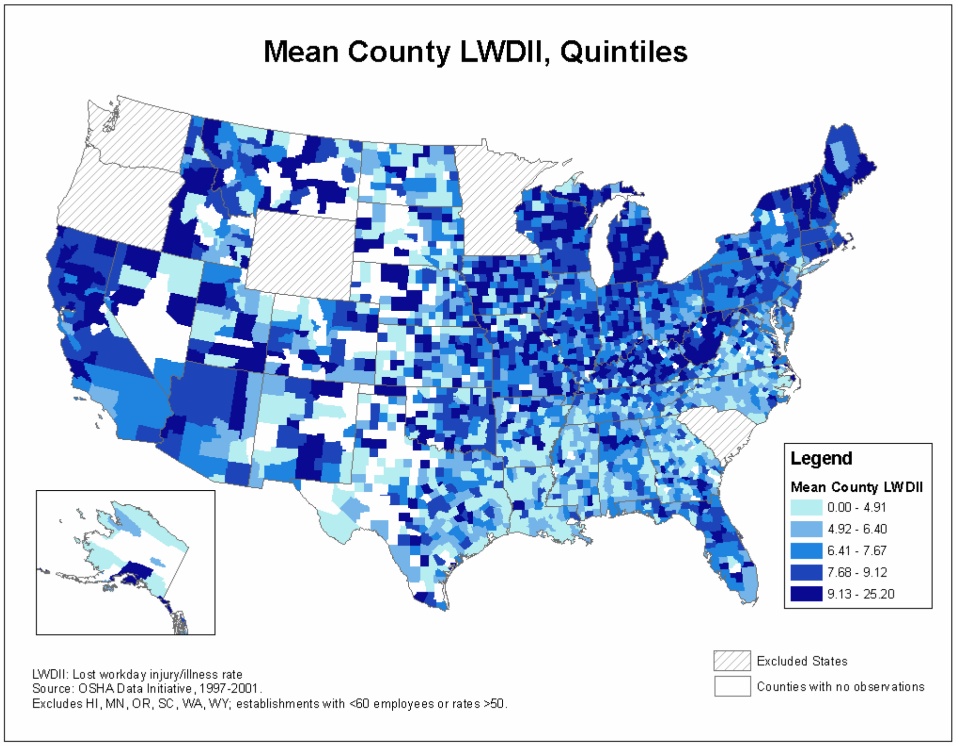
5. CART Analyses: Classification and regression tree (CART) modeling was used to seek out the strongest patterns in the data, to assure that important relationships or interactions were not overlooked. CART models select key risk factors from among those presented, to create a "tree" pattern that splits the data at points where the two branches diverge most (Breiman, 2001; Venables & Ripley, 2002).

6. Computing: Computing tools used in this analysis were: Stata for calculations and exploratory analyses; Stata's Generalized Linear Latent and Mixed Models (GLLAMM) package for multilevel models (Ribeiro & Diggle, 2001; Ribeiro & Diggle, 2001); ArcGIS version 9.0 for mapping (ESRI, 2005); R and GeoR for some computations; code by Curriero, C (Appendix 12) for computing the semivariogram; and R's Tree package for the CART model (R Development Core Team, 2005; Ribeiro & Diggle, 2001; Ripley, 2006). The CART code was adapted from a script by Ingo Ruczinski.

IV. RESULTS

1. Exploratory Analyses: 2,657 of 3,141 U.S. counties were included in the sample. The mean county LWDII was 7.22 (SD: 6.38). Map 15 shows the geographic distribution of the outcome. Rates were relatively low in the South except Florida and especially high in West Virginia and other parts of Appalachia, the upper Midwest, Northern California, and New England. Due to industry distribution, sampling is somewhat sparse in the middle of the country.

Map 15: Mean county LWDII, Quintiles



2. Statistical Modeling: Table 17 presents results from the bivariate, county-level and multilevel regression models. The final model was:

LWDII = 2.83 – 0.03(% African American[C]) + 0.02(% below poverty[C]) + 0.08(% union[S]) - 0.02(rank: ratio of unemployment benefits to average weekly wage[S]) + 0.79(Index of Area Industry Hazard[C]) -0.11(% rural-farm[C]) - 0.96(South Y/N[S]) + county random effect + state random effect.

[C and S identify county- versus state-level variables.]

The log likelihood for the final model was -6218.39. The model condition number was 338.78, showing that the model was reasonably well identified. This model represented a compromise with models including additional risk factors that had theoretical interest but caused the condition number to become very high or the log likelihood to become much more negative. The coefficients of the final model variables changed little between these models. Appendix 11 lists all risk factors discussed below in the results section, including means, ranges, coefficients from bivariate regressions with county mean LWDII, and data sources.

Table 17: Effect of state and county risk factors on county mean LWDII, 1997-2001: bivariate, county-level and multilevel models

Risk factor	Bivariate (95% CI)	County-level linear regression model (95% CI)	Multilevel model: county and state-level fixed effects; county and state error terms (95% CI)
Demographics			
% African American [C]	-0.055 (-0.062, -0.047)	-0.033 (-0.042, -0.024)	-0.031 (-0.041, -0.021)
Local Economy/SES			
% below Poverty [C]	-0.048 (-0.066, -0.031)	0.035 (0.018, 0.053)	0.024 (0.005, 0.044)
Policy, Culture, Values			
% unionized 1999 [S]	0.16 (0.14, 0.18)	0.09 (0.06, 0.12)	0.08 (0.02, 0.15)
Rank: ratio of unemployment benefits to avg. weekly wage [S]	-0.02 (-0.03, -0.02)	-0.02 (-0.03, -0.01)	-0.02 (-0.04, -0.002)
Industry hazard			
Industry hazard: “Expected” county lost workday injury/illness rate if every ODI establishment had its industry mean [C]	0.75 (0.65, 0.85)	0.76 (0.66, 0.85)	0.79 (0.71, 0.89)
Control			
% Rural, Farm [C]	-0.017 (-0.041, 0.006)	-0.121 (-0.144, -0.098)	-0.107 (-0.134, -0.081)
South [S]	-2.06 (-2.28, -1.84)	-1.03 (-1.36, -0.69)	-0.96 (-1.77, -0.16)
Constant		2.82 (2.08, 3.56)	2.83 (1.73, 3.93)
Model evaluation	*Varied*	*R^2 = 0.23*	*Log likelihood=* -6218.39. Condition number=338.78
Number observations	*Varied*	*2,657*	*2,657 counties; 45 states*

C: County; S: State

Demographics: For every point increase in county percent African American, there was a 0.03 point drop in LWDII (95% confidence interval (CI): -0.04, -0.02), controlling for other risk factors. This finding of a negative association held in many models, even after controlling for whether a state was in the South. Other results add a further dimension to the study, even though they could not be included in the final model. The segregation index of dissimilarity [C] and four measures related to immigrant status [C] all were negatively associated with high rates after controlling for other factors. They were excluded from the model due to their negative impact on the model condition number (related to the model's precision). The segregation index of exposure [C] and Gini coefficient of economic inequality [C] were also negatively associated with high rates but were nonsignificant in the final model. Including the percent of county population in various age groups did not substantially change other coefficients nor the percent of variation explained; several age categories were significant in bivariate but not multilevel models.

Local Economy and Socioeconomic Status: By most measures, high rate counties had reduced economic status. For every point increase in the percent of county residents living below the poverty line, there was a 0.02 point increase in LWDII (CI: 0.005, 0.04) after controlling for other risk factors. This finding varied somewhat across models. In bivariate analysis, poverty was negatively associated with rates. The association was found to be confounded with race, unionization and Southern status; when all three of these other variables were in a model, the poverty coefficient would be positive; otherwise, it would be negative.

Although they could not be included in the final model, per capita income [C] (nonsignificant), and education [C] (effects on condition number) also showed evidence in many models that counties with weak local economies and low education had higher injury/illness rates. By contrast, and also excluded from the final model (condition number), a long term *increase* in employment rates from 1975-1996 [C] was shown repeatedly to predict high injury/illness rates. The unemployment rate [C] itself showed mixed relationships with the outcome in different models.

Policy and Culture: For every additional state percent unionized, there was a 0.08 point rise in LWDII (CI: 0.02, 0.15). The other measure of state “worker-friendliness” – quality of the unemployment compensation safety net – showed increasing injury/illness with improving benefits. Small, nonsignificant reductions in rates were found both for measures of county-level OSHA enforcement, and for states that ran their own OSHA programs (versus those using federal OSHA.) County-level measures of environmental emissions were not significantly associated with the outcome, although there was a low number of included counties.

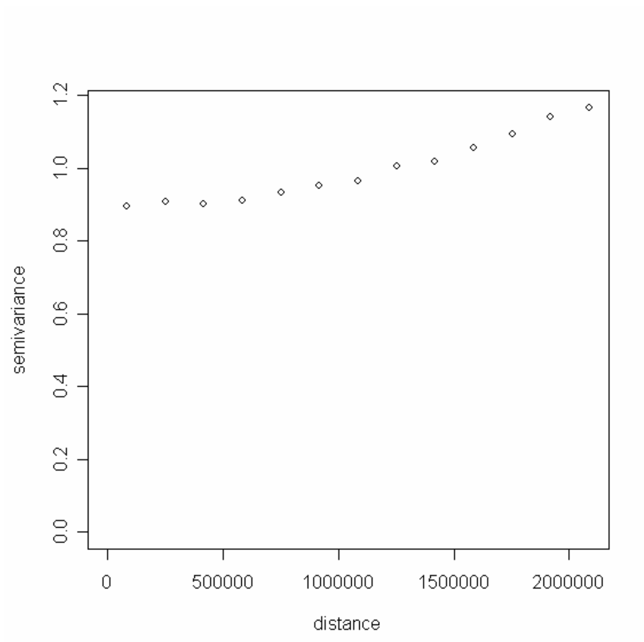
Industry Hazard: For every point increase in county “expected” industry hazard, there was a 0.79 increase in the observed LWDII. (CI: 0.71, 0.89).

Control: Injuries and illnesses were negatively associated with county rural status, especially “rural-farm” areas (coef= -0.11 (-0.13, -0.08)). Southern states were associated with a 0.96 point drop in LWDII versus non-Southern states, controlling for race and other risk factors (CI: -1.77, -0.16). Smaller establishment size (among this sample of establishments with over 60 employees) [C] was associated with higher rates but was not included in the final model due to the effect on log likelihood. Finally, counties with greater numbers of ODI establishments were found to have higher rates although the variable dropped out of the final model due to statistical significance.

No interaction terms between model variables were kept in the final model, due to lack of statistical significance.

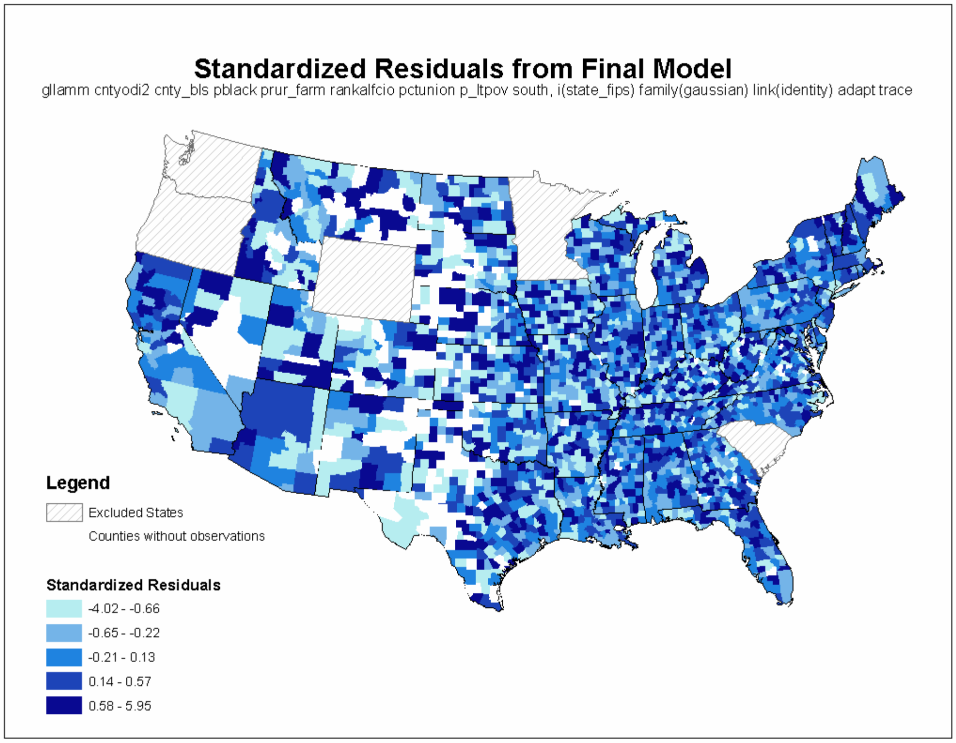
3. Model Diagnostics: A semivariogram was computed for standardized residuals (Figure 6), excluding Alaska due to its distance from the mainland.

Figure 6: Semivariogram of standardized residuals from final GLLAMM model showing little remaining spatial dependence



The plot shows little if any remaining spatial dependence. Even at close distances, this semivariogram shows a large difference between values of residual pairs (the left side of the plot does not come down toward the origin), and as the distance increases up to half the distance in the map, the difference between pairs of residuals barely increases. Map 16 depicts standardized residuals geographically, reflecting close to a checkerboard pattern with few areas of high or low values that would suggest remaining spatial dependence that had not been captured in the model.

Map 16: Standardized residuals from final model



4. Sensitivity Analyses (Appendices 13-15): For the linear regressions by year and at the establishment level, results were generally consistent with the final model in the magnitude and direction of coefficients. The main difference was that, as in the bivariate analysis, *low* poverty was generally associated with high rates in establishment-level analyses and often did not reach statistical significance. Establishment size also did not reach statistical significance. Results from multilevel regression using the outcome, "average county difference between observed and expected mean LWDII," were highly similar to those with the county mean LWDII outcome, so the original was reported due to its comprehensibility.

A sensitivity analysis was performed comparing counties in Southern vs. non-Southern states (Appendix 15). It found that county % African American was much more negatively associated with injury/illness rates in non-South than Southern counties, although the association was significant and negative in both. In non-South counties, there was a strong, significant relationship between high rates and county percent below the poverty line, whereas in the South the relationship was negative and nonsignificant. Unionization had a slightly stronger association with high rates in the South than elsewhere. Finally, in non-Southern areas, “percent rural-farm” had a strong negative association with high rates, whereas in the South, this was mitigated.

A second sensitivity analysis (Appendix 13) compared establishment-level findings in the top ¼ of highest-rate industries with those in the bottom ¾ of industries by rate. High rate industries seemed to drive the regression findings for all variables but the Index of Area Industry Hazard. For example, in high rate industries, % African American was much more negatively associated with high rates than in other industries. Percent below the poverty line was positively associated with the outcome in high rate industries, but had a negative, nonsignificant relationship in other industries. For unionization, rural-farm, and South outcomes, the top ¼ high rate industries also had a closer relationship to the full sample than did the bottom ¾ .

5. CART Models:

Because the analysis results were counter to the directions of initial hypotheses, especially regarding demographics and culture/policy, the CART tool was utilized to suggest strong patterns in the data that might have been missed by the regression models. CART models are highly sensitive to the risk factors used and the order in which they are entered, so no definitive model is presented here. Based on multiple models, a list of especially influential risk factors was developed, and insight was gleaned about their relationships. This list suggested new risk factors to test. “South” was one of the strongest predictors identified in the CART analyses and CART showed that in both Southern and non-Southern states, counties with high percent African-American had relatively low mean LWDIIs (not shown). Figure 7 and associated Table 18 present a sample tree using the variables from the final regression model.

Figure 7: CART regression tree based on the risk factors in the final regression model
Numbers at bottom of "branches" represent mean LWDII among included counties. See Table below for variable names.

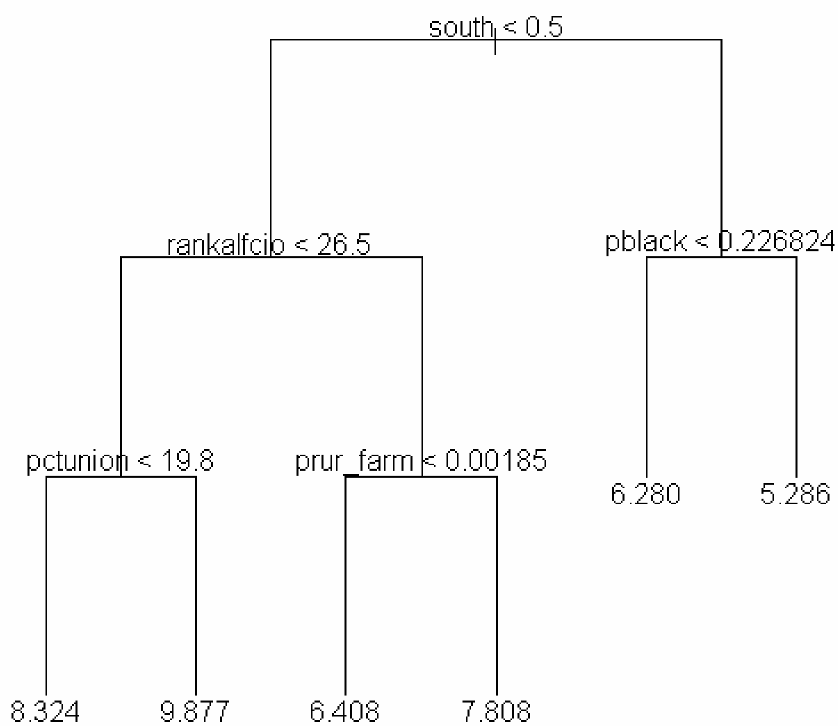


Table 18: Tabular description of CART regression tree based on risk factors in final model. ***Each branch is split at a cutpoint into two smaller branches, one containing all observations below the cutpoint, and the other all the observations above it.***

Risk factor	Cutpoint	Mean LWDII in branch	# of counties in branch
Tree "root" – all counties		7.24	1,962
South	Non-South	8.09	1,206
State rank: Ratio of unemployment benefits to average wage ("rankalfcio"; low rank shows stronger safety net)	<26.5	8.49	674
State % union (pctunion)	<19.9%	8.32	604
State % union	>19.9%	9.88	70
State rank: Ratio of state unemployment benefits to average wage	>26.5	7.59	532
% rural farm (prur_farm)	< 0.19%	6.41	83
% rural farm	> 0.19%	7.81	449
South	South	5.90	756
% African American (pblack)	<23%	6.28	463
% African American	>23%	5.29	293

V. DISCUSSION

Overall, this analysis found that high rates of occupational injury and illness were associated with poverty, white race, unionization, strong safety nets, and industry hazard, as well as location. The findings support the conceptual framework that risk factors in the domains of economy, demographics, culture, policy, and industry hazard are related to occupational injury/illness. The directions of some of the findings run counter to the hypotheses, but they were consistent, and sensitivity analyses and the CART analysis support the analytic choices. As will be discussed, differential underreporting is a real possibility, especially given the ODI's enforcement usage.

Comments on Findings:

Demographics: Our conceptual framework suggests that high rates would occur in high percent African American counties, due to workplace and structural discrimination resulting in workers having less power in such counties compared to business (Murray, 2003; Robinson, 1988). Yet perhaps the strongest finding in the analysis was the *negative* association between LWDII and both percent African American and Southern states. Is there support for this finding in the literature?

Many studies have shown increased risk of *fatal* occupational injury for African Americans (Leigh et al., 1997; Loomis & Richardson, 1998; Loomis et al., 2003; Murray, 2003; Richardson et al., 2004; Stout et al., 1996), although at least one study found equal risk (McGwin et al., 2002). By contrast, the literature on race and *non-fatal* occupational injury/illness is more consistent with our findings, with mixed results often including nonsignificant or protective effects (Dembe et al., 2004; Murray, 2003; Oh & Shin, 2003; Robinson, 1989; Simpson & Severson, 2000; Smith et al., 2005). Looking within-industry in the ODI, our meat processing case study also found increased risk in more white counties, even after controlling for type of meat (Neff, Curriero and Burke, 2006). Oh and Shin (2003) attribute the diversity to use of differing data, methods, and model specifications. It is frequently suggested that the differential findings on race and fatal versus non-fatal injury may reflect the relative ease of underreporting nonfatal compared to fatal injuries. (Other reasons for the differences include the fact that fatal injury databases cover more establishments, differences in inherent hazard and type of event between fatal and nonfatal events, and access to health care.)

Below, we discuss six possible explanations for the observed inverse association between high rates and African American counties.

1) Underreporting. The ODI is collected for use in enforcement, providing a substantial incentive for underreporting. Our conceptual framework provides reason to suspect underreporting would be stronger in areas with more discrimination, as businesses would take more advantage of workers, and workers would feel more insecure about their jobs. The U.S. Government Accountability Office (2005) has raised concerns about underreporting in the ODI, and numerous studies have documented underreporting in occupational injury/illness data generally (Azaroff et al., 2002, 2004; Conway & Svenson, 1998; Leigh & Robbins, 2004; Leigh et al., 2004; Pollack & Keimig, 1987; Pransky et al., 1999; Rosenman et al., 2006; Smith GS, 2003; Smith et al., 2005). Our meat processing study also finds indication of underreporting in the ODI (Neff, Curriero & Burke, 2006). However, it is an open question whether the magnitude of underreporting could be high enough to yield the observed findings. Further, controlling for race and Southern status does little to change the coefficient for the association between rates reported to OSHA and those reported to the non-enforcing Bureau of Labor Statistics – which we suggest is an indicator of underreporting.

2) Sampling: establishment size and industry: It is possible that the ODI sample excludes many of the high risk establishments in high-African American counties due to the size criterion of 60 employees. (Davis, 2000). It is also possible that variations in county industry account for the difference, with more hazardous industries located in whiter counties. This could occur because, as theorized in the "compensating wage hypothesis," across the industries with larger size establishments such as those in the

ODI sample, high hazard work may pay more (Shrader-Frechette, 2002). Jobs that pay better may be more likely to be in areas accessible to white workers than to African Americans due to structural discrimination. Anderton (2000) found a strong association between metropolitan segregation of African Americans from hazardous industry and from whites (Anderton & Egan, 2000). This explanation would seem to contradict extensive evidence that African American workers work in more hazardous industries than whites (Murray, 2003). There might be no contradiction, however, if the data on which the evidence is based includes small establishments whereas our sample excludes establishments with fewer than 60 employees. The Index of Area Industry Hazard *should* account for industry distribution in the regression, but it is possible that it does not fully.

3) <u>Wrong aggregation:</u> While this study's county level aggregation was chosen to capture the regions from which workers commute to establishments, it is possible that this grouping was too broad to capture actual effects. In the environmental justice and other public health geography literatures, studies at smaller aggregations more commonly find effects than those at larger ones (Krieger et al., 2002; Ringquist, 2005).

4) <u>Measure does not capture discrimination well:</u> Workers in high percent African American counties might experience less workplace-relevant racial discrimination than others because management and other social leaders are more likely themselves to be African American. Due to the ecological nature of this study, it is not known *who* in a county is getting injured. African American workers in low-African American counties might be especially at risk, reflecting an interaction of individual

and county racial status. As a partial way of assessing this, we tried controlling for segregation (Williams & Collins, 2001) using two different measures. The widely used index of dissimilarity had a low correlation with % African American (r=0.01) and was positively associated with injury/illness (bivariate coef=2.30 (CI: 1.67, 2.94)), although it was not significant in the final model.

5) <u>Race is proxy for another unidentified risk factor</u>, or the controls in the model were inadequate. Potentially correlated factors include county political progressiveness, pro-worker attitudes, and lack of hazardous businesses due to low social resources or urbanicity.

6) <u>Risk *is* lower</u>. It is also possible that risk truly is lower for African Americans than whites and others.

Economy/SES: The finding of increased risk associated with poverty fit with expectations based on the literature such as Dembe et al.'s (2004) analysis of National Longitudinal Survey of Youth data. One county-level analysis was found: in their *Injury Fact Book*, Baker et al (1992) compared low versus high income county death rates for conditions with frequent occupational causation, finding that low income counties had ten times the fatality rates of high income counties for machinery and four times those for falling objects, electric current, and explosions.

Workers who have fewer options may be more willing to accept risk, and their reduced bargaining power reduces business incentive to provide safety (Kahn, 1991; Robinson, 1986; Robinson, 1988). To the extent personal poverty also reflects area business finances, prevention capacity is also lower (Robinson, 1988). Poverty increases worker vulnerability, due to factors like stress and pre-existing medical conditions.

Moreover, studies have shown that workers experiencing job insecurity and precarious employment have greater injury/illness risk and suffer physical and psychological sequelae (D'Souza et al., 2003; Dekker & Schaufeli, 1995; Probst & Brubaker, 2001; Quinlan et al., 2001; Schmitt, 2002).

Beyond the aggregated individual effects, low income areas are also associated with deindustrialization; remaining establishments may be older and/or hazardous. New plants tend to be attracted to economically more vibrant areas (Loomis et al., 2004; Richardson et al., 2004; Roe et al., 2002). While Conway and Swenson’s (1998) analysis found that export and automation were not significant factors in recent nonfatal injury declines across time, Loomis et al.’s (2004) more sophisticated analysis found that deindustrialization explained about 10-15% of the decline in occupational injury fatalities from 1980-1996, and within-state studies of deaths found similar results.

It is valuable to note that while poverty was found to be significantly associated with high rates in a variety of models tested, the direction of effect was not fully consistent. When race, unionization *and* Southern status were controlled for in county-level analyses, poverty was positively associated with high rates. These variables are highly entwined.

Policy and Culture: In contrast to our findings, a body of evidence supports the idea that unions reduce occupational injuries and illnesses (Gray et al., 1998; Human Rights Watch, 2005; Litwin, 2000; O'Neill, 2002; Reilly et al., 1995; D. Weil, 1997; D. Weil, 1991). We suspect that our finding on this issue may be an artifact, if area unionization is a proxy for industry hazard or reporting rather than worker-friendly social climate. While in this database, the correlation between state percent unionized and the state version of the Index of Industry Hazard was only 0.01, it is possible that unions reflect hazard *within* industries. Due to globalization and anti-union drives, unionized plants may be older and in worse shape than non-unionized ones in their same industries. Furthermore, due to the link with older manufacturing industries, high area union concentration is expected to be associated with longterm job loss and insecurity. Unions may also increase reporting, due to factors like better knowledge of rights, safety committee activities, and more job security. Finally, there is misclassification because the unionization variable includes unionization in industries such as construction and mining that are excluded from the sample, as well as in establishments excluded for size reasons.

The positive association between unemployment safety net quality and increased injury/illness also was contrary to the hypothesized relationship. Economic theory suggests that improved safety nets increase workers' bargaining power . However, as with unionization, the improved safety net could help workers feel safer in *reporting* events. It is also possible that this policy reflects historic union advocacy in a state; and if unions are partly a marker for industry hazard, this safety net variable may be one too.

The South: The "South" risk factor was examined due to the map showing low LWDII rates in the South – a particularly striking finding since Southern states have relatively high *fatal* occupational injuries.[7](Richardson et al., 2004) Our sensitivity analysis segmenting Southern states from others suggested that different risk factors may be operating in the South compared to elsewhere. The inverse relationships between high rates and a)percent African American, b) percent rural-farm were much weaker in the

---

[7] Although, as noted above, there are many differences between fatal and nonfatal injuries, and fatal injury databases would not have the exclusions we had, such as of small establishments and certain industries.

South than the North. The association between poverty and high rates was positive in the North but negative and nonsignificant in the South. Unionization was associated with even higher rates in the South than elsewhere.

South may be a marker for racial composition, culture, anti-union attitudes, political conservatism, different industry, or other factors. In their textbook on industrial location, Harrington and Warf (1995) note that areas of the South appeal to large companies with a variety of location options, due to an area lack of support for workers' rights, disfavor of unions, and lack of established unions. Accordingly, workers may experience more disincentives to report injury/illness, and employers may be more antiregulatory and less likely to report or respond to the ODI survey. Finally, the above six explanations related to race (underreporting, establishment size/industry, proxy, wrong aggregation, measurement/discrimination, true effect) may apply here as well.

Industry Hazard: Our sensitivity analysis by industry hazard found that high rate industries drove the regression findings for most variables. If, as suggested above, the analysis was excluding many of the high rate industries in the South, this would contribute to explaining north-south differences in findings.

Several other risk factors are worth comment. The nonsignificant association of LWDII with OSHA regulatory activity could have been caused by enforcement due to high rates early in the five year study time frame, leading to rate reductions that obscure the association. Establishment size has been associated with injury/illness outcomes and employer safety capacity in the literature such as (Stokols et al., 2001), but averaging size across establishments in a county may have prevented the association from being seen in this dataset. The theory behind including environmental contamination risk factors in this study was that the same general forces that affect decisions about protecting the outside community would also affect decisions about protecting workers. Possible reasons why the risk factors did not show statistically significant associations include substantial missing data, the fact that environmental regulation is far stronger than occupational regulation, and the fact that most of the LWDII events are injuries rather than illnesses, meaning that the two databases reflect a different set of establishments and costs of exposure prevention. There is much research suggesting that immigrants are at increased risk of occupational injury and illness (Dembe et al., 2004; Loh & Richardson, 2004; Pransky et al., 2002). Our finding of a negative, non-significant association suggests that underreporting may be a factor (Azaroff et al., 2002; Azaroff et al., 2004).

Study Limitations: The OSHA Data Initiative is not intended as a surveillance database and is limited by its use in enforcement, lack of unique establishment identifiers that would enable combining multi-year observations by establishment, and data quality. Further, because establishments reporting low rates are sometimes excluded from subsequent samples, mean rates in counties with many such establishments may be biased, and the number of sampled establishments would vary more by year. Another concern is that the outcome variable does not take account of the level of confidence in the mean (related to the number of ODI establishments in a county.) Further, this study is essentially cross-sectional and does not address the time it may take for contextual factors to influence business and worker behavior. This study is also ecological and collects no exposure data. The choice of county aggregation has drawbacks; in addition to those discussed above, counties are often not cohesive or homogenous areas. The ecologic fallacy of inferring establishment risk based on county risk must be avoided.

Strengths: This analysis may be the first to articulate and test a broad theory or conceptual framework for why occupational injury/illness risk would vary geographically. It brings multilevel regression and spatial statistics tools to a field that has made little use of them. While the study is cross-sectional, the LWDII outcome probably means that many of the events are injuries with relatively contemporaneous onset, as opposed to longer latency illnesses that could introduce further noise. The study takes a broad look at potential predictors, starting with a large, theory-based database of 90 county and state-level potential explanatory risk factors. Despite the many 'potentials,' the final model included many of the top key risk factors of interest based on theory and literature: poverty, race, unionization, and industry hazard. While county aggregation is a limitation, it is also a strength, in that it can capture much of the area from which workers commute and the broader area social and policy conditions in which businesses operate, and further, in that county findings have county-level policy implications.

This quantitative analysis supports the project of mapping for surveillance purposes by demonstrating that the observed geographic variations are more than random coincidence, more than the distribution of population hazard, more than the distribution of industry hazard. It demonstrates that the geographic distribution reflects associations with social factors as well. Sensitivity analyses and replications in multiple years of the database increase confidence in the findings. The findings support the conceptual framework categories, though not always in the expected direction of effect. The fact that these risk factors were significantly associated with the outcome – *in any direction* – supports the concept that these area-level social risk factors are important in understanding the determinants of location and risk.

Replication of this study in other databases such as state workers' compensation is warranted. Additional in depth studies on industries such as our companion meat industry case study will enable more specific findings.

VI. CONCLUSION

Some might question the benefits of research demonstrating associations between risk and social factors, because social factors often seem intractable. Yet, we *can* change the following. 1) Change area incentives and capacity for prevention, through policy changes such as the living wage, increased minimum wage, and universal health insurance, enforcement and compliance assistance, new standard-setting and legal action, prosecution of executives, increased enforcement of employment law, social marketing, tax breaks, closer connection between workers' compensation premiums and costs or prevention activities, and local publicity about the cost savings from safety interventions (Dorman, 2000). 2) Change the way we intervene, to take account of area factors like language, literacy, or population density that may influence intervention success. While local program staff may already be aware of local social factors, geographic tools help them communicate about it to policymakers and help determine the extent and level of needs. 3) Change the way we target, to focus on areas at risk, or, under a "justice" framework (Morello-Frosch et al., 2002), on areas suffering from both social injustices and high injury/illness rates. Beyond these changes, these analyses also point to a need to change the way we measure. There is a critical need to evaluate and take action against underreporting, particularly given the ODI's use in enforcement.

