USING GIS TO LINK SEER-MEDICARE AND CALIFORNIA PESTICIDE DATA: A POPULATION-BASED CASE-CONTROL STUDY OF PESTICIDE EXPOSURE AND HEPATOCELLULAR CARCINOMA RISK

by

Trang Minh VoPham

BA, Sociology and Environmental Sciences, University of Virginia, 2009

MPH, Epidemiology, University of Pittsburgh, 2010

MS, Geographic Information Science and Technology, University of Southern California, 2014

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Trang Minh VoPham

It was defended on

December 4, 2014

and approved by

Jian-Min Yuan, MD, PhD, Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh

Maria Mori Brooks, PhD, Associate Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh

Evelyn O. Talbott, DrPH, Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh

Joyce Chung-Chou H. Chang, PhD, Professor, General Internal Medicine, University of Pittsburgh

Darren Ruddell, PhD, Assistant Professor, Spatial Sciences Institute, University of Southern California

Dissertation Chair: Joel L. Weissfeld, MD, MPH, Associate Professor, Epidemiology, Graduate School of Public Health, University of Pittsburgh

Copyright © by Trang Minh VoPham

2014

USING GIS TO LINK SEER-MEDICARE AND CALIFORNIA PESTICIDE DATA: A POPULATION-BASED CASE-CONTROL STUDY OF PESTICIDE EXPOSURE AND HEPATOCELLULAR CARCINOMA RISK

Trang Minh VoPham, PhD

University of Pittsburgh, 2014

ABSTRACT

Geographic information systems (GIS), used to analyze spatial data, represent a powerful method to study human health. This research demonstrates the usage of GIS in (1) designing a pesticide exposure metric and (2) linking population-based data sources to conduct an epidemiologic study examining the association between pesticide exposure and hepatocellular carcinoma (HCC).

The first study presents a new GIS method to estimate individual-level agricultural pesticide exposure in California. Landsat remotely sensed satellite images were classified into crop fields and matched to California Pesticide Use Report (PUR) agricultural pesticide application data. Pesticide exposure was calculated using pesticide-treated crop fields intersecting a 500-meter buffer around geocoded locations. Compared to the standard GIS method of matching PUR data to infrequently updated crop land use surveys (LUS's), our method was able to match significantly more PUR temporary crop pesticide applications to Landsat vs. LUS crops (65.4% vs. 52.4%; n=2,466; McNemar's $p<0.0001$).

The second study explored different ways of scaling up Public Land Survey System (PLSS) section pesticide data, the geographic level of reporting for PURs, to the ZIP Code level. We observed substantial agreement between area-weighted ZIP Code pesticide application rates and gold standard census block rates in rural areas (weighted kappa 0.63; 95% confidence interval [CI] 0.57, 0.69). Area weighting was used to estimate pesticide exposure in the third study.

The third (and primary) study was a population-based case-control study examining the association between agricultural pesticide exposure and hepatocellular carcinoma in California via implementing a novel data linkage between Surveillance, Epidemiology, and End Results (SEER)-Medicare and PURs using Medicare ZIP Codes in a GIS. Among rural California residents, previous annual ZIP Code exposure to over 0.06 applied organochlorine pounds per acre significantly increased the risk of developing HCC after adjusting for liver disease and diabetes (odds ratio 1.52; 95% CI 1.02, 2.28; p=0.0415). This is the first epidemiologic study using GIS to examine pesticide exposure and HCC.

The public health significance of this research is related to using epidemiologic, GIS, and biostatistical methods to form a better understanding of pesticides as a potential risk factor for HCC, which is increasing in incidence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

My epidemiology doctoral degree represents an absolute dream come true for me. This is one of the greatest honors of my life. I have had the great privilege of working with brilliant intellectuals and scholars at both the University of Pittsburgh and University of Southern California in executing my dissertation. These individuals display a great rigor and passion for science and expanding the bounds of knowledge through sound and innovative methods and technologies – inspiring me to pursue the same. These individuals have shaped my understanding of epidemiology, biostatistics, and geographic information science (GIS), which will continue to guide my research and practices throughout my career.

First and foremost, I would like to thank my dissertation chair and doctoral mentor, Joel L. Weissfeld. I have had four great years of learning from him with respect to cancer epidemiology and GIS and hope to continue conducting research and science with the same excitement, careful attention, and passion as I have witnessed from him. I would like to thank my doctoral committee – Jian-Min Yuan, Maria Mori Brooks, Evelyn O. Talbott, Joyce Chung-Chou H. Chang, and Darren Ruddell – for their invaluable guidance and support throughout my doctoral studies. I would especially like to thank Maria Brooks for her support while working for the Study of Women's Health Across the Nation (SWAN), which represented a wonderful research experience filled with exceptional faculty and staff. I would like to thank the University of Pittsburgh Cancer Institute (UPCI) Cancer Epidemiology Prevention Program (CEPP) for all

of their help, in particular, Brenda Diergaarde, Brooke Spencer, Pam Sufka, Melissa Turner, and Mary Yagjian. I would like to thank the University of Pittsburgh Department of Epidemiology Small Grants Program and UPCI (in particular Jian-Min Yuan) for financial support in acquiring SEER-Medicare data to execute the primary study of my dissertation.