This analysis demonstrates that occupational injuries and illnesses vary by geography. Examination of social determinants can yield a new comprehension of the etiology of occupational injury and illness and focus attention on the more important and broader-impact distal causes that contribute to proximal causes. Thus we focus on the goal: to improve intervention targeting and design, and ultimately, prevention.

5

Occupational Injury and Illness in Meat Processing: A Geographic Cut

Roni A. Neff, ScM, Frank Curriero, PhD, and Thomas A. Burke, PhD, MPH

ABSTRACT

BACKGROUND: Meat processing is one of the most hazardous industries in the United States. This research explores the association of county- and state-level social factors with plant location and injury/illness risk.

METHODS: Meatpacking, sausage making and poultry processing establishment occupational injury/illness rates in the OSHA Data Initiative (1997-2001, n=1553) were analyzed. Maps present establishments, high injury/illness establishments, co-location with selected risk factors, and indicators of potential underreporting. Linear and multilevel regressions examine associations quantitatively.

RESULTS: Overall, counties with meat processing establishments were relatively high percent African American, non-college educated, and urban, had longterm job gain, and were in states with medium union membership, anti-union policy, and slightly fewer OSHA inspections. By contrast, meat processing establishments with high occupational injury/illness rates were in counties that were relatively white, low per capita income, high school educated, and had longterm job loss.

DISCUSSION: Contrasts between predictors of establishment location and of high rate establishments are discussed. Geographic approaches are useful for intervention design and targeting, surveillance, and communicating about hazards and social issues in industrial meat production. Improved targeting of enforcement and intervention may ultimately help incorporate more of the social costs of meat production into the price of meat.

*　　*　　*　　*　　*

I. INTRODUCTION

The three meat processing industries (meatpacking, the making of sausage and other prepared meats, and poultry processing) are among the most hazardous in the United States. In 2004, all had occupational injury/illness rates more than double the national average (U.S. Bureau of Labor Statistics, 2006), and in 2002 meatpacking had the highest incidence rate of any industry in the country[8] (U.S. Bureau of Labor Statistics, 2004). This study uses mapping and regression to identify counties with particularly elevated rates and to identify area level predictors of both meat establishment locations and establishments with elevated occupational injury/illness rates. Geographic

[8] Specifically, this statement refers to "DART" cases of injury and illness, or those with days away from work, restricted work activity, or job transfer.

variation in potential underreporting is also examined.

Spatial statistics and mapping are useful for intervention targeting, surveillance, evaluation of area-level risk factors and creation of visual tools for discussing hazards and their distribution. Such approaches are now widespread throughout public health (Cromley & McLafferty, 2002; Hillemeier et al., 2003; Krieger et al., 2002; Krieger, et al., 2003; Waller & Gotway, 2004). They have been used much less to understand occupational injury and illness. Geographic analysis has been used in studies of meat animal *production* (Edwards & Ladd, 2001; Roe et al., 2002; Wilson et al., 2002; Wing et al., 2000), but only two studies were identified looking at the geographic distribution of meat *processing* (Drabenstott, 1998, 1999). Further, while research has examined meat processing occupational *injury/illness risk factors* at the individual and plant levels, no known studies examine area-level determinants.

Spatial analysis tools include Geographic Information Systems (GIS) and spatial statistics. GIS applications can be used to glean information from data in a variety of ways beyond simply making maps. In this paper and its associated appendices, we map point locations, compare areas with high and low rates, examine associations between social risk factors and outcomes, and visualize geographic variation in regression residuals. The maps we present illustrate a key benefit of creating maps rather than just looking at non-geographic information; they show that to the extent a relationship exists between two factors, its geographic features can contribute important information. Our regression analysis for industry location takes account of the spatial nature of the data – that is, that nearby areas are likely to be more similar than those farther away and that counties within states all have the same values for state risk factors. Using multilevel methods for this analysis helps improve the accuracy of results.

II. BACKGROUND

"Meat slaughter and processing" comprises three industries: meatpacking (Standard Industrial Classification (SIC) 2011); sausage and other prepared meats (SIC 2013); and poultry slaughter and processing (SIC 2015).[9] There are approximately 5,700 meat processing plants nationwide (U.S. Government Accountability Office, 2005). Excluding most establishments with fewer than 11 employees, our database identifies approximately 1,533 plants, most with over 100 employees. The industry has become highly integrated in the past 25 years, with a few owners controlling most of the large plants, giving them considerable power over industry conditions (MacDonald et al., 2000). Another change is that the dominant areas for plants in each of the three industries have moved, with an increasing portion in the Midwest and South and in more rural areas (Drabenstott, 1999). The industry operates on small profit margins, meaning that production pressures are high (Human Rights Watch, 2005).

Meat processing workers earn on average only $21,320, or about 63% of the average U.S. manufacturing wage; workers commonly consider the wages desirable (Human Rights Watch, 2005; U.S. Government Accountability Office, 2005). An estimated 42 percent of the industry's 527,000 workers are Hispanic, a 17 percent rise since 1994. About one in four workers are estimated to be foreign-born non-citizens, including individuals from many countries and many without documentation (U.S. General Accounting Office, 1998; U.S. Government Accountability Office, 2005). Companies commonly use recruiting and word of mouth to attract workers from other

---

[9] The SIC system was replaced by NAICS in 2002, however, the data used in this analysis were collected under SIC.

states and even from abroad, due to the paucity of willing locals (Franklin, 2005; Olsen, 2003; Rodriguez, 2003; Stull & Broadway, 2004). Unionization in the industry has plummeted; about 46 percent of meat products workers were unionized in 1980, dropping to 21 percent in 1987 and staying at a similar level at least through 1997 (MacDonald et al., 2000), although some energetic organizing campaigns are ongoing today (United Food and Commercial Workers, 2005a; United Food and Commercial Workers, 2005b).

The seemingly inherent hazards in these industries include: sharp tools; repetitive tasks; slippery blood and fat; heavy carcasses; extreme room temperatures necessary for food safety; and airborne dusts, allergens and infectious agents. Complementing these are avoidable risks that might vary geographically with social conditions. Heavy production pressures lead to risks like mandatory overtime, inadequate rest breaks, and disassembly line speeds faster than workers report they can safely handle (Human Rights Watch, 2005; U.S. Government Accountability Office, 2005). Employee turnover can be as high as 100% per year (MacDonald et al., 2000; U.S. Government Accountability Office, 2005) – and studies (outside the meat industry) often show new workers to be at the highest risk for injury/illness (Robinson, 1988). Inadequate industrial hygiene controls, from personal protective equipment to engineering, lead to unnecessary exposures (Human Rights Watch, 2005). There are also widespread “safety incentives” and other pressures on workers not to report injuries/illnesses nor to be absent from their posts (Human Rights Watch, 2005; Schlosser, 2001; Stull & Broadway, 2004; U.S. Government Accountability Office, 2005). Accordingly, hazards that cause minor injuries or near-misses remain unaddressed, and small injuries and illnesses may escalate without rest or treatment. Finally, despite these conditions and the low pay, workers know there is a ready supply of replacements willing to take their jobs, increasing their acceptance of hazards and pain (Schlosser, 2001; Stull & Broadway, 2004).

Reported injury and illness rates in the meat processing industry have plummeted since the early 1990's, far faster than rates for private industry as a whole. For example, in 1992, meatpacking had a rate of 23.3 Lost Workday Injuries/Illnesses (LWDII) per 100 workers, and by 2001, it was 11.0, whereas for all manufacturing, the rate dropped from 5.4 to 4.1 during that time (U.S. Bureau of Labor Statistics, Undated). These drops may be partly due to recent attention and resultant prevention efforts. Both the poultry and the meatpacking industries have worked jointly with the Occupational Safety and Health Administration (OSHA) to develop voluntary ergonomic standards (OSHA, 1993; OSHA, 2003) and other guidelines.(OSHA, Undated a; OSHA, Undated b) OSHA and trade association officials report these have been widely and effectively used (U.S. Government Accountability Office, 2005). During our research period, OSHA had a voluntary partnership related to safety with ConAgra Refrigerated Foods and since 2002 has had one with the American Meat Institute. In the one identified program focused on a geographic area at risk, OSHA's Omaha area office developed a voluntary partnership program with meatpacking plants. They met together regularly to discuss safety, leading to a reported 39 percent drop in total injury/illness cases (Franklin, 2005).

Beyond voluntary programs, there are few firm pressures on the meat industry to reduce injury/illness rates, other than the costs of workers' compensation and replacing workers and the small risk of OSHA inspection and fines. OSHA conducted 1,900 inspections of meat processing plants between 1995 and September, 2004 -- accounting for less than one percent of the agency's inspections, despite the industry's hazards (U.S.

Government Accountability Office, 2005). OSHA's fines are low in general; a *New York Times* investigation found that even when a worker dies due to a (rarely issued) "willful" violation, the median fine is only $30,240; further, OSHA sought prosecution in only seven percent of "willful death" cases (Barstow, 2003). Finally, while after the fall of OSHA's ergonomics standard the agency promised to cite employers for ergonomics hazards using its General Duty Clause, only 28 citations have been made since January 2001 and none within the meat processing industries (OSHA, 2006). Overall, the message may be that it is cheaper for employers to deal with the consequences of injuries and illnesses than to prevent them, given that they can externalize many of the costs to workers and society (Boden et al., 2001; Dembe, 2001; Dorman, 2000; Leigh & Robbins, 2004).

The speed and magnitude of the drops in reported rates have led some to question the data's validity (GAO, 2005). While as many as 2/3 of all occupational injuries in the United States may be undercounted, concerns about underreporting in this industry have been particularly strong (HRW, 2005; Leigh et al., 2004; Rosenman et al., 2006; Schlosser, 2001). Reasons for concern about this industry include the high percent of immigrant workers, pressures not to report, and the fact that musculoskeletal conditions, ubiquitous in the industry, are among the most underreported conditions nationwide. OSHA does try to address underreporting by verifying reported rates during targeted inspections, randomly inspecting 200 worksites in high-hazard industries that report low rates, and conducting comprehensive record-keeping audits, however, these programs reach only a small percentage of establishments in the industry (GAO, 2005)

Conceptual Framework: This analysis is based on a conceptual framework (elaborated in a companion paper) that defines four key area conditions that affect occupational injury/illness risk: demographics, local economy/socioeconomic status (SES), policy/culture/values, and industry hazard (Neff, Curriero, & Burke, 2006). Through impacts on the balance of power between business and workers and other mechanisms, these geographic risk factors affect business and worker incentives and capacity to provide safe workplaces. In the context of meat processing, it is hypothesized that weaker local economy, minority, immigrant, and less-educated demographics, pro-business policy/culture/values, and more meatpacking than poultry or sausage establishments will be associated with high injury/illness rates.

In this study we consider meat processing establishment locations themselves as conditions of potential concern for worker health. Human Rights Watch stated that there are "systematic human rights violations embedded in meat and poultry industry employment." These include a failure to use known injury/illness prevention methods, denial of workers' compensation, interference with unionizing, and mistreatment of immigrant workers (HRW, 2005). Further, studies have documented environmental injustice in siting animal *production* establishments, and there could be parallels in meat *processing* (Edwards & Ladd, 2001; Wilson et al., 2002; Wing et al., 2000).

Meat establishment locations are expected to be associated with many of the same social risk factors as high rate establishments in general. Meat plants need to be in areas with pools of workers willing to accept the jobs they offer at the wages they pay, in enough quantity to cover for their high turnover rates (Roe et al., 2002). This suggests they will gravitate to urban areas of moderately weak economy. (Studies suggest areas with the highest percentages of poverty and African Americans lack the infrastructure to attract any industries (Morello-Frosch et al., 2002; Wilson et al., 2002)). Establishments that are choosing locations (i.e., newer and more mobile ones typically owned by large firms) are also expected to seek out areas with relatively low worker power and with pro-business policies (Harrington & Warf, 1995; HRW, 2005; Wilson et al., 2002).

Literature on meat establishment location suggests that other predictors are urbanicity (especially small towns, although newer establishments are in more remote rural areas),(Drabenstott, 1999) locations of animal growers, distribution centers, and corporate headquarters, transit routes, and area location incentives (Roe et al., 2002). Animal grower location is affected by lack of community opposition, lax environmental regulations, low population density, and inexpensive land and labor, so processing facilities might be likely to be in urban areas near sites with those characteristics (Drabenstott, 1999; Roe et al., 2002).

III. METHODS:

A. DATA

i) Meat Occupational Injury/Illness Database:

The primary database for this analysis was the OSHA Data Initiative (ODI), an annual survey of establishments in most high injury/illness industries, conducted for use in enforcement. A detailed description may be found in a companion paper (Neff, Burke, & Curriero, 2006). OSHA provided a complete 1997-2001 listing of responding facilities. Unfortunately it did not collect unique establishment identifiers, and many establishments were surveyed in multiple years.

In 2005, OSHA responded to a Freedom of Information Act request by releasing establishment names with exact LWDII rates. This more exact database for 1997-2001, now available on the Internet (OSHA / The Memory Hole, 2005), was used in most of the analyses in this paper although unlike the data OSHA provided, it lacks information on establishment size. All establishments in this database coded with SICs 2011 (meat), 2013 (prepared meats) and 2015 (poultry) were extracted (n=6,435). The three meat products industries and the five study years 1997-2001 were grouped together because similar area-level risk factors were expected to be operating in each. Sensitivity analyses generally supported these choices. Records in the database were matched by computer followed by hand-matching to unite all records for each establishment across years. (Appendix 16 describes matching methods.)

The database covers only plant employees, excluding contract employees like most cleaners and maintenance workers – groups believed to be at highest risk in the meat industries.(GAO, 2005) The LWDII outcome in this analysis covers all reported injuries and illnesses and cannot distinguish among event types. The Bureau of Labor Statistics found that knife injuries and repetitive strain disorders (technically an illness) dominated reporting in these industries (BLS, Undated).

Establishment locations were geocoded to zip codes. A county-aggregated file was created based on zip code centroids and was merged with the below-described social risk factor database. The decision was made to perform analyses at the county and state rather than zip code level because the area-level exposures had most meaning (and were measurable) at these broader levels. In the median U.S. county 69.4 percent of employees live in the county where they work (U.S. Bureau of the Census, 2000).

The outcome measures were: (a) establishment location: counties that had meat industry establishments; and (b) county mean injury illness rates for meat processing establishments.

ii) Social risk factor database:

Based on the conceptual framework categories, a database of 90 risk factors at the state and county level was collected from the U.S. Census and other sources.(Appendix 10) Table 19 lists risk factors that were included in models, categorized by construct, with data sources and comments on why they were included.

Table 19: Meat Risk factor database

RISK FACTOR	SOURCE	WHY INCLUDED
Demographics		
Percent African American	Census 2000. Calculated.	Due to institutional discrimination, counties with high % African American may have more individuals willing to accept meat industry jobs.
Change in percent foreign born, 2000 – 1990	Census 2000, Census 1990. Calculated.	An estimated 1/4 of meat industry workers are foreign born, and many immigrants are known to be attracted to meat industry establishments from other areas in search of jobs.
Percent not completing high school	Census 2000. Calculated.	Suggests extent to which residents may be willing to accept meat industry jobs based on education and access to other options.
Percent with bachelor's degree	Census 2000. Calculated.	Same as above.
Local Economy		
Per capita income	Census 2000.	Suggests residents' economic vulnerability/strength.
Change in unemployment, 1996 – 1975	Census "Counties 1998" file. Calculated. Positive number means increased unemployment.	Same as above. Also may reflect unmeasured aspects of industry hazard if older establishments are riskier.

RISK FACTOR	SOURCE	WHY INCLUDED
Policy/Values/Culture		
Percent unionized 1999 [divided into low, medium, and high splines]	Unionstats copyright 2002 by Barry T. Hirsch and David A. Macpherson. Data source: Current population survey. Available at: http://www.trinity.edu/bhirsch/unionstats/ Accessed 9/30/04. Calculated. Numerator includes public and private industry. Denominator = state employment 1999.	State level risk factor initially included as a measure of worker power. Can also reflect industry hazard and reporting.
Right to work policy (Y/N)	Right to Work States. Available at: http://www.nrtw.org/rtws.htm. Accessed: January 17, 2006.	State level risk factor. "Right to work" policies allow non-union workers to work in unionized establishments, and are sometimes used as a marker of anti-union policy.
Inspections per (employee * 100,000)	OSHA's Inspection Management Information System, Establishment Search Page: http://www.osha.gov/pls.imis.establishment.html. Accessed December 7, 2004. Total # inspections by state from 1/1/97 – 12/31/01. Calculated, denominator = state employment 1999.	State level risk factor reflecting inspections in all industries for all reasons. Used as an indicator of the intensity of OSHA activity in a state.
Industry Hazard		
# Meatpacking establishments	OSHA Data Initiative	Meatpacking has the highest mean LWDII of the three industries.
# Sausage establishments	OSHA Data Initiative	
# Poultry establishments	OSHA Data Initiative	Poultry has the lowest mean LWDII of the three industries.
Control		
Number of years an establishment was surveyed	OSHA Data Initiative	Plants reporting low rates were often not resurveyed in subsequent years.

RISK FACTOR	SOURCE	WHY INCLUDED
% living in an “urbanized area” (UA)	Census 2000. Calculated	UA is defined as a densely settled area with >50,000 people.
% living in an “urban cluster” (UC)	Census 2000. Calculated	UC is defined as a densely settled cluster with population between 2,500 and 50,000.

B) ANALYSIS:

1) Exploratory Analyses

Exploratory data analysis and exploratory spatial data analysis were performed to examine the databases. Maps were created to depict point locations of establishments by industry, size, and reported injury/illness rate (Appendices 17, 18). Correlations between variables were evaluated (Appendix 19.

2) Associations between Social Risk Factors and Outcomes

A simple, exploratory visualization tool was developed in collaboration with Susan Baker (2006), to show risk factors vary within a) counties that have meat processing establishments; (b) the top 1/3 counties based on mean injury/illness rates in the meat industries. For establishment location maps, the map categories are:

- Has establishments, LOW tertile of risk factor
- Has establishments, MEDIUM risk factor
- Has establishments, HIGH risk factor

The process is repeated for the second set of maps, with the difference that the first variable is “top tertile of mean county injury/illness rates.” In these maps, only the 700 counties with meat establishments (out of 3141 nationally) were eligible to be on the map, making for a relatively disjointed and sparse picture.

To focus on the key information, a white background is used for states not in the database, counties without establishments, and in the second set of maps, counties with mean LWDII rates in the bottom and middle tertiles. We emphasize that these maps simply depict co-location of potential risk factor and outcome; they do not represent controlled analysis.

The following county-level maps are presented, selected for their interest and the contrast between findings for the two outcomes:

- Distribution of longterm job loss and gain among counties with meat establishments.
- Distribution of longterm job loss and gain among high rate meat counties
- Distribution of percent African American among counties with meat establishments
- Distribution of percent African American among high rate meat counties

See (Neff, Burke, & Curriero, 2006) for further methodological details. Additional maps are available from the authors.

3) Quantitative Analyses of Relationships:

Regressions were used to examine social risk factors for both establishment location and high rates.

A. Establishment location: A multilevel logistic regression model was built examining predictors of U.S. counties (n=3141) having meat establishments. Multilevel modeling was used to adjust the standard errors of the prediction for the fact that there is less opportunity for variation than the logistic regression model assumes (Snijders & Bosker, 1999). This is because counties within a state necessarily have identical values for the state level risk factors and because the landscape of risk factors varies more gradually than can be explained at the county level.

First, a multivariate logistic regression was built at the county level (outcome=

"county has at least 1 meat establishment" (Y/N)). County-level risk factors were added based on the conceptual framework and statistical significance criteria ($p<0.05$). Next, state level risk factors were added and a multilevel logistic model was run, with error terms at the state and county levels. Model selection and testing of parameter significance was based on reduction in model deviance (Snijders & Bosker, 1999). See (Neff, Curriero, & Burke, 2006) for further methodological details.

The model was evaluated using nonspatial and spatial methods, including examination of a standardized residuals map and examination of correlation and variance inflation factors to assess multicollinearity. An analysis was performed using the original database provided by OSHA, both aggregated across years and separately by year, and the magnitude, direction, and statistical significance of results changed little.

B) High Rates

Linear regression analyses were conducted to examine county-level predictors of mean establishment LWDII among the 700 counties with meat establishments. Multi-level models were not used because no state-level predictors were statistically significant in linear regressions. By-industry analyses are also presented for comparison, although few coefficients reached statistical significance, presumably due to low numbers.

4) Analyses of Possible Underreporting

Several analyses were conducted to examine possible underreporting. Two subjectively-defined indicators of potential problems were used: establishments reporting LWDII of zero, and establishments reporting drops of more than six points from year to year. For the latter, six was chosen because when the year to year drops were averaged by county, 6.4 was the standard deviation, reflecting the spread of data. This measure captures about the 10th to 18th percentiles of establishments, varying by year. While many establishments accurately report both of these findings, patterns in either indicator may suggest patterns of inaccurately reporting these outcomes. Establishments with these indicators were mapped and examined.

Computing: Quantitative analyses were performed in Stata (Intercooled version 9.1) (StataCorp, 2005) and Excel 2003. ArcGIS v. 9.0 was used for mapping (ESRI, 2005).

IV. RESULTS

1) Exploratory Analyses

The matching identified 1553 meat processing establishments located in 47 U.S. states. Meatpacking had the highest LWDII of the three, and poultry, the lowest. Table 20 below describes the database. The meatpacking industry had 519 establishments and a mean LWDII of 9.78. The sausage industry had 461 establishments and a mean LWDII of 8.15. There were 573 poultry plants, with a mean LWDII of 7.91.

Table 20: OSHA Data Initiative meat industry data

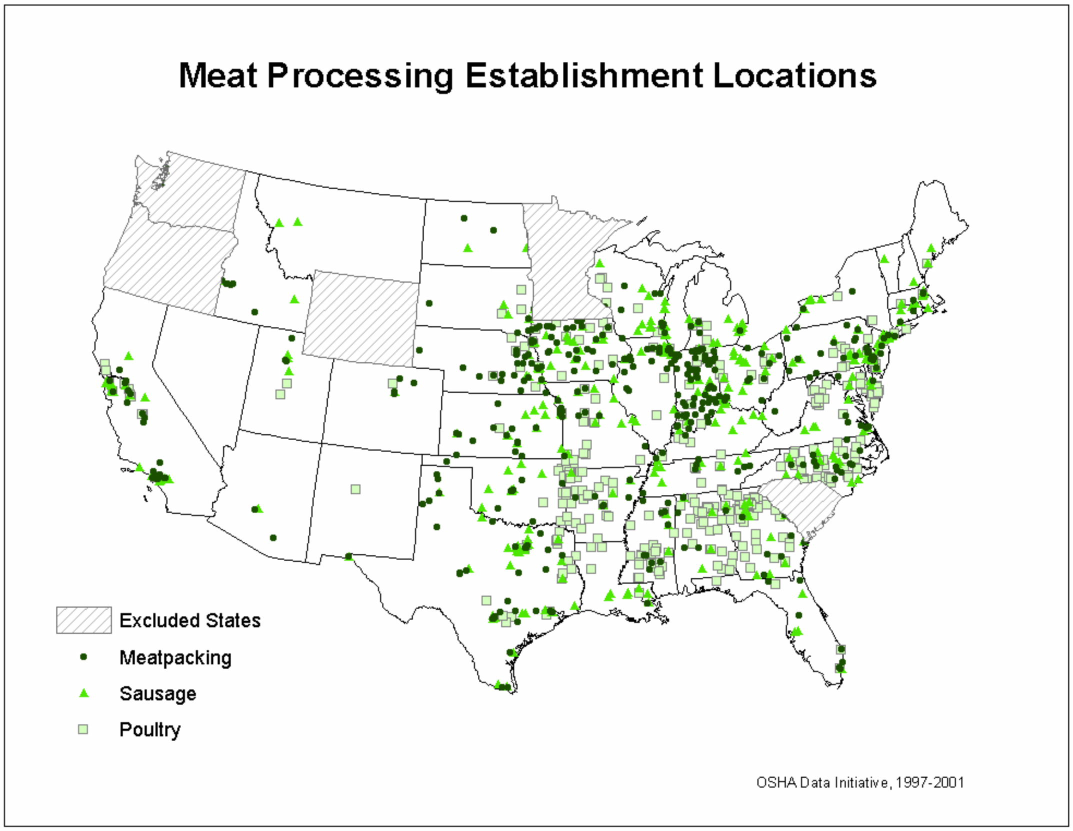
	Meat # Estabs	Mean LWDII (range)	Sausage # Estabs	Mean LWDII (range)	Poultry # Estabs	Mean LWDII (range)
1997	253	14.39 (0,94.4)	208	9.46 (0,39.6)	308	9.11 (0,42.1)
1998	239	13.18 (0,62.1)	215	9.45 (0,59.9)	254	10.51 (0,70.7)
1999	355	9.07 (0,51.2)	323	8.46 (0,40.4)	356	8.1 (0,64.9)
2000	220	12.49 (0,50.5)	206	9.45 (0,35.6)	252	8.24 (0,42)
2001	189	11.82 (0,74.9)	180	8.72 (0,53.7)	229	7.75 (0,29.8)
ALL YEARS*	519	9.78 (0,58.6)	461	8.15 (0,43.1)	573	7.91 (0,128.3)

* Note: the "all years" LWDII results reflect averaging all survey responses for each establishment, and then averaging all establishments, making the single year and all-years means not directly comparable. Approximately 40% of establishments were surveyed in 3 or more years.

2) Analysis of Establishment Location:

Map 17 depicts establishment point locations by industry.

Map 17: Meat processing establishments by industry

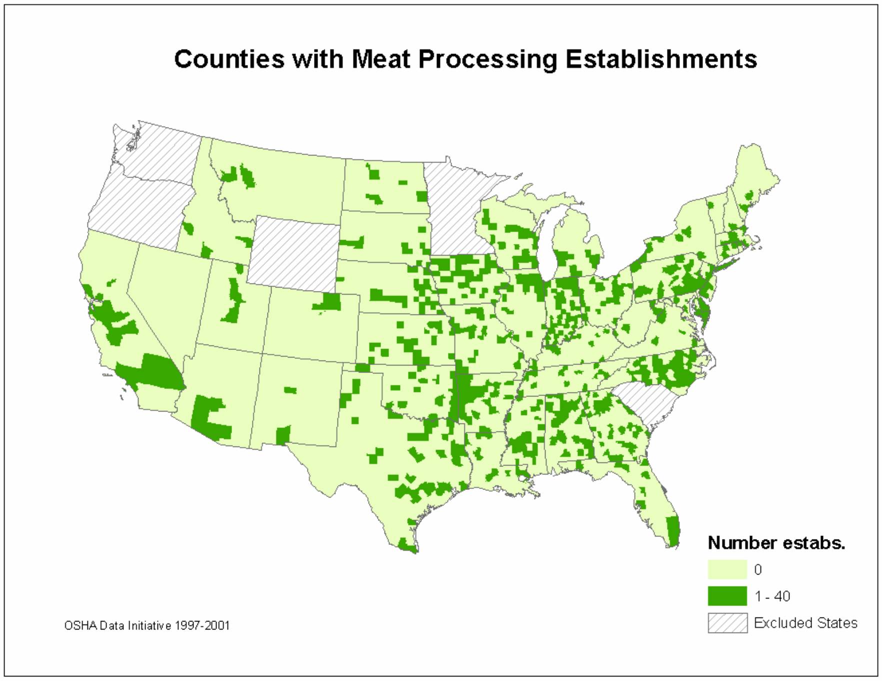


It can be seen that the three industries have different geographic distributions, with meatpacking most prominent in the Midwest, poultry most prominent in the South and East, and sausage more widely distributed. Nearly all establishments in the sample are in the eastern half of the country. The data originally provided by OSHA enabled examination of establishment size, finding that plants were generally larger in the South than elsewhere (Appendix 17). This finding was strongly influenced by the large size of Southern poultry plants. Appendix 18 shows the distribution of high rate establishments (LWDII >=8.0) versus other establishments by industry. It is difficult to visually discern patterns from these maps due to small numbers.

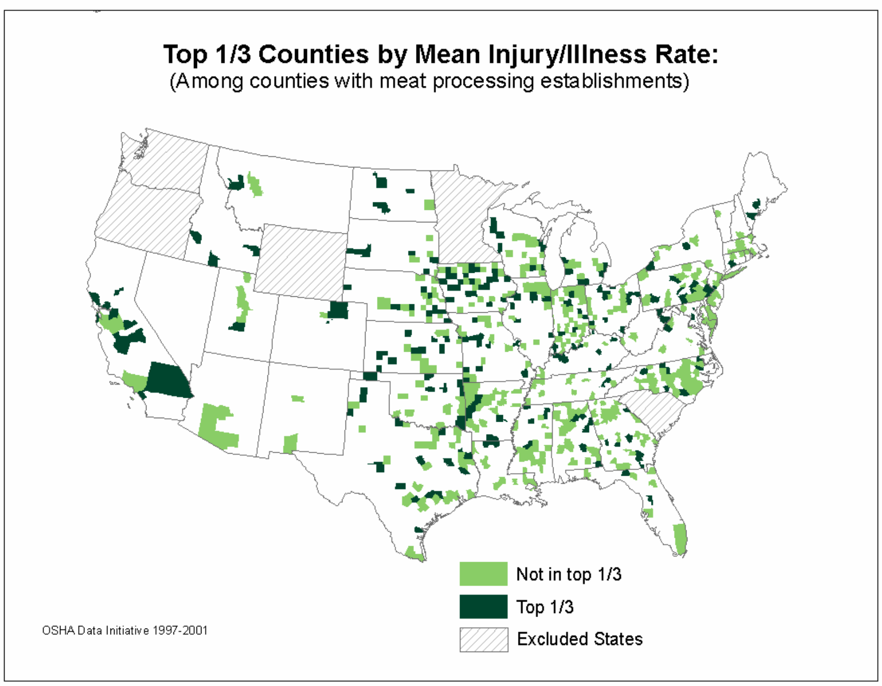
3) Associations between Social Risk Factors and Outcomes

To provide context, Maps 18 and 19 show counties with meat industry establishments, and the top 1/3 of counties by mean injury/illness rates, versus others.

Map 18: Counties with meat processing establishments

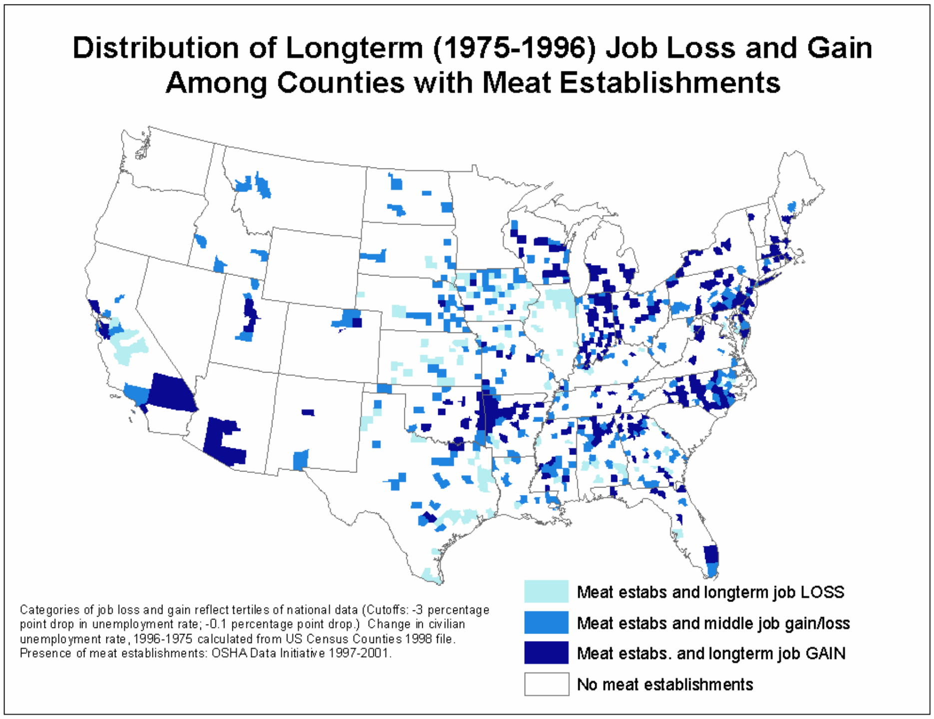


Map 19: Top 1/3 of meat processing counties with highest mean injury/illness rates

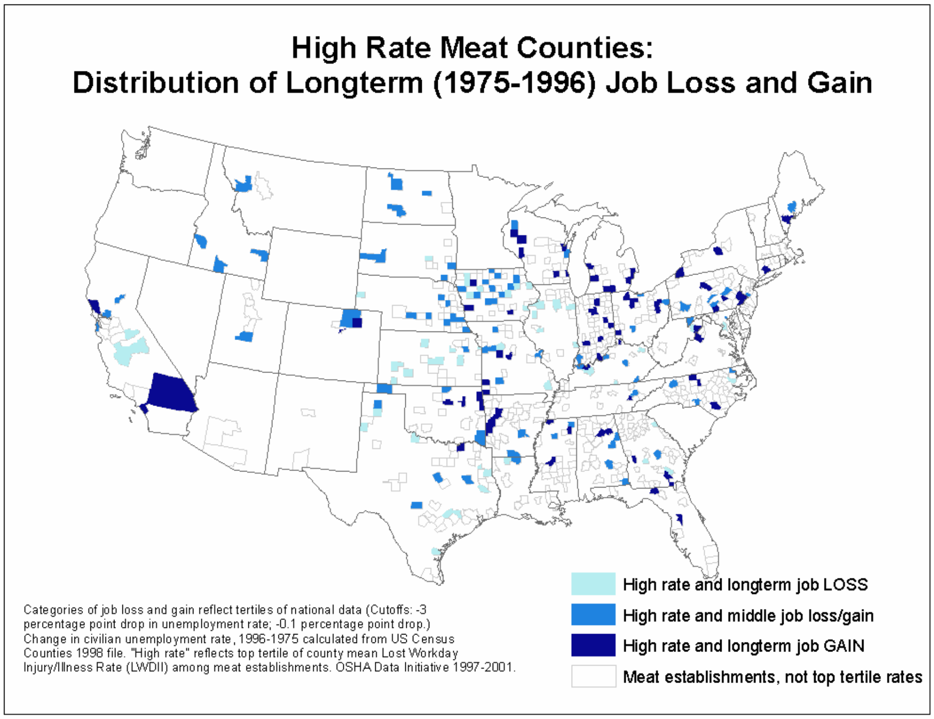


Map 20 shows the relationship between meat establishment locations and county longterm job gain and loss, while Map 21 examines longterm job gain/loss in the top tertile of county mean injury/illness rates. Map 20 shows apparent patterning in the relationship between establishment location and longterm job gain. Areas of overlap include the upper Midwest (Michigan, Wisconsin, Ohio) and Northeast, as well as North Carolina. Map 21 shows that overlaps of longterm job loss and high rate establishments focus in the Midwest, particularly areas near Iowa and Kansas. Small groups of counties with job gain are also distributed through the map. (The large blocks highlighted in the Southwest on both maps reflect small numbers of counties.)

Map 20: Distribution of longterm job loss and gain among counties with meat establishments.

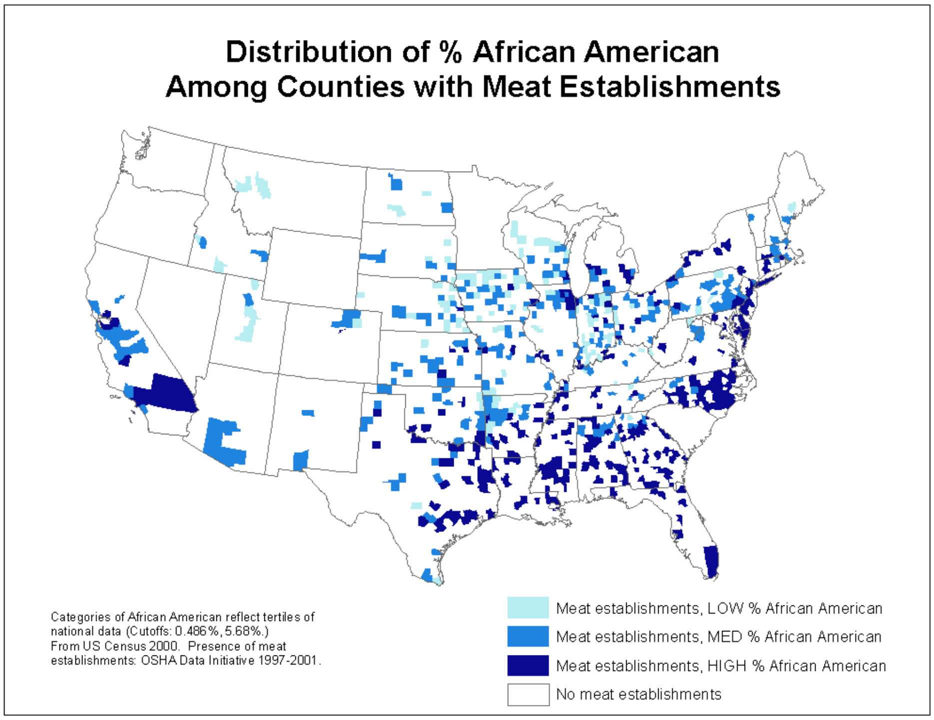


Map 21: Distribution of longterm job loss and gain among high rate meat counties

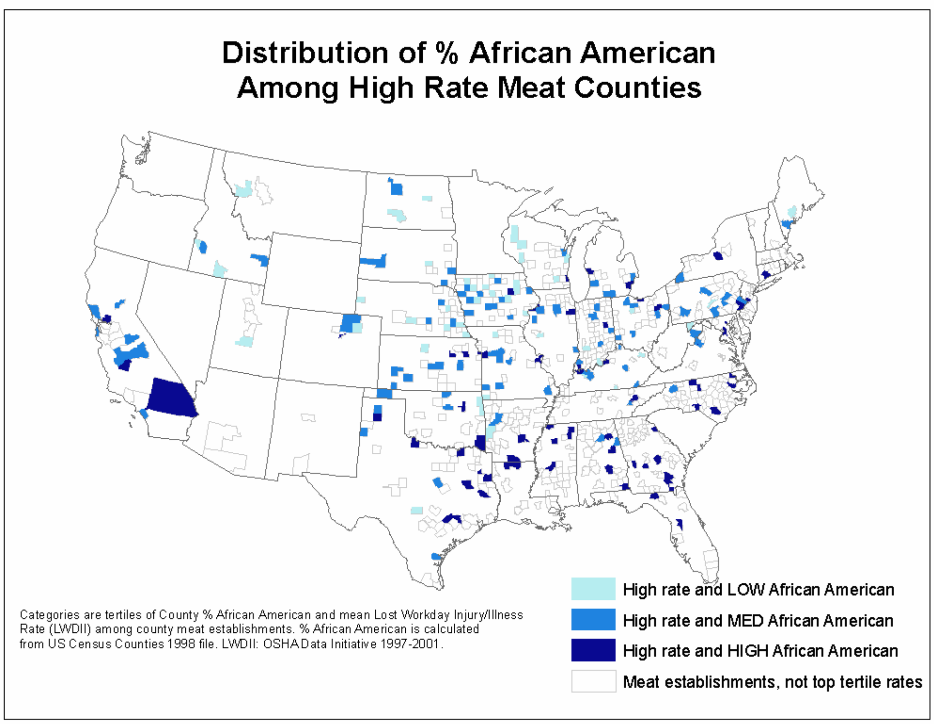


The next two maps show associations between county percent African American and (Map 22) meat establishments, and (Map 23) high mean establishment occupational injury/illness rates. Overlap between African American counties and establishment location focuses in the South and probably includes mostly poultry establishments. Map 23 shows relatively high overlap between *low* percent African American and high injury/illness. The association is strongest in Iowa, and is seen sporadically in nearby and Northern states.

Map 22: Distribution of percent African American among counties with meat establishments



Map 23: Distribution of % African American among meat counties with high mean injury/illness rates



4) Quantitative Analyses of Relationships

Table 21 provides results from the multilevel model analyzing predictors of meat industry locations. Establishments were found in areas that were relatively high percent African American, non-college educated and urban, had longterm job gain, and were in states with medium levels of union membership and slightly reduced levels of OSHA inspections. Because of their relationships with the outcome, the variables describing immigrant status could not be included in these models, yielding coefficients in the thousands. However, it can generally be stated that establishments were more common in counties with high percent immigrant. Poverty was non-significant, but was kept in the model as a control variable that affected other coefficients. Other risk factors from the conceptual framework yielded nonsignificant results.

Table 21: Significant predictors of meat industry locations based on logistic multi-level model*

Construct, Variable	Odds Ratio (95% CI)
Demographics, workforce	
% African American	1.007 (1.003, 1.011)
Percent with bachelor's degree	0.989 (0.98, 0.995)
Local economy	
Increased unemployment 1975-1996**	0.92 (0.87, 0.96)
Social, Policy	
Percent unionized 1999 – low	0.62 (0.5, 0.8)
Percent unionized 1999 – medium	1.19 (1.1, 1.3)
Percent unionized 1999 – high	0.87 (0.8, 0.98)
Right to work policy (Y/N)	2.56 (1.5, 4.3)
Inspections per employee x 100,000	0.999 (0.998, 0.9995)
Control	
% living in an "urbanized area" (UA)	1.013 (1.011, 1.015)
% living in an "urban cluster" (UC)	1.01 (1.008, 1.012)
Percent earning < poverty line	0.991 (0.98, 1.001)
Model Condition Number	*8477.29*
Log Likelihood	*-1413.89*
Variance of State random effect	*0.43 (.13)*

*Outcome = County has at least one meat establishment (Y/N)
** Note: Because the odds ratio was <1, the model predicts increased *employment,* not unemployment.

Map 24 depicts standardized residuals from this regression in counties with meat establishments. The residuals are not randomly distributed across the meat counties, suggesting the model may have missed important geographically varying covariates. The model is more effective (residuals were smaller) in areas of the country where meat establishments are more common.

Map 24: Standardized residuals from multilevel model measuring predictors of meat establishment locations

Lighter colors reflect smaller residuals, meaning the model was more predictive.

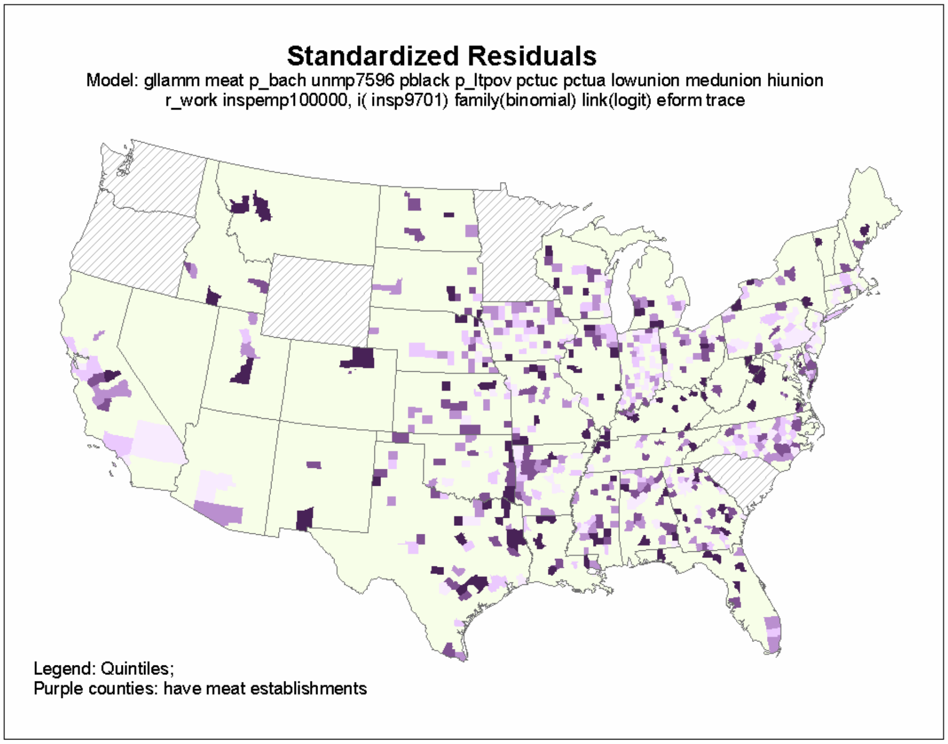


Table 22, next page, presents linear regression results for predictors of establishment injury/illness rates. Note that coefficients from this linear regression cannot be directly compared with those in Table 21, because the latter presents odds ratios. In the three industries combined, every point increase in % African American was associated with a .05 point reduction in mean county LWDII, while each additional percentage point of population not completing high school was associated with a 0.12-

point reduction in injury/illness rates. Areas with longterm increased unemployment, low per capita income, and establishments surveyed in more years of the ODI also had higher injury/illness rates. In the by-industry comparisons, few of the results reached statistical significance, although the coefficients were generally similar. Even where coefficients differed, such as for percent African American in the poultry industry, the 95 % confidence intervals overlapped findings for the combined industry model, suggesting that the true associations in these industries might not differ.

Table 22: Associations between social risk factors and occupational injury illness rates based on county-level linear regression: All meat processing and each industry separately.*

RISK FACTOR	All Meat Processing (95% CI)	Meat	Sausage	Poultry
Demographics, workforce				
% African American	-0.04 (-0.07, -0.01)	-0.02 (-0.08, 0.04)	-0.05 (-0.11, 0.004)	-0.001 (-0.05, 0.04)
Change % Foreign-Born	0.13 (-0.07, 0.33)	0.47 (0.18, 0.76)	-0.29 (-0.6, 0.03)	-0.36 (-0.69, -0.02)
% < High School	-0.12 (-0.21, -0.03)	-0.15 (-0.31, 0.01)	-0.08 (-0.21, 0.05)	-0.06 (-0.20, 0.08)
Local economy/SES				
Longterm increased unemployment, 1975-1996	0.23 (0.1, 0.4)	0.26 (0.01, 0.5)	0.18 (-0.1, 0.4)	-.02 (-0.2, 0.2)
Per Capita Income	-0.0003 (-0.0004, -0.0001)	-0.0003 (-0.0006, -0.00005)	-0.0002 (-0.0004, -0.00004)	-0.0001 (-0.00003, 0.0002)
Control Variables				
Mean number of years in which county establishments were surveyed	1.70 (1.3, 2.1)	2.46 (1.9, 3.0)	1.28 (0.8, 1.8)	0.94 (0.4, 1.5)
Model N (# counties)	699	313	266	297
Model Adj R-squared	0.15	0.26	0.14	0.05

* Outcome = mean of establishment lost workday injury/illness rates (LWDII) in county

The risk factor, "change in percent foreign born between 1990 and 2000," was strongly positively associated with high rates in meatpacking and strongly negatively associated with high rates in the sausage and poultry industries, probably contributing to its not reaching statistical significance in the combined model. Risk factors not listed in these tables – including the number of establishments by industry – were not statistically significant and were thus excluded from the final model.

5) Analyses of Possible Underreporting

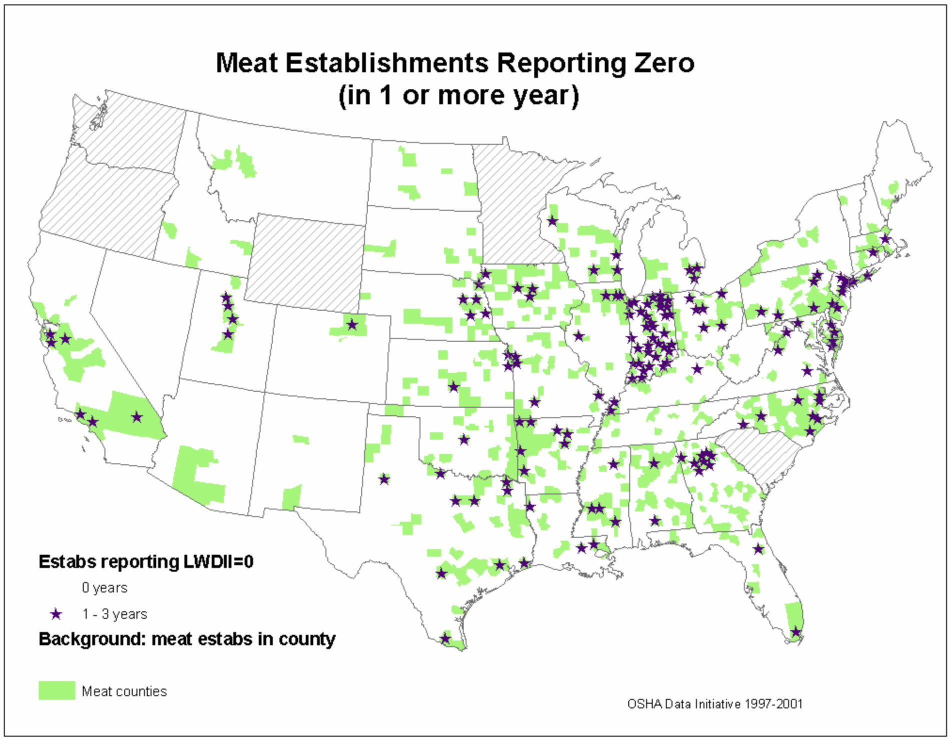
Map 25 depicts establishments reporting LWDII rates of zero by industry. The distribution appears fairly widespread, but geographic patterns are not readily apparent. Possibly zero reports are more common in the Midwest; there is an unexplained cluster of zero reports in Indiana in 1999. Map 26 shows establishments reporting drops of over 6 points in LWDII between any two years. Similar to the other indicator, these sites are fairly widespread and it is difficult to identify a pattern. Table 23 describes further that the frequency of reporting rates of zero *and* that of dropping by over six points increases in opposition to industry LWDII trends. That is, both types of finding are most frequent in meatpacking and rarest in poultry. This finding is in contrast to the trend of mean LWDII rates across industries.[10]

Table 23: Indicators of underreporting: Reports of LWDII rates of "Zero"; Year-to-year LWDII drops of at least six points , By industry

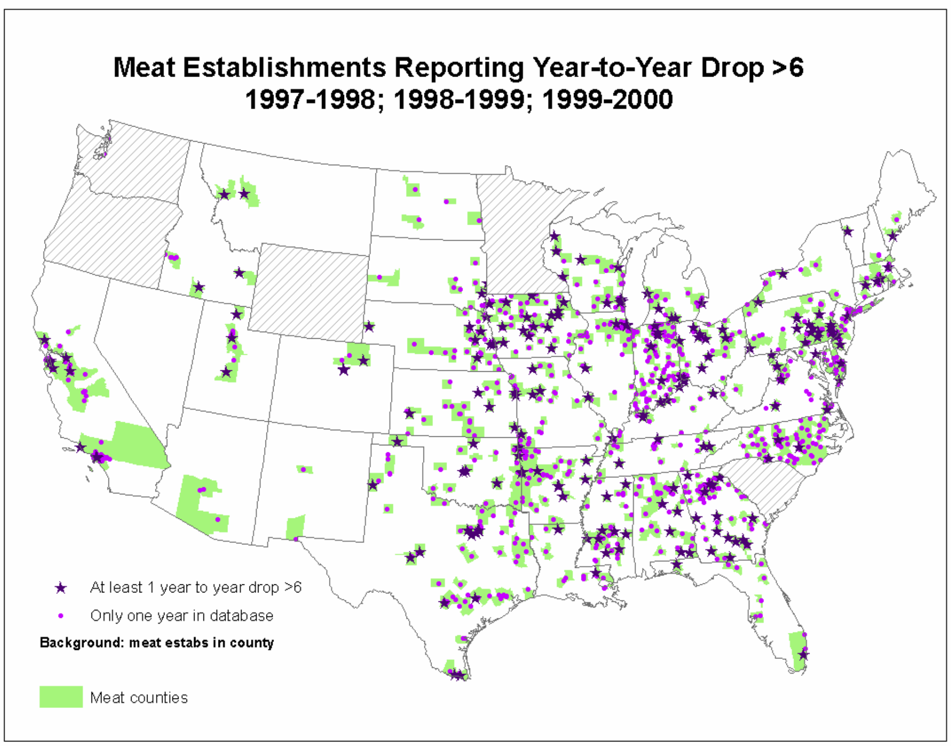
	Meat	Sausage	Poultry
% Reporting "Zero"	15.0	8.03	6.46
% Dropping >=6 in a year	15.6	13.7	12.5
Mean LWDII	9.78	8.15	7.91

[10] In fairness, it should be noted that while meatpacking is consistently found to have the highest LWDII rate of the three industries, poultry and sausage rates do tend to be somewhat similar, and in some years poultry has had higher rates than sausage. (BLS industry data, 2006).

Map 25: Establishments reporting LWDII of "Zero" in one or more years



Map 26: Establishments reporting year-to-year drops >6 LWDII points



V. DISCUSSION

A. FINDINGS

1. Social Risk Factors:

Based on the project conceptual framework we hypothesized that occupational injuries and illnesses would be increased in areas with minority, immigrant, and less-educated demographics, weaker local economies, pro-business policy/culture/values, and more meatpacking than poultry or sausage establishments. Establishment locations were expected to be associated with similar social risk factors, as well as urbanicity.

The finding that occupational injury/illness was associated with low percent African American counties runs counter to the conceptual framework. An initial interpretation is that the more hazardous meatpacking industry is primarily northern, while the lower rate poultry industry is more commonly in the south; however, the finding was consistent across all three industries although not statistically significant. In fact, this finding *is* consistent with the literature's mixed findings on nonfatal occupational injury/illness and race (Dembe et al., 2004; Murray, 2003; Oh & Shin, 2003; Robinson, 1989; Simpson & Severson, 2000; Smith et al., 2005). An analysis also finds rates for the ODI (multi-industry) database overall to be lower in the South and in heavily African American counties than elsewhere (Neff, Curriero, & Burke, 2006). These findings contrast with the literature on *fatal* occupational injury/illness, which tends to match our conceptual framework expectation of risk for African Americans (Loomis & Richardson, 1998; Loomis et al., 2003; Murray, 2003; Richardson et al., 2004; Stout et al., 1996), although at least one study found equal risk (McGwin et al., 2002). The fatal/nonfatal difference may result partly from the relative ease of underreporting nonfatal injury as well as differences in severity, access to health care, and other factors.

Elevated injury/illness was found to be associated with higher completion of high school but non-significantly associated with college completion. A small literature suggests that injury/illness tends to be higher in areas with low education, indicating resident vulnerability on the job market and possibly less safety and rights knowledge (Oh & Shin, 2003; Robinson, 1984).

Numerous studies have documented increased occupational injuries and illnesses among immigrants, especially recent immigrants and those from Mexico (Loh & Richardson, 2004; Pransky et al., 2002). Other studies have shown that employers recruit immigrant workers to their plants – and even the fact that immigrant workers often transform the communities near meat processing establishments. They have also shown the special vulnerability of immigrant workers in these industries due to fears about immigration authorities, lack of workers' compensation coverage or other safety nets, and discrimination (Franklin, 2005; HRW, 2005; Olsen, 2003; Rodriguez, 2003; Stull & Broadway, 2004; GAO, 1998). At the same time, occupational injuries and illnesses are especially likely to be underreported among immigrant workers (Azaroff et al., 2004; HRW, 2005), potentially biasing the findings. In our models, high percent immigrant areas were associated with meat processing establishment locations, and a county's increase in percent of immigrants between 1990 and 2000 was strongly associated with high injury/illness rates in the meatpacking industry. By contrast, counties with especially low increases or even loss of immigrants were associated with high rates in poultry and (nonsignificantly) in sausage-making. In addition to underreporting, effects may have been obscured by the fact that immigrant workers are widespread in the industry and/or the fact that county population may not reflect plant employment.

These analyses showed that injury/illness rates were higher in areas with long term job loss, while meat establishments themselves were more common in areas with long term job gain. The latter areas may have more new plants, which may be safer due to improved technology or because they have tended to be larger as the industry has consolidated – and larger plants are more likely to have capacity to create better safety programs. The literature supports the idea that occupational injury/illness would be higher in areas with weak local economies or deindustrialization, due to worker economic vulnerability and willingness to accept risk, as well as possibly to older facilities (Dembe, 2001; Richardson et al., 2004). Further, counties with many low income residents may have more workers suffering from pre-existing health conditions that reduce their capacity to prevent or withstand injuries and illnesses.

Fewer meat establishments were found in states with both the highest and the lowest percent unionized, whereas more establishments were in states with a middle level. Different mechanisms may drive the findings for high and low unionization rates. There is some evidence that businesses seeking to cut costs seek out areas with pro-business policies, low regulatory enforcement, and unionization (Harrington & Warf, 1995; HRW, 2005), suggesting potentially that management was repelled from areas with high percentages of unionization. Consistent with this, meat establishments were also more likely to be in states with right to work policies and lower levels of OSHA inspection. The finding of fewer establishments in lowest-unionized states might reflect lack of historical presence in those areas. None of the policy/culture/values risk factors significantly affected *injury/illness rates,* though their state aggregation meant numbers were small. A body of literature documents the protective effects of unions for worker safety and health (Gray et al., 1998; HRW, 2005; Litwin, 2000; O'Neill, 2002; Reilly et al., 1995; Weil, 1997; Weil, 1991). However, if unions are strongest in areas with the most industry hazard and if unions increase reporting, this relationship could be obscured.

Examination of residuals from the regression describing establishment location suggests that there may have been missing covariates, including, perhaps, locations of animal growers and of distribution centers, population centers, locations of corporate headquarters, transit routes, and area location incentives (Roe et al., 2002).

The control variables are also worthy of discussion. “Number of years an establishment was surveyed” was included due to OSHA’s policy of not resurveying establishments reporting low rates in subsequent years (in some years). (OSHA Directorate of Compliance Programs, 1997, 1998, 1999, 2000, 2001). The consistent statistically significant association with high rates confirms the expected finding that plants surveyed in more years should have higher rates. Variables representing “number of establishments” in each of the three industries were used to account for the different expected area rates based on industry concentration but, surprisingly, were not statistically significant.

In sum, many aspects of the conceptual framework held. Several findings ran counter to the expected direction of the conceptual framework, such as the finding that high rate establishments were in relatively Caucasian counties and, for poultry, in areas of low change in immigrants. Yet, the fact that these risk factors were significantly associated with the outcome – *in any direction* – provides support for the concept that these area-level social risk factors are important in understanding the determinants of location and risk.

2. Underreporting: In response to questions about underreporting in this industry raised by the United States Government Accountability Office (2005) and others, we examined the industry and geographic distribution of establishments reporting "zero" injuries and illnesses, and of those reporting the steepest drops in injury/illness from year to year. The analyses suggest that reports of unusually low rates were more common in meatpacking than in the other two industries and that such reports may have differed geographically. Further, the findings on race and immigrant status (in the poultry industry) suggest the possibility that underreporting plays a role. OSHA notes that some of the "zero" rates were in surveys returned by corporate headquarters. That or another data issue may explain the high concentration of zeroes in Indiana in 1999.

Establishments with LWDII drops of over six events per 100 workers from year to year were also widespread. Establishment injury/illness rates tend to be highly consistent over time (Hunt, 1993). While "regression to the mean" is an expected phenomenon following high rates, the size and frequency of the observed drops strongly suggests the need for further investigation. The maps of these underreporting indicators suggest the *possibility* of geographic patterning, but further investigation is needed. Collection of unique establishment identifiers by OSHA would enable additional and more detailed longitudinal studies with better reliability.

B. STRENGTHS AND LIMITATIONS

The three chief data limitations are: reporting bias due to the enforcement usage; data quality; and the survey method in which establishments reporting low rates were less likely to be surveyed subsequently (making their counties appear to have worse outcomes than they had). In addition, at least some errors in hand-coding the multi-year database are likely due to data ambiguities. For risk factor variables, their appropriate time frames of effect as “exposures” are a question. The methods of mapping are problematic in that patterns apparent to the eye may not reflect reality, and choices made in map display can affect findings. On maps of the U.S., the disproportionate visual impact of large Western counties can be especially misleading. Use of the county scale is likely to have obscured key relationships. (However, smaller scale analyses would lose some utility and meaning, particularly due to commuting and policy patterns.) Finally, in interpreting the findings, the ecologic fallacy of inferring individual risk from area risk must be avoided.

On the strengths side, the ODI database is a rarity in providing establishment-level data reflecting injuries and illnesses; no articles are known to have analyzed it in the peer reviewed literature. The effort to create establishment identifiers further enriches the data, making it possible to pool across years. Sensitivity analyses supported this effort. This research is built on a new conceptual framework and an interdisciplinary literature review. It uses multiple forms of analysis to address different aspects of the research question. It is the first known geographic examination of occupational injury and illness in meat processing and one of the few geographic analyses of occupational injury/illness overall. Accordingly, the research is useful both for understanding the meat industry and as a case study of a broader phenomenon. This study contributes to the practice of public health tracking, providing an example of how the ODI or other geo-referenced data might be used for surveillance in occupational injury/illness. Officials can use the maps (and additional maps available from the authors) to assist with tracking and intervention design and targeting. Maps provide a visual tool for discussing hazards with the public and policymakers. Further, geographic study can both help identify areas of concern in need of prevention programs and provide information about common local conditions in meat industry areas to help guide program design.

C. CONCLUSIONS

The meat products industry has experienced large reductions in reported injury/illness rates since the early 1990s. However, some establishments continue to have rates far above industry means. While varying establishment injury/illness rates may reflect different levels of inherent hazard, it is also possible that if high rate plants implemented process changes already in use elsewhere, they might see substantial rate reductions.

With today's increased concerns about avian flu and food safety threats, occupational injury and illness must be considered in strategic thinking about protecting the general public's health. Establishments that cannot protect their workers from injury and illness may be less likely to have systems in place to protect workers and the public from other threats.

This paper has demonstrated that social risk factors are associated with establishment locations and that different area risk factors predict high rates. If the business and worker incentives and capacity for safety were altered – in high risk areas and throughout the industry – rates might decline. This can be done with increased enforcement, communicating with the public about social responsibility in purchasing, expanded compliance assistance, increased unionization, and education of workers about safety and their rights. The possibility remains that area risk factors are associated with underreporting. Given that OSHA depends on the ODI to target enforcement, it has the responsibility to invest in a more aggressive program of audits and recordkeeping citations to assure accuracy of its database.

In the ODI sample, over seven percent of meat industry workers *reported* work-related occupational injuries and illnesses each year, more than double the national average in most years. Injured workers must cope with pain, logistical needs, and the struggle for reimbursement (Boden et al., 2001; Dembe, 2001; Dorman, 2000; HRW, 2005; GAO, 1998). In the meat processing industry, injury and illness often result in job loss and a loss of earning power in already low income families (HRW, 2005). The estimated 1/4 of meat processing workers who are immigrants often do not receive workers' compensation and frequently do not know their rights (Escobar, 1999). Local, state, and federal governments commonly pay medical and social welfare costs for injured and ill workers. Relatively little burden is experienced by employers.

Improved targeting of enforcement and intervention and efforts to shift business incentives may ultimately help incorporate more of the worker and social costs of meat production into the price of meat.

6

CONCLUSION

This dissertation began by noting that workplace injuries and illnesses are often attributed to accidental factors – a worker being "in the wrong place." I asked: "what if a place is "wrong" for a lot of workers?" The four papers in this thesis have demonstrated that indeed, such "wrong" places exist. I have examined the geographic variation in occupational injury/illness from different angles: the literature review, demonstration of geographic surveillance methods using the Occupational Safety and Health Administration (OSHA) Data Initiative (ODI), examination of social risk factors for county-level high injury/illness rates in the ODI, and an industry case study focused on meat processing. The results point to three clear conclusions:

1. There is suggestive evidence of substantial, biased underreporting in the ODI.
2. There is geographic variation in occupational injury/illness rates, and it is associated with social risk factors.
3. Geographic Information Systems (GIS) and statistical methods for spatial data provide an important and underused opportunity for improving prevention of

occupational injuries and illnesses through research, communication with the public and policymakers, and intervention targeting.

Each of these is discussed in more detail below, followed by a discussion of social risk factors and intervention, policy recommendations, and concluding remarks.

1. Underreporting in the ODI: Evidence of ODI underreporting comes from several lines of investigation. First is the unexpected association of high Lost Workday Injury/ Illness rates (LWDII) with areas with markers of social privilege, equality, and pro-worker balance of power. This association was seen in maps, bivariate and controlled analyses, for all industries and in meat processing. While explanations can be devised for each finding, biased underreporting is unifying explanation.

Another suggestive piece of evidence comes from comparing maps of nonfatal occupational injury and illness in the ODI with those for occupational fatality and industry hazard. ODI rates – and often other nonfatal injury/illness rates – have a different state distribution from fatal injury rates, which are considered to be more accurate due to the relative difficulty of underreporting fatalities. As would be expected under the conceptual framework, fatal injury rates are higher in the South (NIOSH, 2004). The state map of our Index of Area Industry Hazard, (reflecting the expected county injury/illness rate based on industry)[11] is remarkably similar to that for occupational fatalities. The Index is expected to be less tainted by geographically-biased underreporting than the ODI. While recognizing that fatal and nonfatal events are "apples and oranges," it is possible that the fatality and industry hazard maps suggest something closer to a "true" area risk. If so, then the difference between them and the ODI-reported

[11] More specifically, the Index of Area Industry Hazard assigns each establishment its industry mean as reported to the Bureau of Labor Statistics' Survey of Occupational Injuries and Illnesses. These are then averaged across establishments in a county to calculate "expected" county rates.

rates may partly reflect area social factors and their impact on geographically biased underreporting.(The difference also partly reflects separate databases and other contrasts between fatal and nonfatal events.(Herbert & Landrigan, 2000; Robinson, 1988))

Finally, in the meat case study we were particularly interested in underreporting following the U.S. Government Accountability Office's (GAO, 2005) report raising concerns about its effects in this industry. We created two measures of potential underreporting, both finding suggestive evidence. One was based on the meat industry's high injury/illness rates overall: we mapped establishments reporting the unexpected low rate of zero. In the other, we mapped establishments reporting a drop of at least six points in injury/illness rate from any year to the next. Mapping demonstrated that both measures were widespread; it was difficult to visually discern patterns other than one probably based on a data issue. We compared these measures by industry. Both were most frequent in meatpacking, which has the highest mean injury/illness rates, and least frequent in poultry, with the lowest mean rates.(Table 24) One interpretation is that the higher an establishment's injury/illness rate, the greater the underreporting incentive.

Table 244: Indicators of underreporting: Reports of LWDII rates of "Zero"; Year-to-year LWDII drops of at least six points , By industry

	Meat	Sausage	Poultry
% Reporting "Zero"	15.0	8.03	6.46
% Dropping >=6 in a year	15.6	13.7	12.5
Mean LWDII	9.78	8.15	7.91

None of these findings provides confirmation that biased underreporting exists in the ODI. However, together they provide suggestive evidence that requires further investigation. OSHA annually audited 250 establishments during the study period. The 1997 and 1998 audits found that 19 and 21 percent, respectively, of cases had "major recording errors." Many more errors were in the direction of underreporting that the reverse (OSHA, 2001). Today OSHA's auditing program focuses on 200 randomly selected firms reporting low rates in high rate industries, and the agency also separately has a program of record-keeping audits (GAO, 2005).

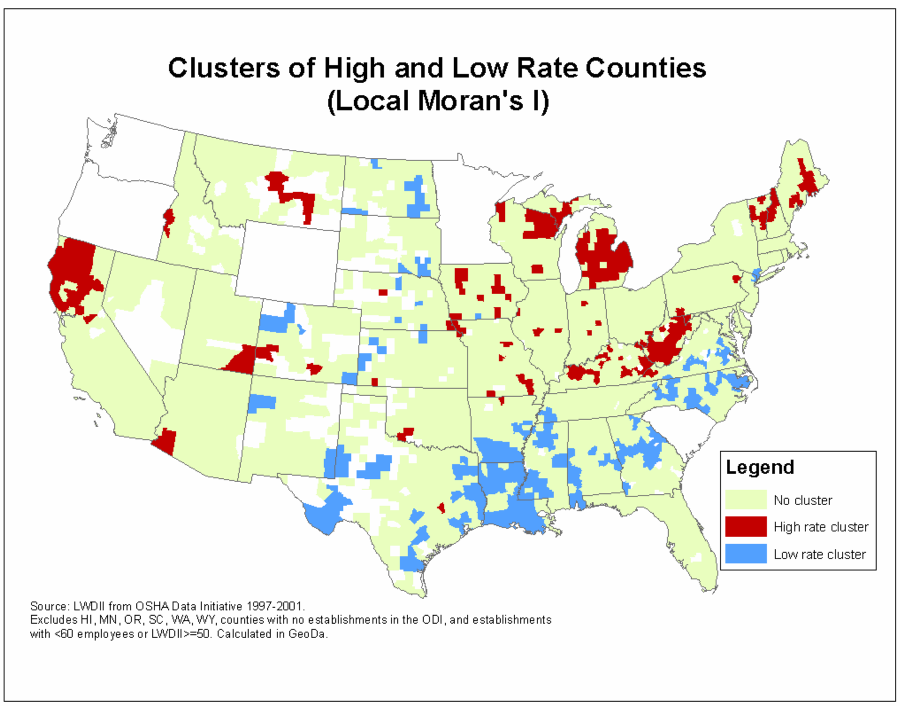
This discussion of underreporting should be placed in context. There is evidence that national surveillance databases miss as many as 2/3 of all nonfatal occupational injuries, including due to underreporting. Underreporting is expected to be worse for illness, due to latency and challenges in proving work-relatedness (Azaroff et al., 2002; Azaroff et al, 2004; Conway & Svenson 1998; Leigh & Robbins 2004; Leigh et al., 2004; Pollack & Keimig, 1987; Pransky et al., 1999; Rosenman et al., 2006; Smith, 2003; Smith et al., 2005). Beyond groups excluded from surveys, reasons for businesses not reporting include concerns about enforcement, workers' compensation premium increases, and denial of government contracts based on high rates. At the worker level, nonreporting may be related to: employer incentive or discipline programs based on injury rates; stigma; lack of access to health care; health care providers or workers not attributing events to work causes; not knowing that they ought to report or how to do so; perceived hassle of reporting; and concerns about legal or job consequences (AFL-CIO, 2006).

Underreporting in the ODI raises both justice and public health concerns. Justice, because it may occur differentially based on establishment injury/illness rates and area social factors; and public health, because OSHA uses the ODI for targeting enforcement, and underreporting may lead to missed opportunities for intervention.

2. There is geographic variation in occupational injury/illness rates, and it is associated with social risk factors. The surveillance paper showed how reported nonfatal occupational injury/illness rates vary by geography in the ODI, even after adjusting for industry hazard. County mean rates in my sample (excluding establishments reporting over 50) ranged from 0 to 25.2 injuries/illnesses per 100 workers.

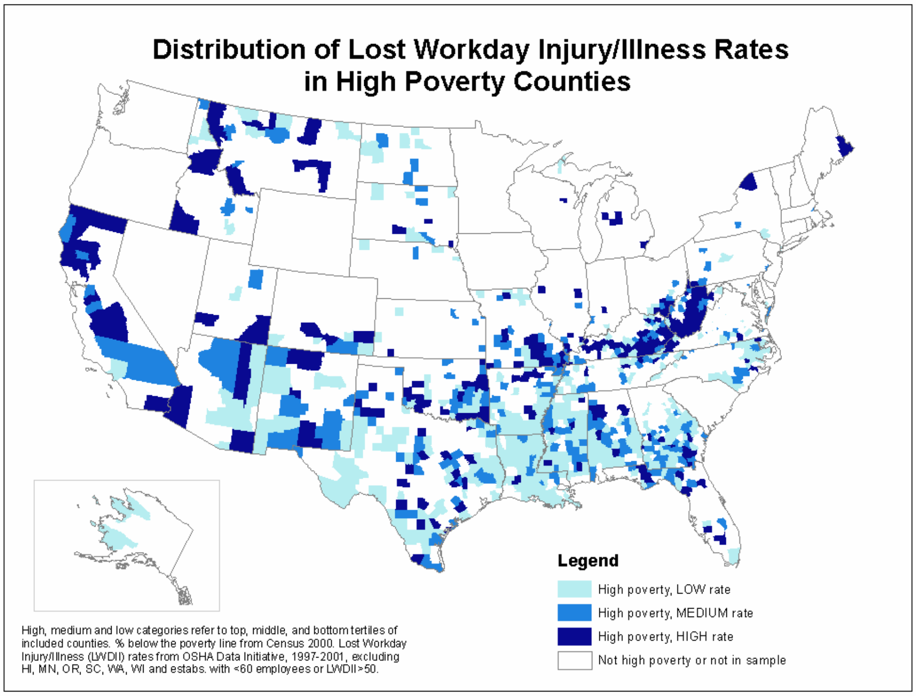
The top five states with the highest mean LWDII rates were Vermont (9.77), West Virginia (9.76), Michigan (9.67), Maine (9.54) and Kentucky (8.99). Maps also highlighted Northern California and Wisconsin. These rates compare with an overall sample mean of 7.22. The states (and district) with the lowest reported rates were Louisiana (4.98), the District of Columbia (5.16), North Carolina (5.45), Delaware (5.67), and Georgia (5.77). Maps showed low rates throughout the South. Map 27 shows geographic clusters of counties with low and high rates (using the Local Moran's I method – details in (Anselin, 2003; Neff, Burke, & Curriero, 2006)). High rate clusters are shown in red; low rate clusters in blue.

Map 27: Clusters of high and low rates in the OSHA Data Initiative 1997-2001 (Local Moran's I)



Both the two-variable maps and the multivariate analyses showed that occupational injury/illness rates varied in concert with social risk factors. For example, Map 28 shows the patterned association between poverty and injury/illness.

Map 28: Distribution of Lost Workday Injury/Illness Rates in High Poverty Counties



Tables 25 and 26 summarize regression results from the cross-industry and meat analyses. These results are not directly comparable, but they do show similar coefficient directions for risk factors that were in both models. The cross-industry multilevel model found that high rates of occupational injury and illness were associated with poverty, white race, unionization, strong safety net, and industry hazard, as well as the controls, non-Southern states and non-“rural farm” areas. In the meat case study, meat establishment locations – which were themselves viewed as conditions of potential concern for worker health – were in counties that were relatively high percent African American, non-college educated, had longterm job gain, were urban, and were in states with medium union membership, anti-union policy, and slightly reduced OSHA inspections. By contrast, meat processing establishments with high occupational injury/illness rates were in counties that were relatively white, low per capita income, high school educated, and had longterm job loss. In meatpacking, high rates were associated with increases in immigrants, whereas in poultry, low increases or decreases in immigrants were predictive.

Table 25: Summary of regression results from all-industry analyses. Outcome: mean county LWDII.

Risk Factor	Coefficient	95% Confidence Interval
Demographics		
Percent African American	-0.03	(-0.04, -0.02)
Local Economy/Socioeconomic Status		
% < Poverty	0.02	(0.01, 0.04)
Policy, Culture, Values		
State percent unionized 1999	0.08	(0.02, 0.15)
State rank: ratio of unemployment benefits to average weekly wage	-0.02	(-0.04, -0.002)
Industry hazard		
Industry hazard: "Expected" county lost workday injury/illness rate if every ODI establishment had its industry mean	0.79	(0.71, 0.89)
Control		
% Rural, Farm	-0.11	(-0.13, -0.08)
South	-0.96	(-1.77, -0.16)

Table 26: Summary of regression results from the meat industry. Outcome: mean county LWDII.

Risk Factor	Coefficient	95% Confidence Interval
Demographics		
Percent African American	-0.04	(-0.07, -0.01)
% < High School	-0.12	(-0.21, -0.03)
Change % Foreign-Born	0.13	(-0.07, 0.33)
Local Economy/Socioeconomic Status		
Per Capita Income	-0.0003	(-0.0004, -0.0001)
Increased Unemployment 1975-1996	0.23	(0.1, 0.4)
Control		
Mean number of years in which county establishments were surveyed	1.70	(1.3, 2.1)

Overall, these analyses supported the project conceptual framework that identified four categories of geographic risk factors: demographics; local economy/socioeconomic status; policy/culture/values; and industry hazard. Every relevant analysis found risk factors in each of these categories. At times, the findings ran counter to the expected direction of the conceptual framework, such as the finding that high rate establishments were in relatively Caucasian counties (one of the strongest and the most consistent associations seen). Yet, the fact that these risk factors were significantly associated with the outcome – *in any direction* – provides support for the concept that area-level social risk factors are important in understanding determinants of location and risk.

Among interpretations for the association of high rates with markers of social privilege, two stand out. First, underreporting, as discussed above. Second, it is possible that the association is partly explained by differences in industry hazard. Higher hazard industries in the sample may pay better and be more unionized and in areas with union-influenced social policy. Further, due to structural discrimination, establishments in these industries may be located in areas that offer more access to white workers than African Americans. While this "compensating wage hypothesis" has been effectively challenged overall, studies suggest it may apply in some cases, such as to white and unionized workers (Shrader-Frechette, 2002). I conjecture that larger size workplaces and those with the honesty to report high rates might also be more likely than others to offer this "hazard pay" due to relatively high worker access to information about hazards. In theory, controlling for area industry hazard should have adjusted for many of these industry differences, and also, analyses separating by high hazard industry did not find evidence that this phenomenon was transpiring, but it remains a possibility worthy of further follow-up.

Replication of this study in other databases such as state workers' compensation is warranted. Additional in depth studies on industries, such as the companion meat industry case study, will also be beneficial.

3. Geographic methods provide an important and underused opportunity for improving prevention of occupational injuries and illnesses through research, communication with the public and policymakers, and intervention targeting.

The literature review identifies many proven benefits of geographic analysis for occupational injury/illness prevention. Findings from cited papers have led to critical new insights in cancer and respiratory disease causation, sparked effective programmatic intervention including improved targeting, clarified and visually dramatized information for policymakers and the public, supported the development of theory, and built understanding of social and other contributions to multiple types of occupational injury and illness. For example, research studies based on the Cancer Atlases have led to insights including the associations between: smelter worker arsenic exposure and lung cancer; shipyard worker asbestos exposure and lung cancer; furniture workers and nasal cavity cancers; and truck drivers and bladder cancer (Devesa et al., 1999). An example of another type of benefit, policy impact, comes from NIOSH’s 1993 chartbook on National Traumatic Occupational Fatality data. Its maps showing Alaska’s fatality rate to be far above the U.S. rate led to a multiagency collaborative on the issue – today considered a major success story for the field (NIOSH, 1993; Smith, 2001).

Although the literature review documents a track record of benefit, the occupational injury/illness field still lags behind other fields in taking advantage of the recent explosion of geographic tools and methods. Despite the consistent finding in geographic research that smaller aggregations improve the ability to detect associations, few publications in the past decade looked at aggregations below the state level. Key occupational injury/illness surveillance databases provide only state level data. Very few articles use spatial statistics and few evaluate more than the most basic geographic risk factors. No identified article articulated a broad theory or conceptual framework for why occupational injury/illness risk would vary geographically.

The Council of State and Territorial Epidemiologists (CSTE) has developed and piloted an impressive set of occupational safety and health indicators for surveillance. Unfortunately, no geographic ones were included, potentially because the committee viewed geography as a second level issue or because they did not believe geographic data would be adequately available.

The field is wide open for additional research and surveillance work, such as developing indicators, using spatial statistics tools, advancing theory, examining and displaying data at aggregations below the state level, and addressing questions specific to industries and occupations, as well as those relevant to broader groups. The literature review paper presents a listing of occupational injury/illness databases that might be applicable to geographic data analysis and provides a list of research questions.

This thesis suggests an approach, a method, and a theory-based conceptual framework for bringing geographic analysis to the task of occupational injury/illness surveillance, research, and ultimately to prevention activities. It may be the first broad-based national examination of geographic predictors of occupational injury and illness, as well as the first to describe the ODI database in the peer reviewed literature. It brings multilevel regression and spatial statistics tools to a field that has scarcely used them. The study takes a broad look at potential predictors, starting with a theory-based database of

90 county and state-level potential explanatory variables. Despite the many possibilities, the final model included many of the top key variables of interest based on theory and literature, such as poverty, race, unionization, and industry hazard. Sensitivity analyses increase confidence in the findings. This dissertation also provides the first known geographic examination of occupational injury and illness in meat processing.

The ODI database is a rarity in providing establishment-level data reflecting injuries and illnesses; the effort to create establishment identifiers for the meat paper further enriches the data, making it possible to pool across years. The quantitative analyses support the project of mapping for surveillance by demonstrating that the observed geographic variations are more than random coincidence, more than the distribution of population, more than the distribution of industry hazard. They demonstrate that the geographic distribution reflects associations with social factors as well.

On Social Risk Factors

Some might question the benefits of research and surveillance methods demonstrating associations between risk and social factors, because the social factors are often so seemingly intractable. Yet, we *can* change the following.

a) Change area incentives and capacity for prevention. Policy changes such as the living wage, increased minimum wage, and universal health insurance can increase worker bargaining power and thus their incentives and capacity to avoid injury/illness. These are all mainstream, feasible changes being explored or used by large jurisdictions including the federal government. Similarly, area business incentives and capacity for prevention can be changed, for instance, through increased enforcement and compliance assistance, new standard-setting and legal action, prosecution of responsible executives or managers, increased enforcement of employment law, social marketing, tax breaks, closer connection between workers' compensation premiums and costs or prevention activities, and local publicity about the cost savings from safety interventions (Dorman, 2000).

Internationally, policy changes related to trade, debt relief, and responsibilities of manufacturers producing (or recycling) for the U.S. market can also have substantial impact on area risk factors and employer incentives and capacity for safety, while targeting improved international aid, worker training, and support for labor movements can improve worker incentives and capacity.

b) Change the way we intervene, to take account of area factors that may influence intervention success. For example, using geographic tools can help document a need for Spanish speaking intervention staff or low-literacy approaches in particular areas. While local program staff may already be aware of such needs, geographic tools help them communicate about it to policymakers and help determine the extent and level of needs. For many such issues, evidence about overall community needs is likely to be more persuasive than analyses of individual risk factors of injured or ill workers.

c) Change the way we target. OSHA currently uses Local Emphasis Programs, intervention and enforcement efforts focused on industries or hazards in particular regions or sub-regions of the country. However, given that OSHA does not currently analyze data geographically, it is not clear how these are selected. Yet, intervention resources are tiny. For example, in FY 2005 federal and state OSHAs together had 2,117 inspectors for eight million workplaces (AFL-CIO, 2006). Geographic targeting, such as focusing on the top-injury/illness counties with at least 30 ODI establishments, can be an efficient use of funds compared to other high-hazard targeting schemes, because of the reduced travel time between establishments and increased staff familiarity with areas. Targeting counties rather than states would save money, both due to the ability to focus more directly on high rate areas and due to their smaller size overall. Further, geographic targeting of enforcement creates a strong incentive for compliance within targeted counties because of the increased likelihood of inspection. Geographic targeting of approaches like OSHA consultation or workers' compensation loss control assistance to employers in improving safety can contribute to an area ethic in which protecting workers is seen as an important corporate responsibility. Geographic targeting also offers the opportunity for local strategies with cross-industry impact, such as worker training in workplace rights and safety, management training related to safety and health programs, and media training.

Beyond targeting areas with the highest rates, targeting schemes might also jointly address industry and area hazards. Finally, targeting schemes could come from a "justice" framework (Morello-Frosch et al., 2002), under which efforts would be directed to areas suffering from both social injustices and high injury/illness rates in order to use limited prevention resources to assist such doubly-harmed areas.

I conclude this dissertation with a short listing of policy recommendations based on the findings.

RECOMMENDATIONS:

1. Underreporting in the ODI

- OSHA should significantly expand its use of audits and recordkeeping inspections to improve ODI accuracy, and should seek additional strategies for addressing underreporting.
- To assure accountability to taxpayers, an outside agency such as the National Research Council, Institute of Medicine, or Government Accountability Office should evaluate OSHA's effectiveness and make further recommendations.

2. Surveillance

- Funds should be made available to increase geographic surveillance and data collection, including through NIOSH's surveillance program and an expansion of Environmental Public Health Tracking Network funds.
- The Council on State and Territorial Epidemiologists should develop county-level geographic occupational injury/illness indicators
- NIOSH and others who maintain occupational safety and health data should expand collection of location information in databases to work towards the *Healthy People 2010* goal (23-3) (U.S. Department of Health and Human Services, 2000).
- Those performing surveillance should seek to present county-level data where possible, and to display information visually using maps and other graphics. They should consider using spatial statistics such as the global Moran's I and Local Indicators of Spatial Autocorrelation to support their maps.
- Geographic occupational safety and health tracking should be integrated with

tracking of environmental health and injury, and with other surveillance efforts.

3. Research

- NIOSH and other funders should encourage expansion of geographic research both within industry sectors or disease groups and as a cross-cutting issue.
- Occupational safety and health researchers should take advantage of databases with location data and should expand research in this area. Research needs and databases for exploration are noted in the literature review.
- Where appropriate, researchers should use methods that address spatial autocorrelation (the fact that nearby areas are often more similar than those farther away) through methods including spatial statistics and multilevel models.
- There is need to further develop theory regarding how area-level risk factors affect worker outcomes, including further evaluating the conceptual framework in this dissertation.

4. Targeting

- OSHA or a state should implement a pilot project to implement and evaluate geographically targeted intervention.

5. Changing business and worker incentives and capacity for prevention

- Policy changes that alter the worker/business balance of power in the direction of increased incentives and capacity for prevention – both in the U.S. and internationally - should be supported.

Program evaluations of these changes can help in evaluating this project's conceptual framework.

In the U.S., there are an estimated 55,000 annual occupational deaths, making these events together the country's eighth leading cause of death. Further, there were 4.26 million reported nonfatal occupational injuries/illnesses in 2004 (Steenland, et al, 2003, U.S. Dept. of Labor, Bureau of Labor Statistics, 2006). Costs are estimated to be in the range of those for cancer and heart disease (Leigh et al., 1997). In many developing countries, the burden is far worse. The International Labour Organization states that internationally, there were 2.2 million reported worker deaths in 2005, and that this figure is probably "vastly underestimated" (International Labour Organization, 2005). International exposure disparities are rooted in area differences including policy choices about occupational injury and illness, trade, taxation, corporate obligations, debt and equity. "The wrong place" is a real place where too many workers spend their days (and/or nights.)

If a geographic targeting scheme succeeded in bringing establishments in high rate areas closer to their industry means, it would impact not only worker health but also costs. Waehrer et al (2004, calculated) document more than a three-fold variation in state occupational injury/illness costs per non-governmental employee. While 73 percent of this variation could be explained by industry, that means 27 percent could be explained by non-industrial factors, including, potentially, factors that can be addressed through the above-discussed interventions.

Budgets for occupational injury/illness prevention are miniscule and face constant political challenges. OSHA must implement optimal targeting if its 2,117 inspectors can hope to make a dent in risk at eight million workplaces.

By helping visualize the problem, geographic tools can help in advocacy to increase the overall level of resources devoted to occupational injury/illness prevention. More directly, they can provide data to improve intervention targeting efforts, identify area risk factors, and make the case for targeting resources to prevention in hard-hit areas so that one day, “the wrong place” can be transformed into just “a place.”

REFERENCES

Adekoya, N., & Pratt, S. G. (2001). *Fatal unintentional farm injuries among persons less than 20 years of age in the United States: Geographic profiles*. DHHS (NIOSH) Publication No. 2001-131. DHHS, CDC, NIOSH.

AFL-CIO. (2006). *Death on the Job: The Toll of Neglect.* (15th ed.). Washington, DC: AFL-CIO. Retrieved June 26, 2006, from http://www.aflcio.org/issues/safety/memorial/doj_2006.cfm

Anderton, D. L., & Egan, K. L. (2000). *Industrial and residential segregation: Employment opportunities and environmental burdens in metropolitan areas (Working Paper).* Social and Demographic Research Institute, University of Massachusetts - Amherst. Retrieved July 5, 2006, from http://www.umass.edu/sadri/pdf/WP/WP3%20-%20Anderton%20Egan.pdf

Anselin, L. (2003). *GeoDa 0.9 User's Guide.* Spatial Analysis Laboratory, University of Illinos, Urbana-Champaign, IL. Retrieved from https://geoda.uiuc.edu/downloadin.php

Anselin, L., & Koschinsky, J. (2006). *GeoDa Help System Glossary of terms (Draft).* Retrieved September 8, 2006 from https://www.geoda.uiuc.edu/support/help/glossary.html

Antao, V. C., Petsonk, E. L., Sokolow, L. Z., Wolfe, A. L., Pinheiro, G. A., & Hale, J. M., et al. (2005). Rapidly progressive coal workers' pneumoconiosis in the United States: Geographic clustering and other factors. *Occup Environ Med, 62*, 670-674.

Azaroff, L. S., Lax, M. B., Levenstein, C., & Wegman, D. H. (2004). Wounding the messenger: The new economy makes occupational health indicators too good to be true. *International Journal of Health Services, 34*(2), 271-303.

Azaroff, L. S., Levenstein, C., & Wegman, D. H. (2002). Occupational injury and illness surveillance: conceptual filters explain underreporting. *Am J Public Health, 92*(9), 1421-9.

Baker, S. P., O'Neill, B., Ginsburg, M. J., & Li, G. (1992). *The Injury Fact Book* (2nd ed.). New York: Oxford University Press.

Barstow, D. (2003, December 22). U.S. rarely seeks charges for deaths in workplace. *The New York Times,* p. A1.

Blot, W. J., Brinton, L. A., Fraumeni, J. F., & Stone, B. J. (1977). Cancer mortality in U.S. counties with petroleum industries. *Science*, *198*, 51-53.

Blot, W. J., & Fraumeni, J. F. (1977). Geographic patterns of oral cancer in the United States: Etiologic implications. *J Chron Dis, 30*(745), 757.

Blot, W. J., & Fraumeni, J. F. (1976). Geographic patterns of lung cancer: Industrial correlations. *American Journal of Epidemiology, 103*(6), 539-550.

Boden, L. I., Biddle, E. A., & Spieler, E. (2001). Social and economic impacts of workplace illness and injury: Current and future directions for research. *American Journal of Industrial Medicine, 40,* 398-402.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199-231.

Brinton, L. A., Stone, B. J., Blot, W. J., & Fraumeni, J. F. (1976). Nasal cancer in U.S. furniture industry counties (letter). *Lancet*, *2*(628)

Bureau of Labor Statistics. (2005). *Table 1. Incidence rates of nonfatal occupational injuries and illnesses by industry and case types, 2004.* Retrieved June 13, 2006 from http://www.bls.gov/iif/oshwc/osh/os/ostb1487.txt

Bureau of Labor Statistics. (2006). *Occupational injuries and illnesses: Industry data (1989-2001).* Retrieved August 17, 2006 from http://data.bls.gov/PDQ/outside.jsp?survey=sh

Cattledge, G. H., Schneiderman, A., Stanevich, R., Hendricks, S., & Greenwood, J. (1996). Nonfatal occupational fall injuries in the West Virginia construction industry. *Accident Analysis and Prevention, 28*(5), 655-663.

Conway, H., & Svenson, J. (1998). Occupational injury and illness rates, 1992-96: Why they fell. *Monthly Labor Review,* 36-58.

Council of State and Territorial Epidemiologists. (1999). *The Role of the States in a Nationwide Comprehensive Surveillance System for Work-Related Diseases, Injuries and Hazards.* Retrieved August 12, 2005, from http://www.cste.org/Occupational%20health/NIOSH.pdf

Council of State and Territorial Epidemiologists in collaboration with NIOSH. (2004). *Occupational Health Indicators: A Guide for Tracking Occupational Health*

Conditions and their Determinants. Retrieved August 12, 2005, from http://www.cste.org/pdffiles/Revised%20Indicators_6.24.04.pdf

Cressie, N. (1991). *Statistics for Spatial Data*. New York: John Wiley & Sons.

Cromley, E. K., & McLafferty, S. L. (2002). *GIS and Public Health.* New York: The Guilford Press.

Davis, J. (2000). Social Stress and Mental Health in the Southern Black Belt. *Sociological Spectrum*, *20*, 465-494.

Dekker, S. W. A., & Schaufeli, W. B. (1995). The effects of job insecurity on psychological health and withdrawal: A longitudinal study. *Australian Psychologist, 30*(1), 57-63.

Dembe, A. E. (2001). The social consequences of occupational injuries and illnesses. *American Journal of Industrial Medicine, 40*(4), 403-417.

Dembe, A. E., Erickson, J. B., & Delbos, R. (2004). Predictors of work-related injuries and illnesses: National survey findings. *Journal of Occupational and Environmental Hygiene*, 542-550.

Devesa, S. S., Grauman, D. J., Blot, W. J., Pennello, G., Hoover, R. N., & Fraumeni, J. F., Jr. (1999). *Atlas of Cancer Mortality in the United States, 1950-94.* No. (NIH) 99-4564). Washington, DC: US Govt Print Off.

Dorman, P. (2000). *Three preliminary papers on the economics of occupational safety and health. (International Labor Organization SafeWork Papers.)* Geneva:

International Labour Organization. Retrieved from: http://www.ilo.org/public/english/protection/safework/papers/ecoanal/index.htm

Drabenstott, M. (1999). Where have all the packing plants gone?: The new meat geography in rural America. *Economic Review (Federal Reserve Bank of Kansas City),* (3rd Quarter), 65-82.

Drabenstott, M. (1998). This little piggy went to market: Will the new pork industry call the heartland home? *Economic Review (Federal Reserve Bank of Kansas City),* (Q3), 79-97.

D'Souza, R. M., Strazdins, L., Lim, L. L., Broom, D. H., & Rodgers, B. (2003). Work and health in a contemporary society: demands, control, and insecurity. *J Epidemiol Community Health, 57*, 849-854.

Edwards, B., & Ladd, A. (2001). Race, poverty, political capacity and the spatial distribution of swine waste in North Carolina, 1982-1997. *NC Geogr, 9*, 59-70.

ERG. (2004). *Evaluation of OSHA's Impact on Workplace Injuries and Illnesses in Manufacturing Using Establishment-Specific Targeting of Interventions: Programmed Inspections and High Hazard Notification Letters.* Lexington, MA.

Escobar, G. (1999, November 30). Workers answer to multiple names: Culture of fraud flourishes in Delmarva chicken towns. *The Washington Post*, pp. A01.

ESRI. (2005). *ArcGIS Version 9.0*. Redlands, CA.

Franklin, S. (2005, January 17). "Jobs that Americans won't do" filled by desperate migrants. *Chicago Tribune.*

Frazier, T. M., & Sundin, D. S. (1986). Industrial demographics and population at risk for silica exposures. In D. F. Goldsmith, D. M. Winn & C. M. Shy (Eds.), *Silica, Silicosis and Cancer* (pp. 3-9). Philadelphia: Praeger Scientific.

Frazier, T. M., Lalich, N. R., & Pedersen, D. H. (1983). Uses of computer-generated maps in occupational hazard and mortality surveillance. *Scand J Work Environ Health, 9*(2 Spec No), 148-54.

Froines, J. R., Dellenbaugh, C. A., & Wegman, D. H. (1986). Occupational health surveillance: A means to identify work-related risks. *American Journal of Public Health, 76*(9), 1089-1096.

Grabowski, J. G., Curriero, F. C., Baker, S. P., & Li, G. (2002). Exploratory spatial analysis of pilot fatality rates in general aviation crashes using Geographic Information Systems. *American Journal of Epidemiology, 155*(5), 398-405.

Gray, G. R., Myers, D. W., & Myers, P. S. (1998). Collective bargaining agreements: Safety and health provisions. *Monthly Labor Review,* (May), 13-35.

Harrington, J. W., & Warf, B. (1995). *Industrial Location: Principles, Practice, and Policy* New York: Routledge.

Herbert, R., & Landrigan, P. J. (2000). Work-related death: a continuing epidemic. *American Journal of Public Health, 90,* 541-545.

Hillemeier, M. M., Lynch, J., Harper, S., & Casper, M. (2003). Measurement issues in social determinants: Measuring contextual characteristics for community health. *Health Services Research*, *38*(6p2), 1645-1717.

Hoover, R., & Fraumeni, J. F. (1975). Cancer mortality in U.S. counties with chemical industries. *Environmental Research, 9*, 196-207.

Hoover, R., Mason, T. J., McKay, F. W., & Fraumeni, J. F. (1975). Cancer by county: New resource for etiologic clues. *Science, 189*, 1005-1007.

Horton, D. K., Berkowitz, Z., & Kaye, W. E. (2004). Surveillance of hazardous materials events in 17 states, 1993–2001: A report from the Hazardous Substances Emergency Events Surveillance (HSEES) system. *American Journal of Industrial Medicine, 45*, 539-48.

Human Rights Watch. (2005). *Blood, Sweat, and Fear: Workers' Rights in U.S. Meat and Poultry Plants*. Retrieved June 26, 2006, from http://www.hrw.org/reports/2005/usa0105/

Hunt, H. A. (1993). *Analysis of Persistence in Employer Injury Rates.* No. N. 93-002. E. Upjohn Institute for Employment Research.

International Labour Organization. (2005). *Global workplace deaths vastly under-reported, says ILO. (Press Release, September 18.)*. Retrieved July 5, 2006 from http://www.ilo.org/public/english/bureau/inf/pr/2005/36.htm

International Labour Organization. (2004). *Occupational accidents, 2002: The world, according to ILO regions*. Retrieved April 20, 2005, from http://www.ilo.org/public/english/protection/safework/accidis/globest_2002/reg_world.htm

Jemal, A., Grauman, D., & Devesa, S. (2000). Recent geographic patterns of lung cancer and mesothelioma mortality rates in 49 shipyard counties in the United States, 1970-1994. *American Journal of Industrial Medicine, 37*, 512-521.

Joines, J. D., Hertz-Picciotto, I., Carey, T. S., Gesler, W., & Suchindran, C. (2003). A spatial analysis of county-level variation in hospitalization rates for low back problems in North Carolina. *Social Science and Medicine, 56*, 2541-2553.

Kahn, S. (1991). Does employer monopsony power increase industrial accidents?: The case of Kentucky coal mines. *NBER Working Paper # 3897,* (November)

Kim, J. H. (1998). *Atlas of respiratory disease mortality, United States: 1982-1993* No. DHHS (NIOSH) Number 98-157. Division of Respiratory Disease Studies, NIOSH.

Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2003). Race/ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures--The public health disparities geocoding project. *American Journal of Public Health*, *93*(10), 1655-1671.

Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M. J., Subramanian, S. V., & Carson, R. (2003). Choosing area-based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *Journal of Epidemiology and Community Health*, *57*, 186-99.

Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M. J., Subramanian, S. V., & Carson, R. (2002). Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? *American Journal of Epidemiology, 156*, 471-482.

Lange, J. L., Schwartz, D. A., Doebbeling, B. N., Heller, J. M., & Thorne, P. S. (2002). Exposures to the Kuwait oil fires and their association with asthma and bronchitis among Gulf War veterans. *Environ Health Perspect, 110*(11), 1141-6.

Leigh, J. P., Marcin, J. P., & Miller, T. R. (2004). An estimate of the US government's undercount of nonfatal occupational injuries. *Journal of Occupational and Environmental Medicine, 46*(1), 10-18.

Leigh, J. P., Markowitz, S. B., Fahs, M., Shin, C., & Landrigan, P. J. (1997). Occupational injury and illness in the United States - Estimates of costs, morbidity, and mortality. *Archives of Internal Medicine, 157*(14), 1557-1568.

Leigh, J. P., & Robbins, J. A. (2004). Occupational disease and Workers' Compensation: Coverage, costs and consequences. *The Milbank Quarterly, 82*(4)

Levinson, M.E. & Haddon, W. (1965). The area adjusted map: An epidemiologic device. *Public Health Reports, 80*(1), 55-59.

Litwin, A. S. (2000). *Trade unions and industrial injury in Great Britain.* The Centre for Economic Performance. Retrieved August 27, 2004, from http://cep.lse.ac.uk/pubs/download/dp0468.pdf

Loh, K., & Richardson, S. (2004). Foreign-born workers: Trends in fatal occupational injuries, 1996-2001. *Monthly Labor Review, June*, 42-53.

Loomis, D., & Richardson, D. (1998). Race and the risk of fatal injury at work. *American Journal of Public Health, 88*, 40-44.

Loomis, D., Richardson, D. B., Bena, J. F., & Bailer, A. J. (2004). Deindustrialization and the long term decline in fatal occupational injuries. *Occup Environ Med, 61*, 616-621.

Loomis, D., Bena, J. F., & Bailer, A. J. (2003). Diversity of trends in occupational injury mortality in the United States, 1980-96. *Injury Prevention, 9*, 9-14.

Louis, T. A. (2006). *Gaussian Empirical Bayes estimates using the EM algorithm (Code written for R computer software).*

Louis, T. A., & Shen, W. (1999). Innovations in Bayes and Empirical Bayes methods: Estimating parameters, populations and ranks. *Statist. Med., 18,* 2493-2505.

MacDonald, J. M., Ollinger, M. E., Nelson, K. E., & Handy, C. R. (2000). *Consolidation in US meatpacking*. Agricultural Economic Report No. 785. Food and Rural Economics Division, Economic Research Service, US Department of Agriculture.

Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., & Fraumeni, J. F., Jr. (1976). *Atlas of Cancer Mortality among U.S. Non-Whites: 1950-1969.* DHEW Publ. No. (NIH) 76-1204. Washington, DC: US Gov. Printing Office.

Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., & Fraumeni, J. F., Jr. (1975). *Atlas* *of Cancer Mortality for U.S. Counties: 1950-1969.* DHEW Publ. No. (NIH) 75-780. Washington, DC: US Gov. Printing Office.

McGlashan, N. D., Harington, J. S., & Chelkowska, E. (2003). Changes in the geographical and temporal patterns of cancer incidence among black gold miners working in South Africa, 1964-1996. *British Journal of Cancer, 88*, 1361-1369.

McGwin, G., Valent, F., Taylor, A. J., Hower, H. J., Davis, G. G., & Brissie, R. M., et al. (2002). Epidemiology of fatal occupational injuries in Jefferson County, Alabama. *Southern Medical Journal, 95*(11), 1300-1311.

McNabb, S. J., Ratard, R. C., Horan, J. M., & Farley, T. A. (1994). Injuries to international petroleum drilling workers, 1988 to 1990. *J Occup Med., 36*(6), 627-30.

Miller, T. R., & Levy, D. T. (1999). Geographic variation in expenditures for workers' compensation hospitalized claims. *American Journal of Industrial Medicine, 35*, 103-111.

Miller, T. R., & Levy, D. T. (1997). Geographic variation in expenditures for workers' compensation in physician claims. *American Journal of Industrial Medicine, 32*, 27-34.

Morello-Frosch, R., Pastor, M., Jr., & Sadd, J. (2002). Environmental justice and regional inequality in Southern California: Implications for future research. *Env H Perspec, 110*(Supplement 2), 149-153.

Muntaner, C., Lynch, J. W., Hillemeier, M., Lee, J. H., David, R., & Benach, J., et al.

(2002). Economic inequality, working-class power, social capital, and cause-specific mortality in wealthy countries. *International Journal of Health Services, 32*(4), 629-56.

Murray, L. R. (2003). Sick and tired of being sick and tired: Scientific evidence, methods, and research implications for racial and ethnic disparities in occupational health. *American Journal of Public Health*, *93*(2), 221-226.

Neff, R. A. (2006). *All over the map: A review of literature on geography and occupational injury and illness. (Dissertation chapter).* Unpublished PhD, Johns Hopkins Bloomberg School of Public Health.

Neff, R. A., Burke, T. A., & Curriero, F. (2006). *In the wrong place?: Spatial tools for occupational injury/illness surveillance. (Dissertation chapter).* Unpublished PhD, Johns Hopkins Bloomberg School of Public Health.

Neff, R. A., Curriero, F, & Burke, T. A.. (2006). *Social predictors of county occupational injury/illness rates. (Dissertation chapter).* Unpublished PhD, Johns Hopkins Bloomberg School of Public Health.

Neff, R. A., Burke, T. A., & Curriero, F. (2006). *Occupational Injury/Illness in the Meat Processing Industries: A Geographic Cut." (Dissertation chapter).* Unpublished PhD, Johns Hopkins Bloomberg School of Public Health.

New Jersey Department of Health and Senior Services. (2005). *Fatal occupational injuries in New Jersey: Ten year report 1993-2002.* Retrieved April 17, 2006 from http://www.state.nj.us/health/eoh/survweb

National Institute of Occupational Safety and Health (NIOSH). (2006a). Division of Respiratory Disease Studies, Surveillance Branch. *National Occupational Respiratory Mortality Surveillance System (NORMS).* Retrieved April 23, 2006 from http://webappa.cdc.gov/ords/norms.html

NIOSH. (2006b). *NIOSH-TIC2 Database.* Retrieved May 20, 2006 from http://www2.cdc.gov/nioshtic-2/Nioshtic2.htm

NIOSH. (2004). *Worker Health Chartbook 2004*. Cincinnatti: DHHS. Retrieved June 29, 2006, from http://www.cdc.gov/niosh/docs/chartbook/

NIOSH. (2001). *Tracking Occupational Injuries, Illnesses, and Hazards: The NIOSH Surveillance Strategic Plan*. Retrieved June 29, 2006, from http://www.cdc.gov/niosh/2001-118.html

NIOSH. (2000). *Worker Health Chartbook, 2000.*

NIOSH. (1993). *Fatal Injuries to Workers in the United States, 1980-1989: A Decade of Surveillance.* DHHS (NIOSH) Pub.No. 93-108

NIOSH. (1991). *Work-Related Lung Disease Surveillance Report.* DHHS (NIOSH) Pub. Number 91-113.

NIOSH. Division of Respiratory Disease Studies, Surveillance Branch. *National Occupational Respiratory Mortality System (NORMS).* Retrieved April 21, 2006 from http://webappa.cdc.gov/ords/norms.html

Oh, J. H., & Shin, E. H. (2003). Inequalities in nonfatal work injury: The significance of race, human capital, and occupations. *Social Science and Medicine*, *57*(2173-2182)

Ohsfeldt, R. L., & Morrisey, M. A. (1997). Beer taxes, workers' compensation, and industrial injury. *The Review of Economics and Statistics, 79*(1), 155-160.

Olsen, J. (2003, August 4). Migrant pipeline fills meatpackers' needs: Word-of-mouth offers have reduced the reliance on risky recruiting trips. Nebraska's oversight law. *Omaha World-Herald,* p. 1A.

O'Neill, R. (2002). *The union effect.* 78. Retrieved from http://www.hazards.org/unioneffect/

Occupational Safety and Health Administration (OSHA). (Undateda). *OSHA Assistance for the Meat Packing Industry*. Retrieved June 30, 2006 from http://www.osha.gov/SLTC/meatpacking

OSHA. (Undatedb). *Poultry Processing Industry E-tool.* Retrieved July 7, 2004 from http://www.osha.gov/SLTC/etools/poultry/index.html

OSHA. (2001). *Occupational Injury and Illness Recording and Reporting Requirements*. Federal Register 66:5916-6135. Retrieved July 12, 2006 from http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_id=16312&p_table=FEDERAL_REGISTER

OSHA. (2006). *General Duty Standard Search.* Retrieved June 15, 2006 from http://www.osha.gov/pls/imis/generalsearch.html

OSHA. (2003). *Ergonomics for the Prevention of Musculoskeletal Disorders: Draft*

Guidelines for Poultry Processing. Retrieved June 30, 2006 from http://www.osha.gov/ergonomics/guidelines/poultryprocessing/index.html

OSHA. (1993 (originally published 1990)). *Ergonomics Program Management Guidelines for Meatpacking Plants.* Retrieved June 29, 2006 from http://www.osha.gov/Publications/osha3123.pdf

OSHA / The Memory Hole. (2005 November). *OSHA's Lost Workday Injury and Illness Database.* Retrieved February 2, 2006 from http://www.thememoryhole.org/osha/lwdii.htm

OSHA Directorate of Compliance Programs. (July 13, 2001). *Site Specific Targeting 2001 (SST-01).* OSHA Directive Number 01-01 (CPL 2). Appendix A

OSHA Directorate of Compliance Programs. (April 19, 1999). *Site Specific Targeting.* OSHA Directive Number 99-3 (CPL 2). Appendix A

OSHA Directorate of Compliance Programs. (April 15, 2002). *Site Specific Targeting 2002 (SST-02),* OSHA Directive Number 02-02 (CPL 2). Appendix A.

OSHA Directorate of Compliance Programs. (April 10, 1998). *Interim Plan for Inspection Targeting,* OSHA Directive Number 98-1 (CPL 2)

OSHA Directorate of Compliance Programs,. (September 8, 2000). *Site Specific Targeting 2000 (SST-MM), OSHA Directive Number 00-05 (CPL 2).* Appendix A.

Pickle, L. W., Mason, T. J., Howard, N., Hoover, R., & Fraumeni, J. F., Jr. (1990). *Atlas of U.S. Cancer Mortality among Nonwhites: 1950-1980.* DHHS Publ. No. (NIH) 90-

1582). Washington, DC: US Gov. Printing Office.

Pickle, L. W., Mason, T. J., Howard, N., Hoover, R., & Fraumeni, J. F., Jr. (1987). *Atlas of U.S. Cancer Mortality among Whites: 1950-1980*. DHHS Publ. No. (NIH) 87-2900. Washington, DC: US Gov. Printing Office.

Pollack, E. S., & Keimig, D. G. (1987). *Counting Injuries and Illnesses in the Workplace: Proposals for a Better System.* Washington, DC: National Academy Press.

Population Studies Center, University of Michigan. *Racial residential segregation measurement project.* Retrieved June 29, 2006 from http://enceladus.isr.umich.edu/race/calculate.html

Pransky, G., Moshenberg, D., Benjamin, K., Portillo, S., Thackrey, J. L., & Hill-Fotouh, C. (2002). Occupational risks and injuries in non-agricultural immigrant Latino workers. *American Journal of Industrial Medicine*, *42*, 117-123.

Pransky, G., Snyder, T., Dembe, A., & Himmelstein, J. (1999). Underreporting of work-related disorders in the workplace: A case study and review of the literature. *Ergonomics*, *42*, 171-182.

Probst, T. M., & Brubaker, T. L. (2001). The effects of job insecurity on employee safety outcomes: Cross-sectional and longitudinal explorations. *Journal of Occupational Health Psychology*, *6*(2), 139-159.

Quinlan, M., Mayhew, C., & Bohle, P. (2001). The global expansion of precarious employment, work disorganization, and consequences for occupational health: A review of recent research. *International Journal of Health Services, 31*(2), 335-414.

R Development Core Team. (2005). *R: A Language and Environment for Statistical Computing. Version 2.1.1.* Vienna, Austria.

Reilly, B., Paci, P., & Hall, P. (1995). Unions, safety committees, and workplace injuries. *British Journal of Industrial Relations, 33*(2), 276-288.

Ribeiro, P. J., Jr., & Diggle, P. J. (2001). geoR: A package for geostatistical analysis. *R-NEWS, 1*(2)from http://cran.r-project.org/doc/Rnews

Richardson, D., Loomis, D., Bailer, A. J., & Bena, J. (2004). The effect of rate denominator source on U.S. fatal occupational injury rate estimates. *Am J Ind Med, 46*(3), 261-270.

Richardson, D. B., Loomis, D., Bena, J., & Bailer, A. J. (2004). Fatal occupational injury rates in Southern and non-Southern states, by race and Hispanic ethnicity. *American Journal of Public Health, 94*(10), 1756-1761.

Ringquist, E. J. (2005). Assessing evidence of environmental inequities: A meta-analysis. *Journal of Policy Analysis and Management, 24*(2), 223-247.

Ripley, B. (2006). *tree: Classification and regression trees (R Program.)* Retrieved from http://cran.r-project.org/src/contrib/Descriptions/tree.html

Robinson J.C. (1989). Exposure to occupational hazards among Hispanics, Blacks, and non-Hispanic Whites in California. *American Journal of Public Health, 79*(5), 629-30.

Robinson J.C. (1986). Hazard pay in unsafe jobs: Theory, evidence, and policy implications. *The Milbank Quarterly, 64*(4), 650-677.

Robinson J.C. (1984). Racial inequality and the probability of occupation-related injury or illness. *Milbank Memorial Fund Quarterly/Health and Society, 62*(4), 567-590.

Robinson, J. C. (1988). The rising long-term trend in occupational injury rates. *American Journal of Public Health, 78*(3), 276-81.

Robinson, J. C. (1986). Job hazards and job security. *Journal of Health Politics, Policy and Law, 11*(1), 1-18.

Rodriguez, R. (2003, March 2). Untraditional immigration: Hispanics flocking to different states. *San Antonio Express-News*, p. 12A.

Roe, B., Irwin, E. G., & Sharp, J. S. (2002). Pigs in space: Modeling the spatial structure of interior US hog production from 1992-97.[Electronic version]. *American Journal of Agricultural Economics, 84*, 259-278.

Rosenman, K. D., Kalush, A., Reilly, M. J., Gardiner, J. C., Reeves, M., & Luo, Z. (2006). How much work-related injury and illness is missed by the current national surveillance system? *JOEM, 48*(4), 357-365.

Schlosser, E. (2001). *Fast Food Nation: The Dark Side of the All-American Meal* Boston: Houghton Mifflin.

Schmitt, J. (2002). Labor markets and economic inequality in the United States since the end of the 1970's. *Paper for the VI Curso Laboral at the Universidad Nacional*

Autonoma De Mexico.

Schulte, P. A. (2005). Characterizing the Burden of Occupational Injury and Disease. *J Occup Environ Med, 47*(6), 607-622.

Shannon, H. S., & Vidmar, M. (2004). How low can they go?: Potential for reduction in work injury rates. *Injury Prevention, 10*, 292-295.

Shen, W., & Louis, T. A. (2000). Triple-goal estimates for disease mapping. *Statist. Med., 19*, 2295-2308.

Shrader-Frechette, K. (2002). Trading jobs for health: Ionizing radiation, occupational ethics, and the welfare argument. *Science and Engineering Ethics, 8*, 139-154.

Simpson, C. L., & Severson, R. K. (2000). Risk of Injury in African American Hospital Workers. *JOEM, 42*(10), 1035-1040.

Smith, G.S. (2003). Injury prevention: Blurring the distinctions between home and work. *Injury Prevention*, *9*, 3-5.

Smith, G. S. (2001). Public health approaches to occupational injury prevention: Do they work? *Injury Prevention, 7 (Suppl I),* i3-i10.

Smith, G. S., Wellman, H. M., Sorock, G. S., Courtney, T. K., & Pransky, G. S. (2005). Injuries at work in the US adult population: Contributions to the total injury burden and opportunities for prevention. *Am J Pub Health, 95*(7), 1213-1219.

Smitha, M. W., Kirk, K. A., Oestenstad, K. R., Brown, K. C., & Lee, S. D. (2001). Effect of state workplace safety laws on occupational injury rates. *Journal of Occupational* *and Environmental Medicine, 43*(12), 1001-1009.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* Thousand Oaks, CA: Sage Publications.

Stanbury, M., Largo, T. W., Granger, J., Cameron, L. & Rosenman, K. (2004). *Profiles of Occupational Injuries and Diseases in Michigan.* Retrieved September 28, 2005 from http://www.michigan.gov/documents/mdch_ProfilesReport_94691_7.pdf

StataCorp. (2005). *Stata Statistical Software, Version 9*. College Station, TX.

Steenland, K., Burnett, C., Lalich, N., Ward, E., & Hurrell, J. (2003). Dying for work: The magnitude of US mortality from selected causes of death associated with occupation. *American Journal of Industrial Medicine, 43*, 461-482.

Stokols, D., McMahan, S., Clitheroe Jr., H. C., & Wells, M. (2001). Enhancing corporate compliance with worksite safety and health legislation. *Journal of Safety Research, 32*, 441-463.

Stone, B. J., Blot, W. J., & Fraumeni, J. F. (1978). Geographic patterns of industry in the United States: An aid to the study of occupational disease. *J Occ Med, 20*(7), 472-7.

Stout, N. A., Jenkins, E. L., & Pizatella, T. J. (1996). Occupational injury mortality rates in the United States: Changes from 1980 to 1989. *American Journal of Public Health*, 73-77.

Stull, D. D., & Broadway, M. J. (2004). *Slaughterhouse Blues: The Meat and Poultry Industry in North America* Belmont, CA: Thomson Wadsworth.

Sundin, D. S., & Frazier, T. M. (1989). Hazard surveillance at NIOSH. *American Journal of Public Health*, *79*(Supplement), 32-37.

Takala, J., & International Labour Organization. (2005). *Introductory Report: Decent Work -- Safe Work.* Retrieved September 23, 2005 from http://www.ilo.org/public/english/protection/safework/wdcongrs17/intrep.pdf

Thacker, S. B. (2000). Historical Development. In S. M. Teutsch, & R. E. Churchill (Eds.), *Principles and Practice of Public Health Surveillance* (2nd Edition ed.). New York: Oxford University Press.

Thacker, S. B., Stroup, D. F., Parrish, R. G., & Anderson, H. A. (1996). Surveillance in environmental public health: Issues, systems, and sources. *Am. J. Pub. Hlth.*, *86*(5), 633-638.

U.S. Bureau of Labor Statistics. (Undated). *Occupational injuries and illnesses: Industry data (1989-2001).* Retrieved July 8, 2004 from http://data.bls.gov/cgi-bin/dsrv?sh

U.S. Bureau of Labor Statistics. (2004). *Table SNR02: Highest incidence rates of nonfatal occupational injury and illness cases with days away from work, restricted work activity, or job transfer, private industry, 2002*from http://bls.gov/iif/oshwc/osh/os/ostb1233.txt

U.S. Bureau of the Census. (2000). *Summary File 3.* Retrieved from: http://www.census.gov

U.S. Department of Health and Human Services. (2000). *Healthy People 2010*.

Washington, DC: US Department of Health and Human Services.

U.S. Dept. of Labor, Bureau of Labor Statistics. (Undated). *Food processing industries.* Retrieved July 7, 2004 from http://www.bls.gov/oco/ocos219.htm

U.S. Dept. of Labor, Bureau of Labor Statistics. (2006). *Injuries, Illnesses and Fatalities Home Page.* Retrieved June 17, 2006 from http://www.bls.gov/iif/home.htm

U.S. Dept. of Labor, Bureau of Labor Statistics. (2003). *Fatal Occupational Injuries in the United States, 1995-1999: A Chartbook.* BLS Report 965. Washington, DC:

U.S. General Accounting Office. (1998). *Community Development: Changes in Nebraska's and Iowa's Counties with Large Meatplacking Plant Workforces. GAO/RCED-98-62*. Washington, DC. Retrieved from http://www.gao.gov/archive/1998/rc98062.pdf

U.S. Government Accountability Office. (2005). *Workplace Safety and Health: Safety in the Meat and Poultry Industry, While Improving, Could be Further Strengthened. GAO-05-96*. Washington, DC. Retrieved from http://www.gao.gov/new.items/d0596.pdf

United Food and Commercial Workers. (2005a). *Fighting for poultry justice in Congress and in our communities.* Retrieved January 20, 2005 from http://www.ufcw.org/workplace_connections/meatpacking_and_poultry/workplace_issues/index.cfm

United Food and Commercial Workers. (2005b). *Justice at Smithfield.* Retrieved January

20, 2005 from http://www.ufcw.org/issues_and_actions/justice_at_smithfield/index.cfm

Ussif, A. A. (2004). An international analysis of workplace injuries. *Monthly Labor Review*, 41-51.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S.* (Fourth ed.). New York: Springer.

Volinn, E., Lai, D., McKinney, S., & Loeser, J. D. (1998). When back pain becomes disabling: a regional analysis. *Pain, 33*, 33-39.

Waehrer, G., Leigh, J. P., Cassady, D., & Miller, T. R. (2004). Costs of occupational injury and illness across states. *J Occup Environ Med., 46*, 1084-1095.

Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data* Hoboken, NJ: John Wiley and Sons, Inc.

Watson, A. P., & White, C. L. (1984). Workplace injury experience of female coal miners in the United States. *Archives of Environmental Health, 39*(4), 284-293.

Weil, D. (1997). Implementing employment regulation: Insights on the determinants of regulatory performance. In B. E. Kaufman (Ed.), *Government Regulation of the Employment Relationship*. Madison, WI: Industrial Relations Research Association.

Weil, D. (1991). Enforcing OSHA: The role of labor unions. *Industrial Relations, 30*(1), 20-36.

Wigglesworth, E. C. (2005). Do some US states have higher/lower injury mortality rates than others? *Journal of Trauma: Injury Infection & Critical Care, 58*(6), 1144-1149.

Williams, D. R., & Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports, 116*, 404-416.

Wilson, S. M., Howell, F., Wing, S., & Sobsey, M. (2002). Environmental injustice and the Mississippi hog industry. *Environ Health Perspect, 110 Suppl*(2), 195-201.

Windau, J. A., Jack, T. A., & Toscano, G. A. (Summer 1998). State and industry fatal occupational injuries, 1992-96. *Compensation and Working Conditions Online,*

Wing, S., Cole, D., & Grant, G. (2000). Environmental injustice in North Carolina's hog industry. *Environ Health Perspect, 108*(3), 225-231.

APPENDICES

APPENDIX 1: Top industries by number surveyed in OSHA Data Initiative (ODI) 1997-2001.

Industry	Standard Industrial Classification (SIC)	Number	% of total (218,757)	Mean LWDII (all estabs, 1997-2001)
Skilled Nursing Care Facilities	8051	23,500	10.74%	9.57
Nursing and Personal Care Facilities, Not Elsewhere Classified	8059	6,985	3.19%	8.44
Plastics Products, Not Elsewhere Classified	3089	6,228	2.85%	7.45
Trucking, Except Local	4213	5,969	2.73%	6.69
Lumber and Other Building Materials Dealers	5211	4,134	1.89%	6.73
Department Stores	5311	3,758	1.72%	5.19
Motor Vehicle Parts and Accessories	3714	3,718	1.70%	7.66
General Medical and Surgical Hospitals	8062	3,359	1.54%	4.24
General Warehousing and Storage	4225	3,197	1.46%	9.25
Intermediate Care Facilities	8052	3,006	1.37%	9.24

APPENDIX 2: Top industries by lost workday injury/illness rate (LWDII) in ODI, 1997-2001

The list of industries having the highest mean LWDII rates may be influenced by industries having small numbers of included establishments and one or a few outlier rates. We did not evaluate this possibility, nor the impact on ranks of repeated sampling of the same establishment. As a partial way of dealing with these concerns, the table below excludes industries with fewer than 5 establishments in the 5-year sample.

Industry	Standard Industrial Classification (SIC)	Mean LWDII of all establishments, averaged across 5 years	# Establishments in ODI (1997-2001)
Courier Services, Except by Air	4215	15.9	1,479
Drapery Hardware and Window Blinds and Shades; Furniture and Fixtures, Not Elsewhere Classified	259	13.35	13
Meat Packing Plants	2011	12.70	947
Malleable Iron Foundries	3322	12.46	49
Air Courier Services	4513	12.27	1,022
Gray and Ductile Iron Foundries	3321	11.50	1,016
Logging	241	11.42	8
Special Product Sawmills, Not Elsewhere Classified	2429	11.23	34
Mobile Homes	2451	11.01	825

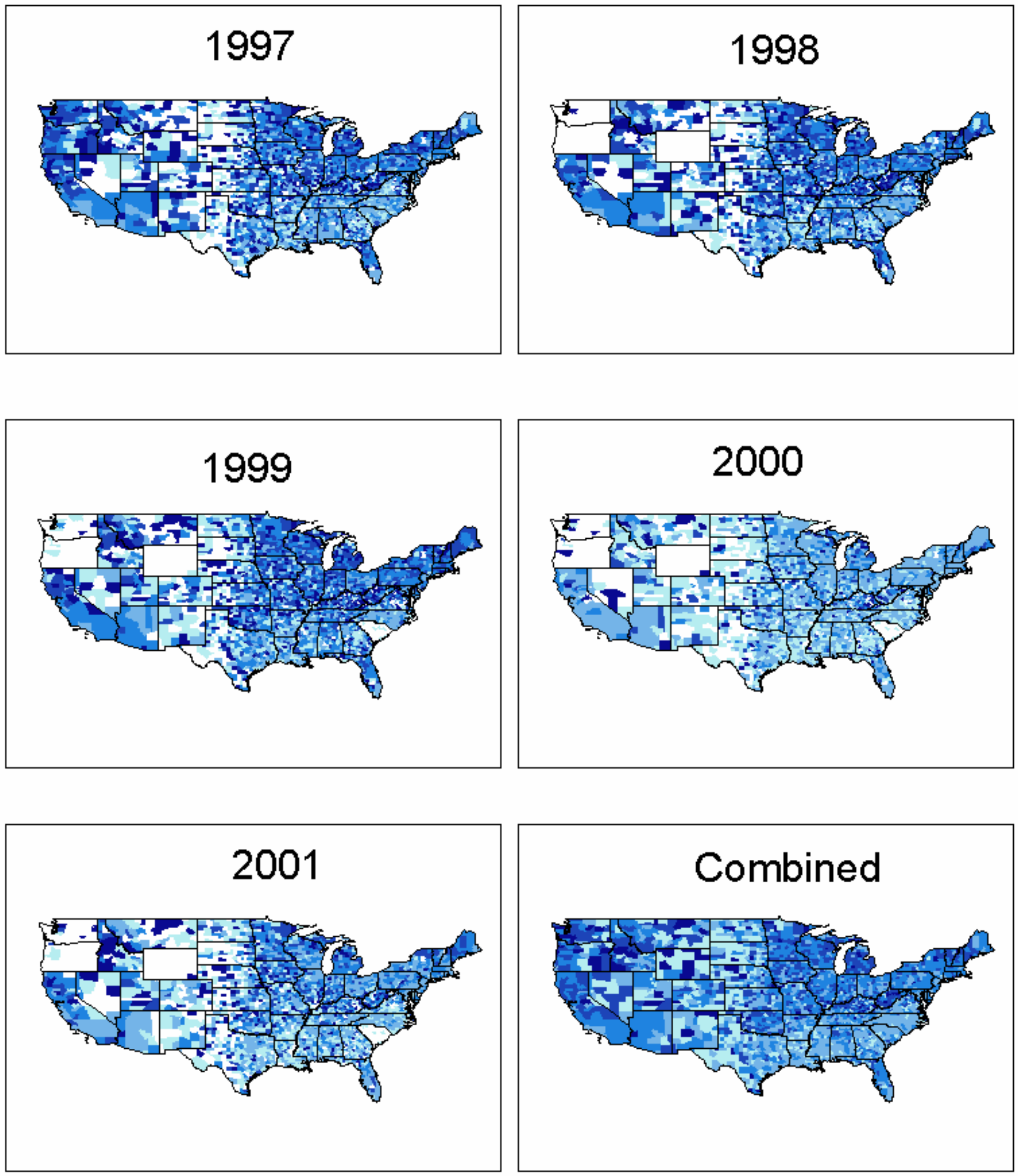
APPENDIX 3: Maps of LWDII by year

The “by year” maps below depict county percent of establishments having “high” LWDII (>8.0 per 100). These maps do not exclude the six states cut from subsequent analyses, nor establishments with fewer than 60 employees or reported rates over 60. The combined map divides counties into quintiles, and the “by year” maps use the same quintile cutpoints to show variation by year.

These maps show that there was variation by year, and that in particular, rates were lower in 2000. However the basic distribution of rates across the map is reasonably consistent across years.

Is There Variation in
Geography of High Rate ODI Establishments By Year?

(% of county establishments with LWDII >=8, quintiles, by year)



APPENDIX 4: ODI sampling criteria by year and discussion/analysis of issue of dropping low rate establishments in subsequent years.

One potential source of bias in the ODI database comes from OSHA's sampling strategy of not resampling establishments in subsequent years if they reported low rates in some years. The ODI is not intended for surveillance, and the aim of improving efficiency in identifying high-rate establishments was more important than that of consistency. This sampling strategy could lead to differential misclassification bias if low rate establishments are not evenly distributed. It could also lead to changes in the county distribution of sampling, if low rate establishments are removed from the sample of one county and their replacements are located in different counties.

A dataset was created of low rate establishments across all years. This was defined as LWDII <=3, or below the national average despite being in a high rate industry.

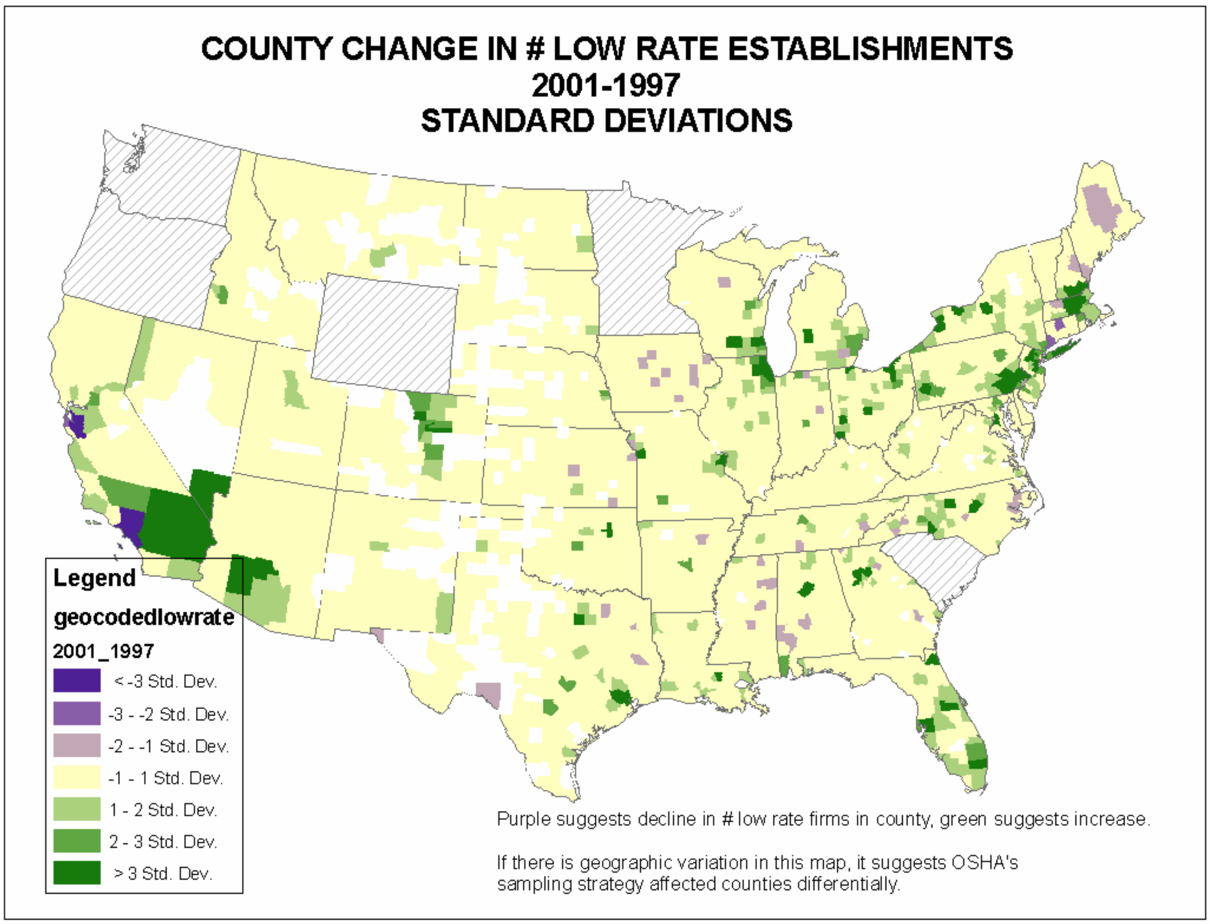
To examine geographic patterns of the above phenomenon, the number of low rate establishments by county was mapped in each year to look for changes. A comparison was performed to see how much county numbers of low rate establishments changed across time. While OSHA's targeted sampling suggests a decrease in low rate establishments would occur, there was also a concurrent national trend of dropping LWDII rates, that should lead to increased low rate establishments. The table below shows the number and percent of establishments with LWDII <=3. Overall, numbers of low-rate firms in the ODI decline for the first three years, then rise dramatically in 2000 and drop back somewhat in 2001.

year	Freq. LWDII <=3	Percent of total
1997	22,542	18.10
1998	21,319	17.12
1999	20,255	16.27
2000	30,448	24.45
2001	29,966	24.06
Total	124,530	100.00

We also examined changes in number of low rate establishments by county, both for each "year 2 minus year 1" pair, and from the beginning to the end of the sample period. The mean difference by year was as follows, suggesting a similar result to that in the table above.

	98-97	99-98	00-99	2001-2000	2001-1997
MEAN	0	0	4	0	3

Next we ask whether counties have DIFFERENTIALLY experienced change in the number of low rate observations across time. A set of maps was created depicting the geographic variation in change in number of "low rate" establishments in the survey. The distribution of these maps varies by years. The map below, "County Change in # Low Rate Establishments, 2001-1997" shows that overall few counties had changes greater than 1 standard deviation. However, there were more counties coded in greens than purples, showing more counties with *increased* low rate firms sampled than not, and further, that these were not randomly distributed across the map, but rather, occurred in small clusters.



This finding suggests either that in green counties, rates legitimately declined at a particularly fast rate; or that there was a particular rise in low rate firms or industries added to their samples in subsequent years.

There was not a good way to account for this bias in the analysis, so it is just noted in discussion.

APPENDIX 5: Comparing Geocoded and Non-geocoded Records in the OSHA Data Initiative

This appendix explores whether establishments that could not be geocoded and included in analyses differed from others in the ODI sample.

Approximately 0.9 percent of the ODI sample could not be geocoded based on zip code locations. Geocoding is a process whereby addresses in a database are matched with known geographic points. Non-geocodes had zip codes that did not occur in the zip code file provided by ESRI in its ARC GIS materials.

Analyses were performed to try to assess the extent to which non-geocoded establishments might differ from others, and whether they might differ geographically.

Table 1 below lists the number and percent ungeocoded by year, suggesting that the rate remained fairly constant across time.

Table: Number and percent ungeocoded by year.

YEAR	N (after exclusions*)	# not Geocoded (%)
1997	44,199	544 (1.2)
1998	36,043	332 (0.9)
1999	42,270	369 (0.9)
2000	45,026	232 (0.5)
2001	49,308	434 (0.9)
TOTAL	216,846	1,941 (0.9)

* Exclusions: HI, MN, OR, SC, WA, WY; establishments with <60 employees; establishments reporting LWDII>50

Tables 2 and 3 compare ungeocoded and tied records with geocoded ones based on LWDII and employer size categories, finding differences to be negligible, with the mean difference for both = 0.000. Overall, there is no evidence of establishment difference based on geocoding status.

Table 2: Comparison of ungeocoded and geocoded records by LWDII category

LWDII Category	% in Ungeocoded	% in Geocoded	Difference
0	0.120491	0.151626	-0.031
0 to 3	0.225521	0.190703	0.035
3 to 6	0.22454	0.214428	0.010
6 to 7	0.066748	0.060034	0.007
7 to 8	0.056319	0.053231	0.003
8 to 10	0.086135	0.086786	-0.001
10 to 15	0.124294	0.133108	-0.009
15 plus	0.095951	0.110084	-0.014
MEAN DIFFERENCE			**0.000**

Table 3: Comparison of ungeocoded and geocoded records by employer size category

Number of employees	% in Ungeocoded	% in Geocoded	Difference
1 – 39	0.097	0.108	-0.011
40 - 49	0.065	0.094	-0.029
50 - 59	0.065	0.086	-0.021
60 - 79	0.101	0.138	-0.038
80 - 99	0.076	0.104	-0.029
100 - 249	0.279	0.307	-0.027
250 - 499	0.131	0.099	0.032
500 - 999	0.091	0.040	0.050
1000 plus	0.097	0.024	0.073
MEAN DIFFERENCE			**0.000**

APPENDIX 6: Index of Area Industry Hazard: Calculations

Following is the code in the statistical package R (by Frank Curriero, 2005) for generating the Index of Area Industry Hazard based on LWDII data by year downloaded from the Bureau of Labor Statistics.

CREATING INDEX OF AREA INDUSTRY HAZARD

1) Adding annual LWDII rates collected from Bureau of Labor Statistics website into the ODI database by 4-digit sic code (if there was no 4-digit match, 3-digit SIC code or even 2-digit was used.)

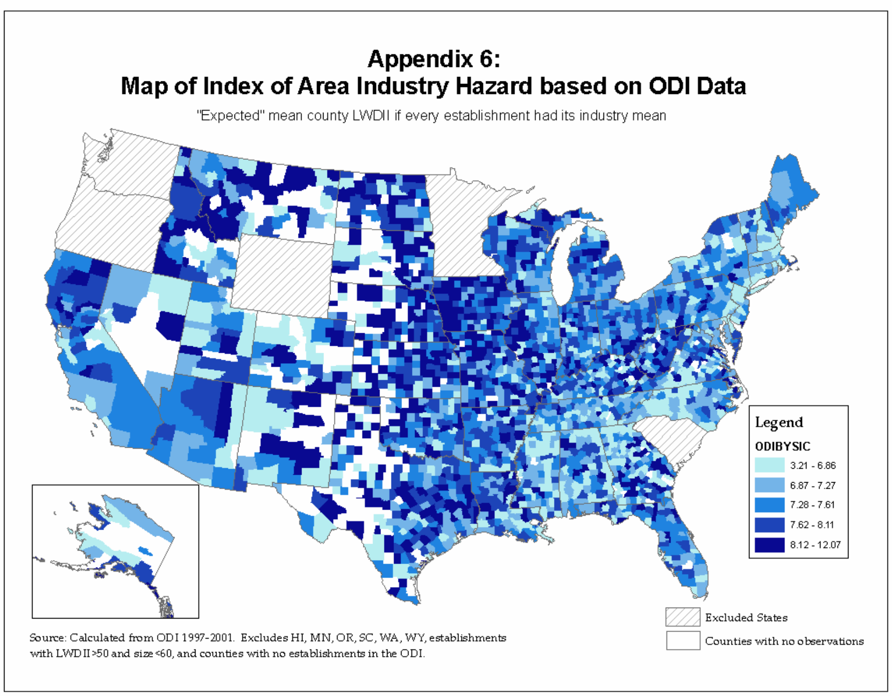
```
temp<-unique(c(LWDII97$SIC,LWDII98$SIC,LWDII99$SIC,LWDII00$SIC,LWDII01$SIC))
LWDII<-data.frame(SIC=temp,Total.Cases97=NA,LWDII97=NA,Total.Cases98=NA,
LWDII98=NA,
Total.Cases99=NA,LWDII99=NA,Total.Cases00=NA,LWDII00=NA,
                           Total.Cases01=NA,LWDII01=NA)

u<-match(LWDII97$SIC,LWDII$SIC)
LWDII$Total.Cases97[u]<-LWDII97$TOTAL.CASES
LWDII$LWDII97[u]<-LWDII97$LWDII

u<-match(LWDII98$SIC,LWDII$SIC)
LWDII$Total.Cases98[u]<-LWDII98$TOTAL.CASES
LWDII$LWDII98[u]<-LWDII98$LWDII

u<-match(LWDII99$SIC,LWDII$SIC)
LWDII$Total.Cases99[u]<-LWDII99$TOTAL.CASES
LWDII$LWDII99[u]<-LWDII99$LWDII

u<-match(LWDII00$SIC,LWDII$SIC)
LWDII$Total.Cases00[u]<-LWDII00$TOTAL.CASES
LWDII$LWDII00[u]<-LWDII00$LWDII

u<-match(LWDII01$SIC,LWDII$SIC)
LWDII$Total.Cases01[u]<-LWDII01$TOTAL.CASES
LWDII$LWDII01[u]<-LWDII01$LWDII

u1<-match(ODI97$sic4,LWDII97$SIC)
u2<-match(ODI97$sic3,LWDII97$SIC)
u3<-match(ODI97$sic2,LWDII97$SIC)

# The R object dataframe called LWDII has all the combined information.  Later
# this file will be exported for potential use elsewhere.


# Adding LWDII to ODI Data
# ------------------------
u1<-match(ODI97$sic4,LWDII97$SIC)
u2<-match(ODI97$sic3,LWDII97$SIC)
u3<-match(ODI97$sic2,LWDII97$SIC)

umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI97<-data.frame(ODI97,LWDII=LWDII97$LWDII[umatch])
```

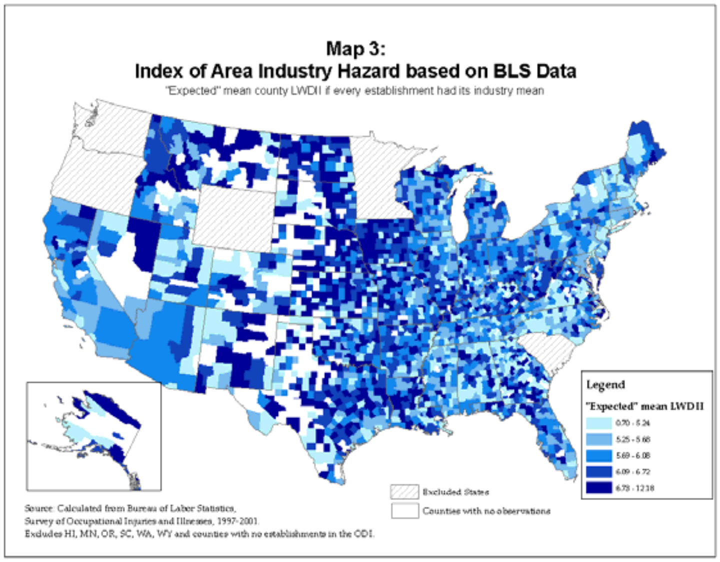
```
u1<-match(ODI98$sic4,LWDII98$SIC)
u2<-match(ODI98$sic3,LWDII98$SIC)
u3<-match(ODI98$sic2,LWDII98$SIC)

umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI98<-data.frame(ODI98,LWDII=LWDII98$LWDII[umatch])

u1<-match(ODI99$sic4,LWDII99$SIC)
u2<-match(ODI99$sic3,LWDII99$SIC)
u3<-match(ODI99$sic2,LWDII99$SIC)

umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI99<-data.frame(ODI99,LWDII=LWDII99$LWDII[umatch])

u1<-match(ODI00a$sic4,LWDII00$SIC)
u2<-match(ODI00a$sic3,LWDII00$SIC)
u3<-match(ODI00a$sic2,LWDII00$SIC)

umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI00a<-data.frame(ODI00a,LWDII=LWDII00$LWDII[umatch])

u1<-match(ODI00b$sic4,LWDII00$SIC)
u2<-match(ODI00b$sic3,LWDII00$SIC)
u3<-match(ODI00b$sic2,LWDII00$SIC)

umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI00b<-data.frame(ODI00b,LWDII=LWDII00$LWDII[umatch])
u1<-match(ODI01a$sic4,LWDII01$SIC)
u2<-match(ODI01a$sic3,LWDII01$SIC)
u3<-match(ODI01a$sic2,LWDII01$SIC)
umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI01a<-data.frame(ODI01a,LWDII=LWDII01$LWDII[umatch])
u1<-match(ODI01b$sic4,LWDII01$SIC)
u2<-match(ODI01b$sic3,LWDII01$SIC)
u3<-match(ODI01b$sic2,LWDII01$SIC)
umatch<-u1
umatch[is.na(u1)]<-u2[is.na(u1)]
umatch[is.na(u1)&is.na(u2)]<-u3[is.na(u1)&is.na(u2)]
ODI01b<-data.frame(ODI01b,LWDII=LWDII01$LWDII[umatch])

# Creating Index
# ------------------------
# Index of Area Industry Hazard is calculated as the average
# LWDII over counties (and the same is done for states).
# Each observation in ODI
# is then given its respective county (and state) average.
t2<-tapply(ODI$LWDII,ODI$fips,mean,na.rm=T)
u<-match(ODI$fips,as.numeric(names(t2)))
ODI$hazind3<-round(as.vector(t2[u]),2)

t2<-tapply(ODI$LWDII,ODI$statefips,mean,na.rm=T)
u<-match(ODI$statefips,as.numeric(names(t2)))
ODI$hazind4<-round(as.vector(t2[u]),2)
```

APPENDIX 7: Comparing the Index of Area Industry Hazard based on the Bureau of Labor Statistics, Survey of Occupational Injuries and Illnesses (SOII) data with an index based on ODI data itself

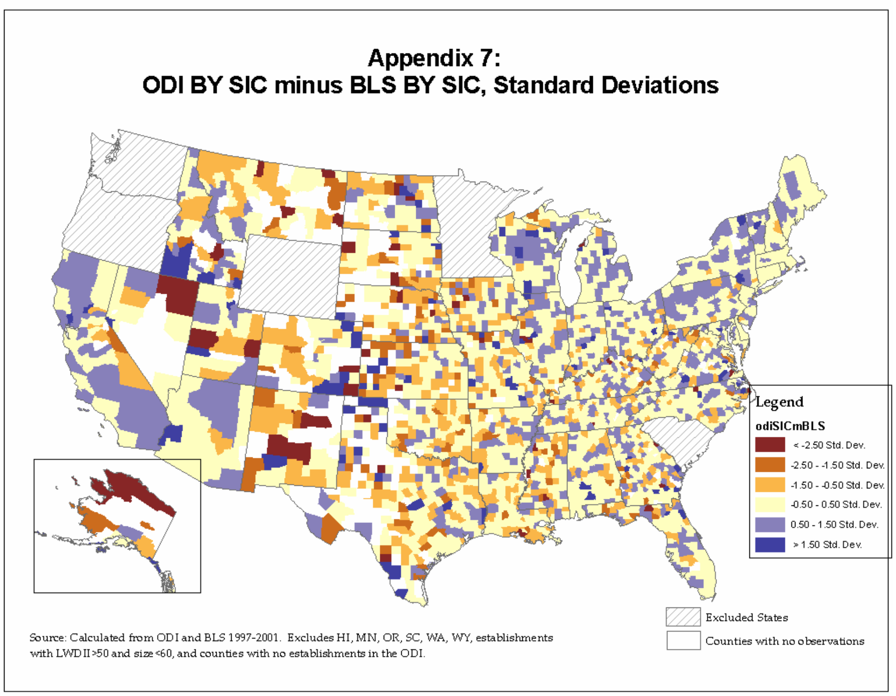
We define the Index of Area Industry Hazard as the "expected" mean county lost workday injury/illness rate (LWDII) if every establishment had the mean rate of its industry (4-digit Standard Industrial Classification code). For the analyses in the thesis, the industry mean is calculated based on rates in the Bureau of Labor Statistics Survey of Injuries and Illnesses (SOII), which may provide more accurate reporting than the ODI (Map 3 in Surveillance paper, presented below for comparison). This Index accounts for area differences in industry hazard and reporting, as well as differences in reporting between the ODI and SOII. An alternate Index uses the industry mean reported in ODI itself, thus focusing more directly on differences in hazard and reporting, removing database differences from the equation.

Geographic distribution of the ODI Index is presented below, showing that it follows a similar pattern to the original Index, with higher expected rates towards the center of the country and lower ones at the coasts. Index rates are lower in parts of the South than elsewhere, suggesting that the finding of low rates in the South may be related to the South being represented by lower rate industries than elsewhere, or that industries common to Southern states are especially likely to have underreporting. Louisiana, which had particularly low rates in the observed sample, is not included in the low-rate area of the South in this map.





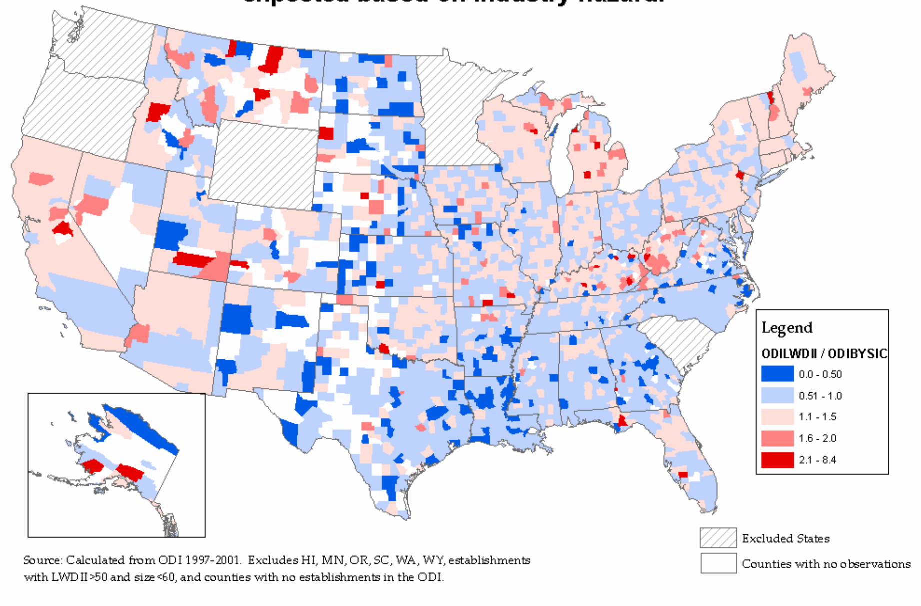
The map below compares the above ODI Index and SOII Index directly by subtracting the latter from the former. Results are presented in standard deviations rather than quintiles, to focus more clearly on areas where the ODI index is higher than the SOII one (blues) and lower (reds/oranges.) While the ODI Index does find higher expected rates in the center of the country than are observed in the actual ODI database (Map 2, Surveillance paper), ODI index rates there are still lower than the SOII index. The ODI index particularly exceeds the SOII in the Northeast, Michigan-Wisconsin, and California, among other areas.



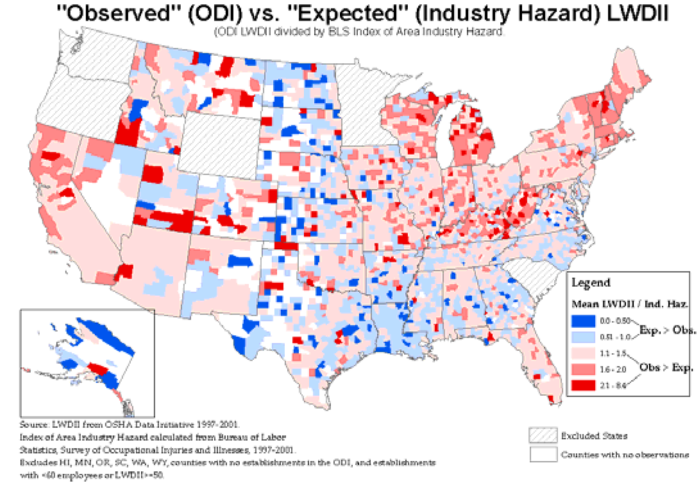
APPENDIX 8: Map of difference between ODI Index ("expected") and ODI LWDII ("observed") rates

This map shows the distribution of the difference between "observed" and "expected" based on industry hazard, when the Index of Industry Hazard based on the ODI sample instead of the BLS sample is used. The map should be compared to Map 4 in the surveillance paper (below), which presents the same comparison but using the SOII Index instead of the ODI one. The distribution in both maps is similar. Contrasting the ODI-based (above) and SOII-based (below) maps, it can be seen that in the ODI map, the expected is > observed (blue) in more areas than the SOII map, and there are fewer areas of extreme difference between observed and expected in either direction (dark blue or red). This is probably partly because using the ODI database takes out the extra muck introduced by having two different databases. But the widespread and geographically differential finding of expected rates being less than observed in the ODI also suggests that underreporting may be more at play in this database than in the SOII.

Appendix 8:
Difference between ODI "observed" LWDII and ODI Index "expected based on industry hazard."



Map 4:
"Observed" (ODI) vs. "Expected" (Industry Hazard) LWDII
(ODI LWDII divided by BLS Index of Area Industry Hazard)



APPENDIX 9: Steps for performing surveillance analyses

Following is an outline of steps to perform the tracking analyses described in this paper. These could be adapted for a different database. The following computer programs were used: Excel, dbase, ARC GIS, Stata, StatTransfer, R Statistical software, GeoDa.

A) CREATING THE DATA FILES

1) Convert the database to a dbf file, preferably using the StatTransfer program. Assign a unique identifier to each record by sorting however you want, then in a new column, typing 1,2,3,4 in the first four rows. Then drag the "fill handle"  through the full range of cells you want to number.
2) Add the file into GIS and geocode based on zip code.
3) Add a shape file showing county boundaries. Join the county file to the ODI file "based on spatial location". Export and name the file. This assigns a county and state name to each establishment.
4) Transfer the data into a statistical analysis package.
5) Exclude observations that will be excluded: We excluded six states, all establishments with LWDII >50, and all establishments with fewer than 60 employees.
6) Create SOII index of industry hazard as follows.

- Copy LWDII rates by industry and year from the Bureau of Labor Statistics Survey of Injuries and Illnesses website (I cut and pasted the file as a text document, then opened it in excel, deleting variables I did not want, then stat transferred to stata.)
- Use a statistical package to assign each observation the LWDII for its 4-digit SIC code by year. If the 4-digit SIC code was not listed in the BLS file, try the 3-digit code, and finally the 2-digit code if needed. Call this the SOII index.

7) To get the ODI Index of Industry Hazard:

- Collapse the main establishment level file by 4-digit SIC code, saving the mean LWDII.
- Name this variable the ODI index, and merge by SIC with the establishment-level file.

8) For any other text variables provided by OSHA, create numeric categorized variables.
9) Merge with data sets of other potential explanatory variables. *This is the establishment level file.*
10) Collapse by county, saving the count of observations in the county, the count of observations >= your threshold for 'high injury/illness,' mean of LWDII's, hazard indices, categorized variables, and county/state level explanatory variables. *This is the county-level file.*
11) Collapse by state, saving the count of observations in the state, the count of observations >= your threshold for 'high injury/illness,' mean of LWDII's, hazard indices, categorized variables, and county/state level explanatory variables. *This is the state-level file.*
12) Construct tables by year indicating number in sample, number and percent "high rate," percent of observations not geocoded, and any yearly variations in sampling.

B) BASIC COUNTY-LEVEL MAPS (mean LWDII, Index of Area Industry Hazard, LWDII divided by Index [the latter variable can be calculated in GIS])

1) Convert the county-level file to dbase and add to GIS. Geocode using a county file, based on "FIPS" (the 5-digit code reflecting the county and state). You are ready to map!
2) To set map projection, right click on "layer" and click Properties, then Coordinate Systems, Projected, Continental, North America, and finally, USA Contiguous Albers Equal Area Conic USGS.
3) Right click on the county level file and click Properties. Go to symbology tab, quantities, and select variables to map. Most of our maps use quintiles. Remove county boundary lines for better visibility.
4) Insert a second data frame for Alaska (Hawaii data were too sparse and inconsistent to use in this analysis), and copy and paste in each layer. Change the extent and size of this layer to fit the state.
5) For all maps: Insert a file with state boundaries (no fill, just the boundaries) as the top layer. Make a selection from the state file to include only excluded states, and export this to be a new layer. I colored these only with diagonal lines. Put this in after your main map layer. Add a layer to show counties that were not part of the sample (use a copy of the county-level file, and in symbology, manually create two levels, one with the count of establishments=0, one with count >0. The symbol for count=0 should be white with no borders, and for count=1 should be clear symbol, clear border, so it doesn't show at all.) This can be the last layer on maps. These layers can be copied and pasted from one map to another.
6) Insert legend, title.

Map count, percent high rate, industry hazard index, percent high rate normalized by industry hazard index in quintiles. Consider using Geoda for exploratory spatial data analysis. This program allows "dynamic linking" so that you can highlight particular observation/s on one map, chart, or graph, and see where it/they fall on a variety of other charts/maps/graphs. Gather data on mean(SD), range, # missing, # coded '0', correlations with % high rate, and data sources into a table.

C) STATE LEVEL MAP

1) Adapt the steps above to the state level file (geocoding is based on "state fips" or state name).

D) MORAN'S I and LISA

1) It is necessary to remove Alaska from the dataset first, due to its distance from the mainland. In ARC GIS with the county file, select Alaska, "switch selection" so you have all the other states, and export to a new file.
2) Moran's I and LISA's were computed using GeoDa. (Geoda was not used for mapping because it has fewer graphic options than ARC GIS.) Open GeoDa and input your new shape file, with "fips" as the "key variable."
3) Before running the analyses, create a "weights" file, by clicking on tools (from tabs on top), weights, create. The input file is your file. Make a new file name for the output, use fips as the identifier. I used queen contiguity and also experimented with other options.

4) From the top tabs, click on Space, Univariate moran, and the outcome variable you are looking at. Enter the new weight file you created and run. Right-click on the output to get the option for randomization, and try 9999 permutations.
5) To run the LISA analysis, click on Space, Univariate LISA, and the outcome variable, and enter in the weight file. Click to get all the maps. Right-click on the output to get the option for randomization, and try 9999 permutations. Save these, especially the cluster map. I reopened the cluster map shape file in ARC GIS and formatted it to match the formatting of the other maps in the thesis.

E) RANKING

1) Using the state-level file in excel, states were sorted on mean LWDII and the top-ranked states were reported with accompanying details.
2) Using the county-level file in excel, counties were sorted on mean LWDII and the top-ranked counties were reported with accompanying details.
3) The file was sorted on number of establishments per county, and counties with at least 30 establishments were chosen. Then they were sorted on mean LWDII and the top-ranked establishments reported.
4) The Bayesian ranking (Louis, 2006; Shen & Louis, 2000) was run in a single year. To select a year, the analysis in (3) was run separately by years and results were compared by hand to see which year delivered top 10 counties that most frequently occurred in other years.

- After a year was selected, a file was created following similar steps to that in (A) above, with only 1998 data. This was saved as a “csv” file. All counties with no observations were deleted. Then the following commands were used in R statistical software.

```
rank1998.data <- read.csv("C:/OSHA 2-06/98ranking.csv",
                  col.names=c("fips","meanodi98","sdodi98","id98"))
attach(rank1998.data)
Y <- meanodi98
V <- sdodi98^2
nreps = 100  #Should be sufficient.
gamma = .7
##########################################
# Initialization
n = length(Y)
D = rep(1,  n)  #= (1-B)
mu= 0
postm=rep(0,n)
postv=rep(1,n)
cb=postm
##########################################
# The recursion
for(reps in 1:nreps)
{
for( k in 1:n)
```

```
{
postm[k]= mu + D[k]*(Y[k] - mu)
postv[k] = D[k]*V[k]
}

mu = mean(postm)
TT = mean(postv) + mean((postm - mu)^2)
D=TT/(TT+V)
}
###########################################
# Compute summaries

cb = mu +sqrt(D)*(Y-mu)  # This is "constrained Bayes"
ranky=rank(Y)
rankpostm=rank(postm)
rankcb=rank(cb)
print(list(mu=mu,TT=TT))

# Rbar
rbar= rep(1, n)

for(i in 1:(n-1))
for(j in (i+1):n)
{
rbar[i] = rbar[i] + pnorm((postm[i]-postm[j])/sqrt(postv[i]+postv[j]))
rbar[j] = rbar[j] + 1- pnorm((postm[i]-postm[j])/sqrt(postv[i]+postv[j]))
}
rhat=rank(rbar)
###########################################
outputall=cbind(fips,id98,Y,postm,cb,V,postv,ranky,rankpostm,rankcb,rhat,rbar)
write.csv(outputall, file = "C:/OSHA 2-06/98rankresults.csv")
```

The top ten Bayesian-ranked counties are reported with their direct ranks juxtaposed.

F. PLACING 'HIGH RATE' COUNTIES IN CONTEXT OF EXPLANATORY VARIABLES

1) Calculate correlations between each explanatory variable and the outcome and view relationships using scatterplots and other nonspatial methods. Try using Geoda's "Explore" tab to run exploratory analyses. It enables "brushing" to select counties on any of the charts or tables and have them highlighted on all associated tables and maps.
2) Create two-variable maps:
    a. Determine cutoff points (experiment with several) for "high rate" and values of the explanatory variables. We used tertiles.

b. Code tertiles. Assign each county three dichotomous variables: “top 1/3 for risk factor x”(coded: 0/1); “middle third mean injury/illness rate” (coded: 0/3); and “top third mean injury/illness rate” (coded 0/5).
    i. In ARCGIS, sort by the variable and select the observations in the top or bottom tertile (e.g., the first 1047 of 3141 US counties). The number of observations will vary if there is missing data or if multiple observations near the cutpoint have the same value. Note the cutpoints. Create the middle tertile variable by selecting observations using the tertile cutpoints you just found.
c. Add a new column for each explanatory variable, called something like “[variable]_sum”. Sum the three variables you just created.
d. Go to the layer’s symbology, and select the new sum variable. Click on “classify” and then “exclusion.” Exclude if the variable = 0, 3, or 5. Then the remaining values can be labeled as follows:
    - *High risk factor, LOW injury/illness rate (sum=1)*
    - *High risk factor, MEDIUM injury/illness rate (sum=4)*
    - *High risk factor, LOW injury/illness rate (sum=6)*
e. Leave the background white both for counties and states not in the database, and for those falling in the bottom 2/3 for the risk factor.
f. Add titles, legends, and variable descriptions to the maps.
g. The number of “overlaps” between top tertile of the risk factor and of LWDII can be counted and compared with what might be expected if both categories were distributed geographically randomly, calculated as (n1/N * n2/N)*N, where n1 and n2 represent the number of observations in the top third of each variable and N is the total number of included counties.
h. For keeping similar format in making multiple maps, it is helpful to: add a second copy of the layer to an existing map, but leave it unchecked so the information doesn’t appear twice. Then, “save as” your new map name and when you go to ‘symbology,’ click on ‘import’ and import the symbology from the other layer. Then you just have to edit the data labels and change the map title, variable notes, and legend.
i.

APPENDIX 10: List of risk factors and sources

	Variable	Source
	Economy	
1.	% Unemployed (self report)	Census 2000
2.	% White collar (managerial, professional, technical, sales, admin support)	Census 2000
3.	% Blue collar	Census 2000
4.	Median household income	Census 2000
5.	Per capita income	Census 2000
6.	Change in # Manufacturing establishments, 1979-1995	Census Counties 1998 CD ROM
7.	Change in # Manufacturing establishments with >100 employees, 1979-1995	Census Counties 1998 CD ROM
8.	Change in unemployment rate 1975 – 1996	Census Counties 1998 CD ROM
9.	% with housing costs at least 35% of household income	Census 2000
10.	Change in unemployment (claims per population), 1997-2001 [used both to evaluate short-term change and for sensitivity analyses regarding combining 5 years of data]	Bureau of Economic Analysis
11.	Median family income	Census 2000
12.	Per capita retail sales 1992	Counties 1998
13.	Sales, shipments, receipts, revenues in 1000s	Calculated from Economic Census, 1997
14.	Sales, shipments, receipts, revenues in 1000s per population	Calculated from Economic Census, 1997 and Census 2000
15.	Sales, shipments, receipts, revenues in 1000s per establishment	Calculated from Economic Census, 1997
16.	Value of manufacturing shipments in 1000s	Calculated from Economic Census, 1997
17.	Wholesale sales	Calculated from Economic Census, 1997
18.	% Unemployed (based on claims filed) 1997	Bureau of Economic Analysis
19.	% Unemployed (based on claims filed) 2001	Bureau of Economic Analysis
20.	# Walmart stores	From file constructed by Robert Cook from Walmart atlas. : http://www.discoveryowners.com/cginfo.htm#textmapfiles
21.	# Walmart stores per population	Calculated from file constructed by Robert Cook from Walmart atlas. : http://www.discoveryowners.com/cginfo.htm#textmapfiles and US Census 2000
22.	Median household income (state)	Census 2000
	Demographics	

	Variable	Source
23.	% with < high school diploma	Census 2000
24.	% with bachelor's degree or greater	Census 2000
25.	% of households that were "linguistically isolated"	Census 2000
26.	% recent immigrants entering US between 1990 and 2000	Census 2000
27.	% minority	Census 2000
28.	% living below poverty line	Census 2000
29.	% white	Census 2000
30.	% African American	Census 2000
31.	Segregation: index of exposure (for black and white groups. Measures the extent to which an individual is likely to be exposed to members of the other racial group.	Calculated based on black vs. white residence in census tracts. For US counties, based on formula provided by the University of Michigan Population Studies Center. http://enceladus.isr.umich.edu/race/calculate.html
32.	Economic inequality: Gini coefficient (a widely used measure based on the Lorenz curve of the distribution (specifically, the area of the difference between that curve and a uniform distribution, divided by the area under the uniform distribution).	Computed based on census tract household income for U.S. counties by Volscho, T. from Census 2000
33.	% "linguistically isolated" and Spanish speaking	Census 2000
34.	% foreign born, 1990	Census 2000
35.	% foreign born, 2000	Census 2000
36.	Change in # foreign born between 1990 and 2000	Census 2000
37.	Change in % foreign born between 1990 and 2000	Census 2000
38.	% who graduated high school but did not complete 4-years of college	Census 2000
39.	Racial segregation: index of dissimilarity. For black and white groups. Measures % of population that would need to move to get an even distribution.	Calculated based on black vs. white residence in census tracts across counties, based on formula provided by the University of Michigan Population Studies Center. http://enceladus.isr.umich.edu/race/calculate.html
40.	County % age 0-15	Census 2000
41.	County % age 16-21	Census 2000
42.	County % age 22-29	Census 2000
43.	County % age 30-44	Census 2000
44.	County % age 45-69	Census 2000
45.	County % age 70-74	Census 2000
46.	County % age 75 +	Census 2000

	Variable	Source
	Policy, Culture	
47.	Local government's % spending on education, 1992	Counties 1998
48.	Local government's % spending on health, 1992	Counties 1998
49.	Local government's % spending on public welfare, 1992	Counties 1998
50.	Total # of violations cited by OSHA	OSHA IMIS – provided database
51.	# violations with initial penalty	OSHA IMIS – " "
52.	# violations with current penalty	OSHA IMIS – " "
53.	$ value of initial penalties	OSHA IMIS – " "
54.	$ value of current penalties	OSHA IMIS – " "
55.	Right to work state (Y/N) [this policy allows non-union workers to work in unionized facilities and is often used as an indicator of anti-union policy]	National Right to Work Legal Foundation. www.nrtw.org/a/rtwempl.htm
56.	# of OSHA inspections between 1997 and 2001	OSHA IMIS – calculated from website
57.	Maximum weekly unemployment benefit, January 2001 (state)	Department of Labor. www.workforcesecurity.doleta.gov/unemploy/pdf.200/ch304.pdf
58.	State rank: ratio of unemployment benefits to average weekly wage (1 reflects strong safety net; 51 is weak).	AFL CIO. www.aflcio.org/issuespolitics/stateissues/ui/upload/stateranking_04072003.pdf
59.	% of workers who were unionized, 1999 (state)	Compiled by Hirsch BT and Macpherson D, from Current Population Survey data www.unionstats.com
60.	Whether state operated its own OSHA program	OSHA. www.osha.gov/oshdir/states.html
61.	State rank: business-friendly tax policy	Tax Foundation: 2003 State Business Tax Climate Index. http://www.taxfoundation.org/bp45.pdf
62.	Number of days workers must wait before obtaining workers' compensation benefits	State Workers' Compensation Laws compiled by the US Department of Labor. http://www.dol.gov/esa/regs/statutes/owcp/stwclaw/tables-html/table-14.htm
	Environmental	
63.	Total air emissions, 2000	Toxics Release Inventory, 2000
64.	Waste Due to Catastrophic or One Time Events – All substances combined, 2000	Toxics Release Inventory, 2000
	Control – Other	

	Variable	Source
65.	Mean number of employees per establishment (categories)	ODI
66.	Industry hazard (county): “Expected” *county* lost workday injury/illness rate if every ODI establishment had its industry mean	ODI and Bureau of Labor Statistics
67.	Industry hazard (state): “Expected” *state* lost workday injury/illness rate if every establishment had its industry mean	ODI and Bureau of Labor Statistics
68.	% of residents who work in the county [to evaluate appropriateness of county as unit of analysis]	Census 2000
69.	Census region	Census 2000
70.	Whether state is in South (TX, LA, MS, AL, NC, GA, VA, AR, TN)	States selected based on standard definitions of “South” but excluding Florida due to its different outcome pattern compared to other S’ern states
71.	Number of establishments surveyed in ODI 1997-2001	ODI
	Control – Area, Density	
72.	Area – square miles	Census 2000
73.	Whether county is part of a Metropolitan Statistical Area (MCMSA) [urban or suburban]	Census 2000
74.	Population 2000	Census 2000
75.	Population density	Census 2000
76.	Number age 16+ in civilian labor force	Census 2000
77.	Number of establishments	Economic Census 1997
78.	% Urban	Census 2000
79.	% Urban Area (densely settled, 50,000 or more people)	Census 2000
80.	% Urban Cluster (contains 2,500-49,999 people)	Census 2000
81.	% Rural	Census 2000
82.	% Rural – Farm	Census 2000
83.	% Rural – No farm	Census 2000
84.	Population density (state)	Census 2000
	Control - Industry	
85.	Number manufacturing establishments	Calculated from Economic Census 1997
86.	Number production workers	Calculated from Economic Census 1997
87.	Number retail establishments	Calculated from Economic Census 1997
88.	Number wholesale establishments	Calculated from Economic Census 1997
89.	% of private non-farm establishments that were retail, 1995	Counties 1998
90.	% of private non-farm establishments that were service, 1995	Counties 1998

APPENDIX 11: Risk factors discussed in regression results section, with: means; ranges; coefficients from bivariate regression versus mean county LWDII. Table of correlation coefficients. Variance Inflation Factors.

Items in bold were included in the final model.

Risk Factor	Mean (range)	Bivariate Coef. versus County Mean LWDII
Demographics		
% African American †	**0.09 (0, 0.86)**	**-5.47*** (div by 100)**
Segregation Index of Dissimilarity *(For African American and Caucasian groups. Reflects the portion of a county's population that would need to change census tracts to get an even distribution.)*	0.33 (0, 0.96)	2.31***
Segregation Index of Exposure *(For African American and Caucasian groups. Measures the extent to which an individual has members of the other racial group in his/her census tract.)*	0.07 (0, 0.85)	-6.44***
% "Linguistically Isolated" †	0.2 (0, 0.33)	-6.79***
% Immigrated between 1990-2000 †	0.02 (0, 0.19)	-6.49**
% Foreign Born 2000 †	0.03 (0, 0.51)	-1.81
Change in % Foreign Born Between 1990 and 2000 †	0.01 (00.13, 0.21)	-8.62**
Gini Coefficient of Economic Inequality *(A widely used inequality measure comparing the observed income distribution to a uniform one. Specifically, reflects the area of the difference between the distribution's Lorenz curve and a uniform distribution, divided by the area under the uniform distribution).*	0.40 (0.29, 0.62)	-11.55***
County % age 0-15 †	22.4 (8.7, 41.7)	-7.9 ***
County % age 16-21 †	8.5 (2.5, 38.1)	-1.7
County % age 22-29 †	9.2 (0, 24.2)	-4.8
County % age 30-44 †	13.9 (2.98, 2.7)	0.9

Risk Factor	Mean (range)	Bivariate Coef. versus County Mean LWDII
County % age 45-69 †	27.3 (6.4, 46.3)	8.3 ***
County % age 70-74 †	3.7 (0.3, 9.2)	0.5
County % age 75 + †	7.0 (0.5, 19.6)	0.4
Local Economy/SES		
% < Poverty †	**0.14 (0.02, 0.51)**	**-4.84*****
Per capita income	17513 (5213, 44962)	0.00003
% Unemployed †	0.23 (0.03, 0.65)	0.66
% < High School †	0.17 (0.05, 0.64)	-4.44***
% High School Grad but no College †	0.61 (0.28, 0.81)	6.57***
% Bachelor's Degree †		0.26
Change in Unemployment between 1975 and 1996	-1.47 (-19.2, 30.3)	-0.06***
Policy/Culture/Values		
% Unionized (state)	**11.46 (3.2, 25.3)**	**0.16*****
Rank: ratio of unemployment benefits to average weekly wage (state)	**26.70 (2, 51)**	**-0.02*****
OSHA – # inspections in county	140.99 (0, 11814)	0.00004
OSHA – total county violations 1997-2001	444.8 (0, 37599)	0.00002
OSHA – county violations with initial penalty, 1997-2001	219.4 (0, 35692)	5.13e-06
OSHA – county violations with current penalty, 1997-2001	194 (0, 27954)	0.0002
OSHA – total $ of county initial penalties, 1997-2001 (among cited counties)	5173.4 (100, 197991.67)	4.56e-06
OSHA – total $ of county current penalties, 1997-2001(among cited counties)	2872.7 (54.5, 81237.9)	3.10e-06
State runs OSHA program instead of federal government (Y/N)	29 states	0.48***
Total air emissions, 2000	190544 (0, 999885)	-1.71e-07
Waste Due to Catastrophic or One Time Events – All substances combined, 2000	1737.95 (0, 98425)	-3.96e-06

Risk Factor	Mean (range)	Bivariate Coef. versus County Mean LWDII
Industry Hazard		
Index of Area Industry Hazard	**7.17 (0, 25.2)**	**0.75*****
Control		
% Rural – Farm †	**0.04 (0, 0.32)**	**-1.72**
South	**35.7% South; 64.7% non-south**	**-2.06**
Establishment Size	Data provided in categories only.	-0.20***
# Establishments in county in ODI	82.3 (1, 5685)	0.00007

* p<0.05; ** p<0.01; *** p<0.001

† Regression coefficient should be divided by 100 because data were presented in decimal rather than percent format.

Correlations Between Variables Used in Final Model

The following table of correlation coefficients and Variance Inflation Factor (VIF) table show that while correlations between a few variables were elevated, overall multicollinearity was not so high as to problematize the final model.

cntyodi2	pblack	p_ltpov	pctunion	rankal~o	cnty_bls	prur_f~m	south	
cntyodi2	1.0000							
pblack	-0.2724	1.0000						
p_ltpov	-0.1056	0.4663	1.0000					
pctunion	0.2959	-0.3363	-0.3068	1.0000				
rankalfcio	-0.0912	0.2173	0.1790	0.1796	1.0000			
cnty_bls	0.2797	-0.1157	0.0100	0.0025	-0.0778	1.0000		
prur_farm	-0.0281	-0.2761	-0.0554	-0.0369	-0.2062	0.2534	1.0000	
south	-0.3370	0.5522	0.3288	-0.6606	0.1894	-0.1225	-0.2688	1.0000

Variance Inflation Factors (VIFs)

Variable	VIF	1/VIF
south	2.76	0.362613
pctunion	2.38	0.420961
pblack	1.73	0.579457
p_ltpov	1.37	0.731900
rankalfcio	1.31	0.764084
prur_farm	1.27	0.789036
cnty_bls	1.08	0.927786
Mean VIF	1.70	

APPENDIX 12: R Code for Calculating Semivariogram

SEMIVARIOGRAM

```
resid<-read.csv("c:\\regression06\\residuals\\countyodi2all_no0rev_321.csv",
header=TRUE)

temp<-as.geodata(cbind(resid$x_coord,resid$y_coord,resid$stdres))

library('geoR')
resid<-
read.csv("c:\\regression06\\residuals\\countyodi2all_no0rev_321.csv",header=TRU
E)
temp<-as.geodata(cbind(resid$ALB_X,resid$ALB_Y,resid$STDRES_320))

vario<-variog(temp)
vario$max.dist
[1] 4339714

vario<-variog(temp,max.dist=4339714/2)
plot(vario)

vario<-variog(temp,max.dist=200000/2)
plot(vario)
```

APPENDIX 13: Linear regressions by year and at establishment level

Below are establishment level linear regression analyses by year using the same variables as were used in the final model; and an establishment level linear regression combining all years, and controlling for “year.”

Overall, these analyses show essentially similar coefficient direction and magnitude to those in the final model (Regression paper, Table 17.) The main difference from the final model is that the percent < poverty measure was negative in all years but 1998, and only reached statistical significance in 1997. The poverty measure consistently showed results like these in establishment level analyses, whereas at the county level it was often statistically significant and positive.

NOTE: Although technically appropriate, multilevel models were not used for these analyses, since the goal was to quickly compare the coefficients with those in the final model, not to get the statistical significance right. Also note that coefficients for “pblack” (% black), “p_ltpov”(% below poverty), and “prur_farm” (% rural farm) should be divided by 100 because data were presented in decimals rather than fractions.

```
. use "C:\OSHA 2-06\odi2all_no0_nog50.dta", clear

. by year: regress cnty_odi pblack p_ltpov pctunion rankalfcio cnty_bls
prur_farm south

-------------------------------------------------------------------------------
-> year = 1997

      Source |       SS       df       MS              Number of obs =   44199
-------------+------------------------------           F(  7, 44191) = 2987.92
       Model |  54291.8988     7  7755.98554           Prob > F      =  0.0000
    Residual |   114710.31 44191  2.59578443           R-squared     =  0.3212
-------------+------------------------------           Adj R-squared =  0.3211
       Total |  169002.209 44198   3.8237524           Root MSE      =  1.6111

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.104498   .0743906   -28.29   0.000    -2.250305   -1.958691
     p_ltpov |  -.6150082    .185701    -3.31   0.001    -.9789853   -.251031
    pctunion |   .0596238   .0020821    28.64   0.000     .0555429    .0637047
  rankalfcio |  -.0303685   .0006768   -44.87   0.000    -.0316951   -.0290419
    cnty_bls |   1.436377   .0146222    98.23   0.000     1.407717    1.465037
   prur_farm |  -11.35437   .3847446   -29.51   0.000    -12.10847   -10.60026
       south |  -.5243433   .0277468   -18.90   0.000    -.5787276   -.4699591
       _cons |  -.0429269    .087765    -0.49   0.625    -.214948    .1290941
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> year = 1998
```

```
      Source |       SS       df       MS              Number of obs =   36043
-------------+------------------------------           F(  7, 36035) = 2380.75
       Model |  41797.6823     7  5971.09747           Prob > F      =  0.0000
    Residual |  90378.3113 36035  2.50807025           R-squared     =  0.3162
-------------+------------------------------           Adj R-squared =  0.3161
       Total |  132175.994 36042  3.66727689           Root MSE      =  1.5837

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.355457   .0825071   -28.55   0.000    -2.517173   -2.193741
     p_ltpov |   .3975179   .2034573     1.95   0.051    -.0012645    .7963004
    pctunion |   .0589147   .0021777    27.05   0.000     .0546463    .0631832
  rankalfcio |  -.0317325   .0007479   -42.43   0.000    -.0331984   -.0302666
    cnty_bls |   1.336705    .016033    83.37   0.000      1.30528     1.36813
   prur_farm |  -9.057146   .4066569   -22.27   0.000    -9.854206   -8.260086
       south |  -.5524307   .0299409   -18.45   0.000    -.6111158   -.4937456
       _cons |   .5180823    .096035     5.39   0.000     .3298508    .7063138
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> year = 1999

      Source |       SS       df       MS              Number of obs =   42270
-------------+------------------------------           F(  7, 42262) = 2802.38
       Model |  48482.8728     7  6926.12469           Prob > F      =  0.0000
    Residual |  104451.313 42262  2.47151845           R-squared     =  0.3170
-------------+------------------------------           Adj R-squared =  0.3169
       Total |  152934.186 42269  3.61811696           Root MSE      =  1.5721

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.461243   .0750427   -32.80   0.000    -2.608328   -2.314158
     p_ltpov |  -.1745643   .1846729    -0.95   0.345    -.5365269    .1873984
    pctunion |   .0529436   .0020163    26.26   0.000     .0489916    .0568956
  rankalfcio |  -.0302674    .000692   -43.74   0.000    -.0316237   -.0289111
    cnty_bls |   1.291496   .0146491    88.16   0.000     1.262783    1.320208
   prur_farm |   -10.1096   .3679538   -27.48   0.000    -10.83079   -9.388399
       south |  -.6382113   .0274443   -23.25   0.000    -.6920026     -.58442
       _cons |   .8465509   .0876495     9.66   0.000      .674756    1.018346
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> year = 2000

      Source |       SS       df       MS              Number of obs =   45026
-------------+------------------------------           F(  7, 45018) = 2802.75
       Model |  54485.7871     7  7783.68387           Prob > F      =  0.0000
    Residual |  125021.988 45018  2.77715554           R-squared     =  0.3035
-------------+------------------------------           Adj R-squared =  0.3034
       Total |  179507.775 45025  3.98684675           Root MSE      =  1.6665

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.188722   .0782782   -27.96   0.000    -2.342148   -2.035295
     p_ltpov |  -.2390021   .1908632    -1.25   0.210     -.613097    .1350929
    pctunion |   .0530037   .0020621    25.70   0.000      .048962    .0570454
  rankalfcio |  -.0309661   .0007044   -43.96   0.000    -.0323468   -.0295854
    cnty_bls |   1.441794   .0155351    92.81   0.000     1.411345    1.472243
   prur_farm |  -8.156972   .4014262   -20.32   0.000    -8.943774    -7.37017
```

```
       south |  -.6200394   .0283045   -21.91   0.000    -.6755168   -.5645621
       _cons |   .0016242   .0926536     0.02   0.986    -.1799784    .1832268
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> year = 2001

      Source |       SS       df       MS              Number of obs =   49308
-------------+------------------------------           F(  7, 49300) = 3573.55
       Model |  62401.3422     7  8914.47746           Prob > F      =  0.0000
    Residual |  122982.317 49300  2.49457033           R-squared     =  0.3366
-------------+------------------------------           Adj R-squared =  0.3365
       Total |  185383.659 49307  3.75978379           Root MSE      =  1.5794

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |   -2.30305    .070023   -32.89   0.000    -2.440296   -2.165804
     p_ltpov |  -.7216988   .1728087    -4.18   0.000    -1.060406   -.3829916
    pctunion |   .0452127   .0018022    25.09   0.000     .0416803    .0487451
  rankalfcio |  -.0309486   .0006299   -49.14   0.000    -.0321831   -.0297141
    cnty_bls |   1.536006   .0141939   108.22   0.000     1.508185    1.563826
   prur_farm |  -8.099304   .3903998   -20.75   0.000    -8.864492   -7.334115
       south |  -.6087671   .0250588   -24.29   0.000    -.6578827   -.5596515
       _cons |  -.3518964   .0841797    -4.18   0.000    -.5168896   -.1869033
------------------------------------------------------------------------------
```

Regression with the full database, controlling for year, found that year was not a significant predictor.

. regress cnty_odi pblack p_ltpov pctunion rankalfcio cnty_bls prur_farm south year

```
      Source |       SS       df       MS              Number of obs =  216846
-------------+------------------------------           F(  8,216837) =12658.52
       Model |  260949.686     8  32618.7107           Prob > F      =  0.0000
    Residual |  558749.772216837  2.57681933           R-squared     =  0.3183
-------------+------------------------------           Adj R-squared =  0.3183
       Total |  819699.458216845  3.78011694           Root MSE      =  1.6052

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.279145   .0339146   -67.20   0.000    -2.345616   -2.212673
     p_ltpov |  -.3040747   .0836363    -3.64   0.000    -.4679997   -.1401497
    pctunion |   .0533759    .000903    59.11   0.000      .051606    .0551458
  rankalfcio |   -.030891   .0003075  -100.45   0.000    -.0314937   -.0302883
    cnty_bls |   1.411548   .0066986   210.72   0.000     1.398419    1.424677
   prur_farm |  -9.497853   .1743755   -54.47   0.000    -9.839625   -9.156082
       south |   -.593157   .0123398   -48.07   0.000    -.6173427   -.5689714
        year |  -.0016817   .0023851    -0.71   0.481    -.0063564     .002993
       _cons |   3.549776   4.768709     0.74   0.457    -5.796773    12.89633
------------------------------------------------------------------------------
```

Establishment-level linear regression model

Below, the regression final model (Table 17) is performed as an establishment level linear regression, to examine the impact of the county aggregation on findings. Coefficients were generally similar in direction and magnitude to those in the county-level analysis, although the % < poverty measure showed the opposite direction of effect. While that variable reached statistical significance in this analysis, it may not have if significance was properly corrected using multilevel models. (Due to the very large sample size, earlier multilevel models took several days to run, so we did not perform that analysis here.) Note that coefficients for "pblack" (% black), "p_ltpov"(% below poverty), and "prur_farm" (% rural farm) should be divided by 100 because data were presented in decimals rather than fractions.

An establishment level analysis is performed using establishment size in categories ("sizecata") as a predictor, but even with this very large sample size, it did not reach statistical significance.

. use "C:\OSHA 2-06\odi2all_no0_nog50.dta", clear

```
. regress cnty_odi pblack p_ltpov pctunion rankalfcio cnty_bls prur_farm south

      Source |       SS       df       MS              Number of obs =  216846
-------------+------------------------------           F(  7,216838) =14466.84
       Model |  260948.404     7  37278.3435           Prob > F      =  0.0000
    Residual |  558751.053216838  2.57681335           R-squared     =  0.3183
-------------+------------------------------           Adj R-squared =  0.3183
       Total |  819699.458216845  3.78011694           Root MSE      =  1.6052

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.278919    .033913   -67.20   0.000    -2.345388    -2.21245
     p_ltpov |  -.3039629    .083636    -3.63   0.000    -.4678875   -.1400384
    pctunion |     .05338    .000903    59.11   0.000     .0516102    .0551499
  rankalfcio |  -.0308929   .0003075  -100.46   0.000    -.0314956   -.0302902
    cnty_bls |     1.4116   .0066982   210.74   0.000     1.398472    1.424729
   prur_farm |  -9.495748   .1743497   -54.46   0.000     -9.83747   -9.154027
       south |  -.5930998   .0123395   -48.07   0.000    -.6172849   -.5689147
       _cons |   .1875169   .0399924     4.69   0.000     .1091329    .2659009
------------------------------------------------------------------------------
```

Including a variable for establishment size:

```
. regress cnty_odi pblack p_ltpov pctunion rankalfcio cnty_bls prur_farm south
sizecata

      Source |       SS       df       MS              Number of obs =  216846
-------------+------------------------------           F(  8,216837) =12658.59
       Model |  260950.667     8  32618.8333           Prob > F      =  0.0000
    Residual |  558748.791216837   2.5768148           R-squared     =  0.3183
-------------+------------------------------           Adj R-squared =  0.3183
       Total |  819699.458216845  3.78011694           Root MSE      =  1.6052

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -2.278361   .0339182   -67.17   0.000     -2.34484   -2.211883
     p_ltpov |  -.3037716   .0836363    -3.63   0.000    -.4676966   -.1398465
    pctunion |   .0533965   .0009032    59.12   0.000     .0516263    .0551667
  rankalfcio |  -.0308973   .0003075  -100.47   0.000    -.0315001   -.0302945
    cnty_bls |   1.411544   .0066985   210.73   0.000     1.398415    1.424673
   prur_farm |  -9.496878    .174354   -54.47   0.000    -9.838608   -9.155149
       south |  -.5927439   .0123453   -48.01   0.000    -.6169405   -.5685474
    sizecata |    -.00145   .0015476    -0.94   0.349    -.0044833    .0015832
       _cons |   .1913203   .0401979     4.76   0.000     .1125336    .2701071
------------------------------------------------------------------------------
```

APPENDIX 14: Regression using the outcome measure, "average county difference between observed and expected mean LWDII" (with expected mean defined as the Index of Area Industry Hazard.)

The main regression analysis used the outcome measure, "observed county mean LWDII." This thesis has highlighted the difference between the observed and "expected" mean rates based on industry mean rates, and here I present the results of a multilevel regression analysis which takes this difference ("odibls") as the outcome. The final multilevel model is presented below for comparison.

ODIBLS is defined as: average county difference between observed and expected lwdii. Other variables are: south (dichotomous, 1=southern state); pblack (county % African American); punemp (% unemployed); unmp7596 (change in unemployment rate between 1975 and 1996); nmanuest (# manufacturing establishments); prur_farm (% rural farm); rankalfcio (AFLCIO's rank of state ratio of unemployment benefits to average wages); pctunion (state % unionized); wcwait (# of days state workers must wait for workers' compensation benefits). Note that coefficients for "pblack", "punemp" and "prur_farm" should be divided by 100 because data were presented in decimals rather than fractions.

```
.  xi: gllamm odibls south pblack punemp unmp7596 nmanuest prur_farm rankalfcio
pctunion wcwait, i(state_fips) family(gaussian) link(identity) adapt trace

number of level 1 units = 1964
number of level 2 units = 47

Condition Number = 30162.531

gllamm model

log likelihood = -4181.3404

------------------------------------------------------------------------------
      odibls |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |  -.7387888   .3284025    -2.25   0.024    -1.382446   -.0951317
      pblack |  -2.841006   .4524536    -6.28   0.000    -3.727799   -1.954213
      punemp |   7.848287   2.371519     3.31   0.001     3.200195    12.49638
    unmp7596 |   -.056767   .0160465    -3.54   0.000    -.0882176   -.0253164
    nmanuest |  -.0002325   .0000829    -2.80   0.005     -.000395     -.00007
   prur_farm |  -11.41677   1.731971    -6.59   0.000    -14.81137    -8.02217
  rankalfcio |    -.02583   .0084074    -3.07   0.002    -.0423081   -.0093518
    pctunion |   .0697139   .0256838     2.71   0.007     .0193747    .1200532
      wcwait |  -.1848946   .0585576    -3.16   0.002    -.2996654   -.0701239
       _cons |   2.799981   .5000012     5.60   0.000     1.819996    3.779965
------------------------------------------------------------------------------
```

```
Variance at level 1
------------------------------------------------------------------------------

  4.0187967 (.12966002)

Variances and covariances of random effects
------------------------------------------------------------------------------


***level 2 (state_fips)

    var(1): .31022568 (.09730361)
```

Final multilevel model, for comparison:

Risk factor	**Multilevel model:** county and state-level fixed effects; county and state error terms
Demographics	
Percent African American †	-3.12 (-4.10, -2.13)
Local Economy/SES	
% < Poverty †	2.41 (0.46, 4.37)
Policy, Culture, Values	
State percent unionized 1999	0.08 (0.02, 0.15)
State rank: ratio of unemployment benefits to average weekly wage	-0.02 (-0.04, -0.002)
Industry hazard	
Industry hazard: "Expected" county lost workday injury/illness rate if every ODI establishment had its industry mean	0.79 (0.71, 0.89)
Control	
% Rural, Farm †	-10.72 (-13.38, -8.07)
South	-0.96 (-1.77, -0.16)
Constant	2.83 (1.73, 3.93)
Model evaluation	*Log likelihood*= -6218.39. Condition number=338.78
Number observations	*2,657 counties; 45 states*

† Coefficient/CI should be divided by 100.

APPENDIX 15: Regression separately for Southern and non-Southern states and for top quarter high-LWDII industry vs. others

A) Comparison of linear, county-level regression results in Southern (south=1) versus non-Southern (south=0) states finds substantial differences.

- County % African American was much more negatively associated with injury/illness rates in non-South than in Southern counties, although the association was significant and negative in both areas.
- The finding that "percent < poverty" was consistently statistically significant across analyses may be related to the strong differential between its effect in Southern and non-Southern states. In non-South states, it has a strong, significant relationship, while in the South, it has a negative, nonsigificant relationship after controlling for race and other variables.
- Percent unionized had a slightly increased association with high rates in the South than elsewhere.
- In non-Southern areas, "percent rural-farm" had a strong negative association with high rates, whereas in the South this association was mitigated.

(Note that coefficients for "pblack" (% black), "p_ltpov"(% below poverty), and "prur_farm" (% rural farm) should be divided by 100 because data were presented in decimals rather than fractions.)

```
use "\\sph.ad.jhsph.edu\jhsph-data\docstorage01\rneff\My Documents\thesis
regression\regression started 2
> -01\COUNTYodi2all_no0rev_321.dta", clear
. by south: regress cnty_odi pblack p_ltpov pctunion rankalfcio cnty_bls
prur_farm

-------------------------------------------------------------------------------
-> south = 0

      Source |       SS       df       MS              Number of obs =    1711
-------------+------------------------------           F(  6,  1704) =   31.97
       Model |  2262.75309     6  377.125515           Prob > F      =  0.0000
    Residual |  20103.3887  1704  11.7977633           R-squared     =  0.1012
-------------+------------------------------           Adj R-squared =  0.0980
       Total |  22366.1418  1710  13.0796151           Root MSE      =  3.4348

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -5.233812   1.423311    -3.68   0.000    -8.025432   -2.442191
     p_ltpov |   5.982728    1.53262     3.90   0.000     2.976713    8.988743
    pctunion |   .0922318   .0206563     4.47   0.000     .0517175    .1327461
  rankalfcio |   -.025335    .007345    -3.45   0.001    -.0397412   -.0109289
    cnty_bls |   .7824255   .0797691     9.81   0.000     .6259698    .9388812
   prur_farm |  -14.50915   1.701669    -8.53   0.000    -17.84673   -11.17157
       _cons |   2.819058   .6181623     4.56   0.000     1.606621    4.031495
------------------------------------------------------------------------------
```

```
--------------------------------------------------------------------------------
----------------------------
-> south = 1

      Source |       SS       df       MS              Number of obs =     949
-------------+------------------------------           F(  6,   942) =   26.89
       Model |  995.589633     6  165.931605           Prob > F      =  0.0000
    Residual |  5813.34216   942  6.17127618           R-squared     =  0.1462
-------------+------------------------------           Adj R-squared =  0.1408
       Total |  6808.93179   948   7.1824175           Root MSE      =  2.4842

------------------------------------------------------------------------------
    cnty_odi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -1.986663   .5295051    -3.75   0.000     -3.02581   -.9475174
     p_ltpov |  -2.720537   1.502725    -1.81   0.071    -5.669612    .2285391
     pctunion |   .1162585   .0529971     2.19   0.029     .0122524    .2202646
  rankalfcio |  -.0208721    .007323    -2.85   0.004    -.0352435   -.0065008
    cnty_bls |   .7852619   .0748274    10.49   0.000     .6384143    .9321096
   prur_farm |  -4.584173   3.868495    -1.19   0.236    -12.17604    3.007693
       _cons |   2.240582   .5543232     4.04   0.000      1.15273    3.328433
------------------------------------------------------------------------------
```

B) Regression separating out the top ¼ highest LWDII industries
Nearly every coefficient changed substantially in magnitude and direction.

- The high rate industries seem to drive the regression findings more than other industries, as their findings are closer to the reported outcomes.
- In high rate industries, % African American was much more negatively associated with the outcome than in other industries. % < poverty was positively associated with the outcome in high rate industries, whereas in other industries the relationship was nonsignificant and negative. Further, in high rate industries, unions were more positively associated with the high rates than in the rest of the industries, and the safety net had no relationship with the outcome in non-high rate industries. Even the relationships seen between rates and % rural farm and south were driven by high rate industries.
- The coefficient on the Index of Area Industry Hazard ("bls") is an exception, where a closer relationship with the outcome was seen in non-high rate industries. The coefficient was 0.38 in the high rate industries, and 0.82 in other industries. This suggests that as industry hazard increased, there was much less increase in reported rates in high rate industries than in other industries, another potential indicator of underreporting.

(Note that coefficients for "pblack" (% black), "p_ltpov"(% below poverty), and "prur_farm" (% rural farm) should be divided by 100 because data were presented in decimals rather than fractions.)

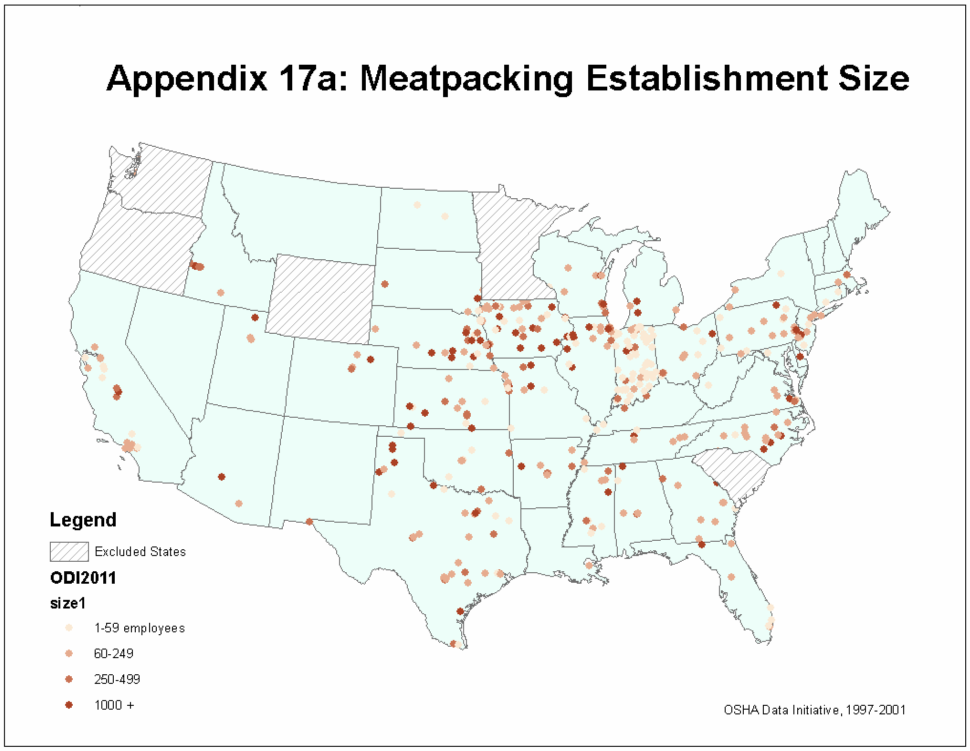
```
. use "C:\OSHA 2-06\topquarterbycty.dta"

. regress odilwdii    pblack p_ltpov pctunion rankalfcio prur_farm bls south

      Source |       SS       df       MS              Number of obs =    2532
-------------+------------------------------           F(  7,  2524) =   79.55
       Model |  5952.68753     7  850.383933           Prob > F      =  0.0000
    Residual |    26979.69  2524  10.6892591           R-squared     =  0.1808
-------------+------------------------------           Adj R-squared =  0.1785
       Total |  32932.3775  2531  13.0116071           Root MSE      =  3.2694

------------------------------------------------------------------------------
    odilwdii |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -4.505446   .5896315    -7.64   0.000    -5.661657   -3.349235
     p_ltpov |    3.87623   1.206648     3.21   0.001     1.510108    6.242352
    pctunion |   .0811586   .0190311     4.26   0.000     .0438404    .1184768
  rankalfcio |  -.0250042   .0056505    -4.43   0.000    -.0360844   -.0139241
   prur_farm |  -20.16514   1.555897   -12.96   0.000    -23.21611   -17.11418
         bls |   .3797503   .0748932     5.07   0.000      .232892    .5266087
       south |  -1.491712   .2258145    -6.61   0.000    -1.934513   -1.048911
       _cons |   6.736815   .6493667    10.37   0.000     5.463469    8.010161


. use "C:\OSHA 2-06\3quartersbycty.dta"
. regress odilwdii    pblack p_ltpov pctunion rankalfcio prur_farm bls south

      Source |       SS       df       MS              Number of obs =    2417
-------------+------------------------------           F(  7,  2409) =   49.18
       Model |  2177.86158     7  311.123082           Prob > F      =  0.0000
    Residual |  15240.1871  2409  6.32635411           R-squared     =  0.1250
-------------+------------------------------           Adj R-squared =  0.1225
       Total |  17418.0486  2416  7.20945722           Root MSE      =  2.5152

------------------------------------------------------------------------------
    odilwdii |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      pblack |  -1.748068   .4606022    -3.80   0.000    -2.651286   -.8448505
     p_ltpov |  -.7965552   .9714141    -0.82   0.412    -2.701449    1.108339
    pctunion |   .0543957   .0148016     3.68   0.000     .0253706    .0834208
  rankalfcio |   .0000824   .0044358     0.02   0.985    -.0086161    .0087808
   prur_farm |  -3.840118   1.309913    -2.93   0.003    -6.408791   -1.271445
         bls |   .8213967    .074389    11.04   0.000     .6755237    .9672698
       south |  -.6328439   .1759296    -3.60   0.000    -.9778329   -.2878549
       _cons |   1.911806   .4091147     4.67   0.000     1.109553    2.714059
------------------------------------------------------------------------------
```

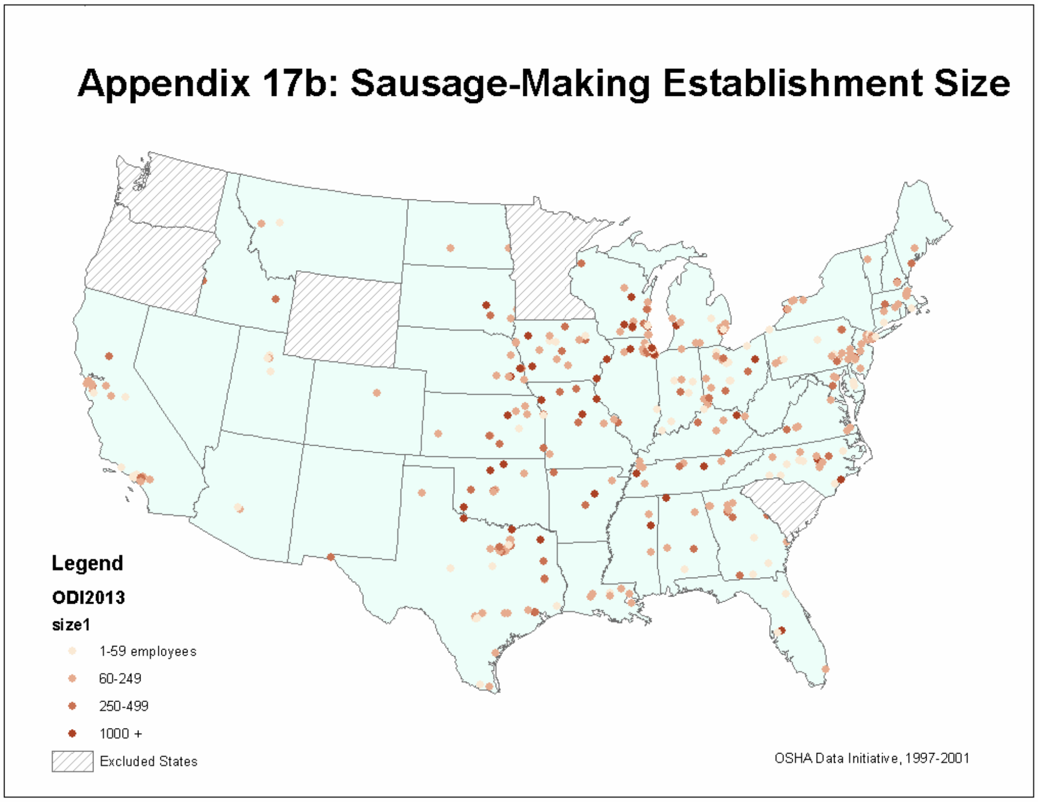
APPENDIX 16: Matching methods for meat database

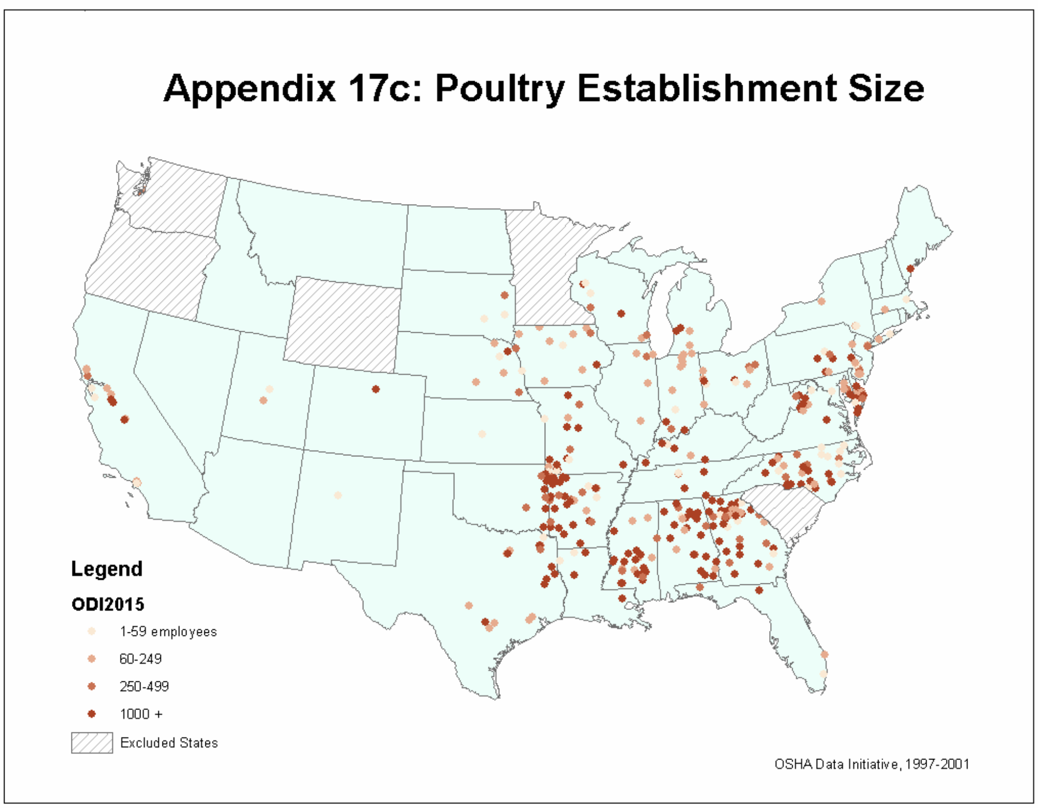
To combine survey years, the meat database of 6,435 survey responses was first collapsed across years to match records with identical names, addresses, and SICs, yielding 2,315 listings. Hand matching combined records with minor differences (such as spelling, data in different fields) to yield 1,553 establishments. Records with different establishment names but the same address were kept separate to account for changed management, except in cases of strong similarity, such as "Waltham Beef & Provision Co" and "Waltham Beef Company Inc." An assistant coder was used, with a sample of 100 records tested to assure intercoder agreement.

APPENDIX 17: Meat: Establishments by size

Meatpacking establishments tended to be especially large in the Midwest and smaller in the South. Poultry establishments were especially large in the South. Sausage establishment size was more widely distributed.

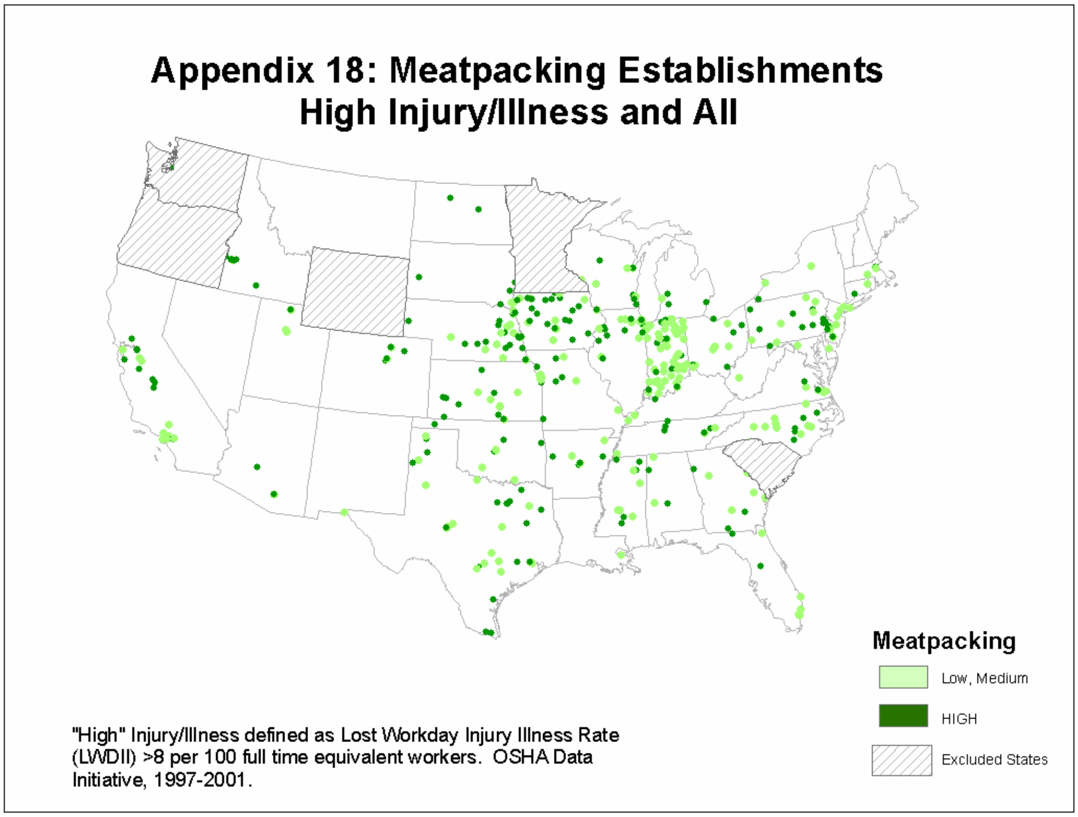


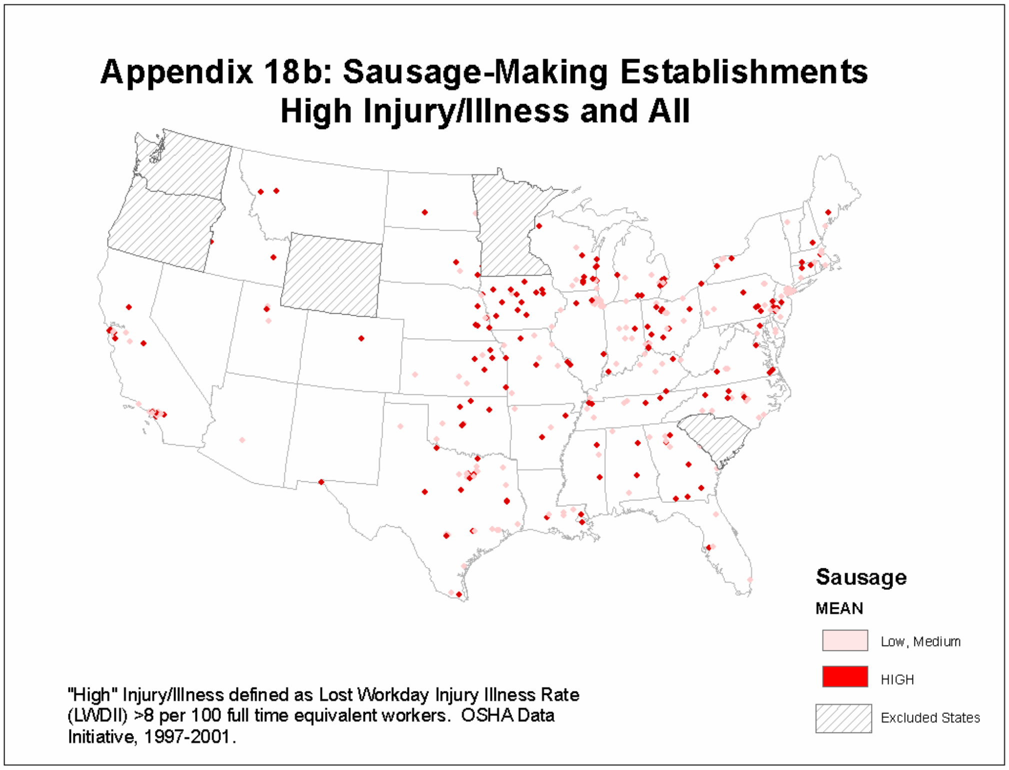

Appendix 17b: Sausage-Making Establishment Size
Legend
ODI2013
size1
1-59 employees
60-249
250-499
1000 +
Excluded States
OSHA Data Initiative, 1997-2001

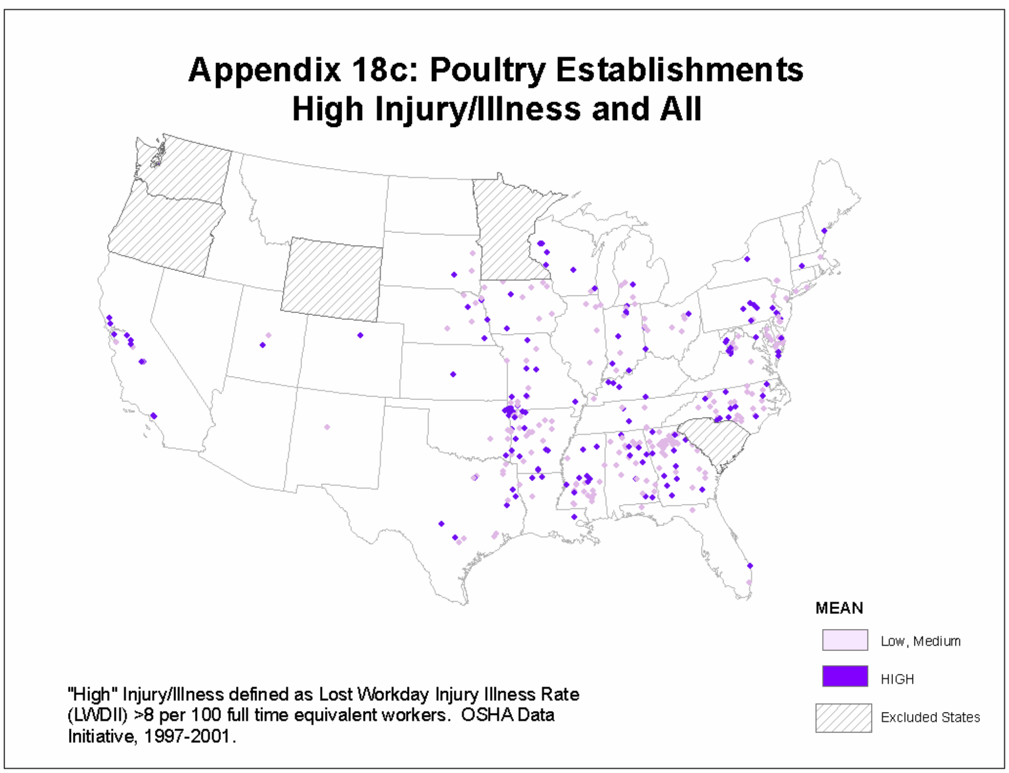

Appendix 17c: Poultry Establishment Size
Legend
ODI2015
1-59 employees
60-249
250-499
1000 +
Excluded States
OSHA Data Initiative, 1997-2001

APPENDIX 18: Meat: Establishments with LWDII >8 and others, by industry

The following maps show high rate establishments versus all establishments, by industry. It is difficult to visually discern patterns in these maps.




Appendix 18b: Sausage-Making Establishments
High Injury/Illness and All
Sausage
MEAN
Low, Medium
HIGH
Excluded States
"High" Injury/Illness defined as Lost Workday Injury Illness Rate (LWDII) >8 per 100 full time equivalent workers. OSHA Data Initiative, 1997-2001.


Appendix 18c: Poultry Establishments
High Injury/Illness and All
MEAN
Low, Medium
HIGH
Excluded States
"High" Injury/Illness defined as Lost Workday Injury Illness Rate (LWDII) >8 per 100 full time equivalent workers. OSHA Data Initiative, 1997-2001.

APPENDIX 19: Correlations between variables used in meat analyses

Analyses were performed to assess the extent to which multicollinearity or high correlation between variables existed in the meat analyses. The VIF and correlation tables below show that while correlations existed, the problem was not so substantial as to preclude the models generated.

A) HIGH RATE ESTABLISHMENTS: VIFs are acceptable and only a few variables are highly correlated.

Variable	VIF	1/VIF
plt_hs	2.31	0.432767
pcapinc	2.20	0.453681
DpctFor	1.22	0.819448
pblack	1.19	0.841398
unmp7596	1.07	0.936130
n	1.01	0.985410
Mean VIF	1.50	

	mean	pblack	pcapinc	plt_hs	unmp7596	n	DpctFor
mean	1.0000						
pblack	-0.1212	1.0000					
pcapinc	-0.1163	-0.1535	1.0000				
plt_hs	-0.0210	0.3577	-0.6509	1.0000			
unmp7596	0.1301	0.0807	-0.2225	0.1044	1.0000		
n	0.3402	-0.0607	-0.0933	0.0486	0.0153	1.0000	
DpctFor	-0.0105	0.0006	0.1999	0.1389	0.0042	-0.0200	1.0000

B) ALL ESTABLISHMENTS
(VIF is a little high when low, med and hi union are all in model)

Variable	VIF	1/VIF
medunion	4.79	0.208731
r_work	3.91	0.255746
pctua	1.95	0.513818
p_bach	1.72	0.582913
p_ltpov	1.66	0.604018
pblack	1.54	0.650786
hiunion	1.41	0.706838
lowunion	1.27	0.786897
pctuc	1.27	0.790085
inspe~100000	1.26	0.796526
unmp7596	1.13	0.882538
Mean VIF	1.99	

Combining the three union categories makes the VIFs all within the acceptable range.

Variable	VIF	1/VIF
pctunion	3.19	0.313147
r_work	3.09	0.323493
pctua	1.94	0.514220
p_bach	1.69	0.592108
p_ltpov	1.57	0.636966
pblack	1.44	0.692607
pctuc	1.26	0.791436
inspe~100000	1.12	0.892384
unmp7596	1.11	0.900754
Mean VIF	1.83	

Several variables have high correlations

	meat	pblack	p_bach	unmp7596	p_ltpov	lowunion	medunion	hiunion	r_work	i~100000	pctua
meat	1.0000										
pblack	0.0917	1.0000									
p_bach	0.1214	-0.0877	1.0000								
unmp7596	-0.1288	0.0216	-0.1636	1.0000							
p_ltpov	-0.1055	$\boxed{0.4222}$	$\boxed{-0.3696}$	0.2564	1.0000						
lowunion	-0.0556	-0.2642	0.0150	0.1007	-0.0312	1.0000					
medunion	0.0603	-0.3673	0.0977	-0.0639	-0.3247	0.2606	1.0000				
hiunion	-0.0092	-0.1289	0.1208	-0.1381	-0.1719	0.0826	$\boxed{0.5076}$	1.0000			
r_work	-0.0100	0.3436	-0.1194	0.0869	0.2151	-0.1946	$\boxed{-0.8514}$	$\boxed{-0.4064}$	1.0000		
inspe~100000	-0.0625	-0.1364	0.1057	-0.1246	-0.1124	-0.2009	0.2862	0.2618	-0.2907	1.0000	
pctua	0.2694	0.1058	$\boxed{0.5611}$	-0.1670	-0.2242	-0.0107	0.0968	0.1063	-0.0969	0.0226	1.0000
pctuc	0.0073	-0.0189	-0.0942	0.0917	0.0556	0.0400	0.0068	-0.0167	0.0026	0.0144	$\boxed{-0.4145}$

CURRICULUM VITAE

Roni Neff, SM (PhD expected 2006)

EDUCATION AND TRAINING

Johns Hopkins Bloomberg School of Public Health, PhD expected, 2006
Department of Health Policy and Management
--Environmental and Occupational Health Policy Concentration

Harvard University School of Public Health, SM 1997
Department of Health and Social Behavior

Brown University, AB 1989
Department of Philosophy

PROFESSIONAL EXPERIENCE

Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD
Research Assistant

- Developed maps and participated in analysis of cumulative risk from toxics following Hurricane Katrina. (2006)
- Supervising team of doctoral students exploring environmental health indicators.(2004-present)
- Researched and co-wrote paper on stakeholder use of scientific information to delay environmental regulation. (2003-2004)
- Interviewed leaders of advocacy organizations regarding priority setting. Co-writing paper. (2002-present)
- Interviewed state and local health department officials regarding their environmental health surveillance activities and assisted with development of influential Pew Environmental Health Commission report on environmental health tracking. (2000-2001, 2004)

Teaching Assistant

- Occupational and Environmental Health Policy (2004)
- Terrorism and Public Health (2004)
- Injury Prevention Summer Institute (2002)

Project Manager (2001-2003)

- Managed research and policy project aimed at developing a framework for public health protection at contaminated military bases.
- Developed case studies of health, environment, and community issues at two contaminated sites. Conducted and analyzed interviews. Designed and conducted survey. Created policy recommendations with principal investigator.

CHANA: Counseling, Helpline & Aid Network for Abused Women, Baltimore, MD (2000 – 2001). ***Assistant Director***

- Collaborated with Director to run small, comprehensive domestic abuse program.

Center for the Advancement of Health, Washington, DC (1997-2000)
Alliance Director

- Directed activities of alliance of scientists working to increase attention to social and behavioral aspects of physical health.

Harvard School of Public Health, Boston, MA (1996)
Research Assistant

- Performed multidisciplinary literature review on economic inequality.
- Conducted media research for project on smoking cessation in pregnancy.

NYC Department of Health, Lead Poisoning Prevention Program, NY, NY (1993-95) ***Senior Public Health Educator***

- Supervised lead poisoning prevention hotline. Coordinated and conducted city-wide presentations for parents, construction workers, educators, property owners, and others.

Voluntary Health Association of India, New Delhi, India. 1993.
AIDS Project Intern

- Wrote materials about AIDS for village health workers.

CHOICE (Concern for Health Options: Information, Care, Education), Phila., PA. 1989-93. ***Reproductive Health/AIDS Hotline Counselor and Resource Specialist; Development Associate***

- Counseled, educated, and provided referrals and case follow-up. Trained staff, supervised calls. Participated in coalitions. Performed qualitative study of impact of policy on clients. Wrote grant proposals, corresponded with donors, planned events.

GRANTS AND FELLOWSHIPS

Institutional Training Grant in Occupational Injury Epidemiology, National Institute for Occupational Safety and Health (NIOSH), stipend and tuition varying by year (2001-2006)

Environmental Public Health Tracking Fellowship, Johns Hopkins Bloomberg School of Public Health, $10,000 (2006)

Innovation Grant, Center for a Livable Future, Johns Hopkins Bloomberg School of Public Health, $20,000 (2004)

Pilot Project Award for thesis research, NIOSH, $10,004 (2004)

PUBLICATIONS – PEER REVIEWED

Neff RA, Goldman LR. Regulatory Parallels to *Daubert*: Stakeholder Influence, "Sound Science," and the Delayed Adoption of Health-Protective Standards. *American Journal of Public Health.* 2005 95: S81-S91.

Litt J, Tran N, Malecki KC, **Neff R**, Resnick B, and Burke T. 2004. Mini-Monograph: Priority Health Conditions, Environmental Data, and Infrastructure Needs: A Synopsis of the Pew Environmental Health Tracking Project. *Environ Health Perspect*. 2004 Oct;112(14):1414-8.

Burke TA, **Neff R**, Chossek K, Tran N. *Strengthening the Role of Environmental Public Health at Contaminated Air Force Bases: Prevention, Evaluation, and Response*. 2003.

Ad Hoc Working Group on Treatment of Tobacco Dependence. Realignment of the Nation's Tobacco Agenda: The Need to Treat Tobacco Dependence. *Preventive Medicine.* 32, 95-100 (2001.) (Coauthor)

Pew Environmental Health Commission. *America's Environmental Health Gap: Why the Country Needs a Nationwide Health Tracking Network*. 2000. (Coauthor)

PRESENTATIONS

Neff R (presenter), Burke T, Curriero F. County-level predictors of occupational injury and illness. Planned presentation, American Public Health Association (APHA), 2006.

Neff, R. Geography of Occupational Injury and Illness in the Meat Products Industries. Center for a Livable Future Research Day. November 30, 2005.

Dreyling EK, Dederick EJ, Chari R, Resnick BA, Malecki K, Burke TA, **Neff R.** Tracking health and the environment: A pilot test of environmental public health indicators. Presentation, APHA, 2005

Neff, R (presenter, session moderator). Geographic mapping: High injury employers and vulnerable workers. Presentation, APHA, 2004

Litt JS, Tran NL, Malecki K, **Neff R**, Resnick BA, Burke TA, Apelberg B, Wismann A, Nachman K. Identifying priority health conditions, environmental data, and infrastructure needs: A synopsis of the Pew environmental health tracking project. Presentation, APHA, 2004

Neff R (presenter), Burke TA, Chossek K, Tran NL. A framework for strengthening the role of public health at contaminated sites. Presentation and Student Award Poster, APHA, 2003

Neff R (presenter), Chossek K, Tran N, Burke TA. Mutual distrust: Case studies in military-community relationships during cleanup of military hazardous waste sites. Presentation, APHA, 2001

Tran N, Burke TA, Chossek K, **Neff R.** Health studies and risk assessments at Military Superfund Sites, is there a disconnect? Presentation, APHA, 2001

Litt JS, Burke TA, Tran NL, Chossek K, **Neff R.** The Environment and Public Health: A Look at the Nation's Ability to Track and Respond to Environmental Hazards, Exposures, and Health Outcomes. Presentation, APHA, 2000

Neff R (presenter). Cultivating Capacity: Advancing NIH Research Training in the Health-Related Behavioral and Social Sciences. Presentation, APHA, 1999

Neff R (presenter). Health-Related Living Conditions in an INS Detention Facility: Report from the Texas-Mexico Border. Presentation, Harvard School of Public Health, 1997.

Neff R (presenter), Chen J, James A, Koch-Weser S, Wiggins N, Kawachi I, Prothrow-Stith D. War on the Poor: Health Hazard. Presentation, APHA, 1996

Neff R (presenter). Public Health Students as Activists, Roundtable, APHA, 1996

Neff R (presenter). RJR's advocacy advertising campaign during the FDA public comment period: a cost, media strategy, and content analysis, Presentation, APHA, 1996

Brown KM, **Neff R.** Development of Targeting Strategies from a New York City childhood lead poisoning prevention media campaign. Poster, APHA, 1995

RECENT HONORS AND AWARDS

Victor P. Raymond Award in Health Policy, Johns Hopkins Bloomberg School of Public Health (2003)

Student Award, American Public Health Association Environment Section (2003)

Honors in doctoral qualifying examination, Johns Hopkins Bloomberg School of Public Health (2002)

PLACE AND DATE OF BIRTH

Philadelphia, PA, May 26, 1967.