Finally, I would like to thank the most important people in my life – my family. I dedicate my dissertation to my husband, Matthew David Weaver, the sweetest and kindest man whom I love more than life and who has supported me throughout my graduate career. Thank you. And I would like to dedicate this work to my mother, Phạm Thị Bạch Tuyết, who has taught me about the great virtues that guide all facets of my life – love, respect, kindness, honesty, modesty, and strength. I fortuitously defended my PhD on December 4, 2014, your birthday. Happy birthday, mẹ. I would not be here today without you.

1.0 LITERATURE REVIEW: PESTICIDES AND HEPATOCELLULAR CARCINOMA

1.1 ABSTRACT

Objective: To review evidence regarding the potential risk of hepatocellular carcinoma (HCC) posed by exposure to pesticides in order to highlight future directions for research. Data Sources: PubMed was searched between 1966 and January 2013 combining the following search terms: (Hepatocellular Carcinoma OR Liver Cancer) AND (Pesticides OR Environmental Exposure) AND Humans [MeSH] AND Risk Factors. Study Selection and Data Extraction: The following inclusion criteria were applied: studies reported in English, primary sources, the outcome of interest as hepatocellular carcinoma or primary liver cancer, an exposure of interest as pesticides, and investigation of risk of disease. Results: Of the 365 search results, 17 studies met inclusion criteria. There were 16 case-control studies and one prospective cohort study. Six studies were conducted in China and Vietnam, three studies in Egypt, four studies across Europe, and four studies in the United States. The majority of the studies used self-reported measures to ascertain pesticide exposure and demonstrated that pesticides moderately increased risk of HCC. The strongest evidence of an association was three case-control studies conducted in China showing significant dose-response relationships where increasing serum levels of pesticides, such as dichlorodiphenyltrichloroethane (DDT), significantly elevated HCC risk—odds ratios ranging

between 2.96 and 4.07. Overall, results were inconsistent in reaching statistical significance. Conclusion: There is evidence to suggest an association between exposure to pesticides and developing HCC. Future research should focus on biomonitoring and geospatial-based methods of ascertaining exposure to specific pesticides, as well as on geographic areas with rapidly increasing HCC incidence such as the United States.

1.2 INTRODUCTION: EPIDEMIOLOGY OF HEPATOCELLULAR CARCINOMA

Hepatocellular carcinoma (HCC) is the sixth most common cancer in the world and the third leading cause of cancer-related death (1). Approximately 70 to 85% of primary liver cancer cases are hepatocellular carcinoma (2). Over 80% of HCC cases occur in East Asia and sub-Saharan Africa (3). Using the World Standard Population, age-adjusted HCC incidence in China is 37.4 per 100,000 among males and 13.7 per 100,000 among females (4). However, in traditionally high-risk areas, HCC incidence rates have been stabilizing, potentially due to the hepatitis B virus (HBV) vaccine and reducing dietary exposure to aflatoxin (AFB_1)-contaminated foods, which is a mycotoxin produced by the *Aspergillus* fungus that forms on corn, rice, and peanuts in moist conditions (2, 3, 5). On the other hand, in traditionally low-risk areas, such as the United States, HCC is both the most commonly occurring type of primary liver cancer (84%) (6) and rapidly increasing in incidence (7). Using the 2000 U.S. Standard Population, age-adjusted incidence of HCC in the United States increased three-fold between 1975 and 2005, from 1.6 per 100,000 to 4.9 per 100,000 (8) and reaching 5.7 per 100,000 in 2010 (9). Rising incidence in the U.S. is potentially associated with increasing obesity, hepatitis C virus (HCV) infection rates peaking in the 1960s and 1970s and clinically manifesting 20 years (or more) later, and

improving survival among cirrhosis patients with chronic HBV and/or HCV (2, 3). There were an estimated 28,720 new cases of HCC diagnosed in 2012 in the United States (7).

HCC is more common among males, a potential reflection of males more commonly exposed to risk factors such as alcohol consumption and higher blood testosterone concentrations and more active androgen receptor alleles in male HBV carriers as HCC risk factors (3, 10). HCC occurs more frequently among individuals of Asian descent in the U.S. (3). Racial and ethnic differences in HCC risk indicate geographic variations in prevalence of and acquisition time of risk factors. The mean age of diagnosis is 65 years old (median 64 years) (3, 6, 11). The majority of adult-onset HCC cases occur sporadically, or among individuals with no similarly affect first-degree relative (3). Among patients with localized HCC, treatment includes liver resection or transplantation (8). Chemoembolization, combining chemotherapy and occlusion of the tumor's blood supply, may improve survival in patients with unresectable HCC. However, most patients are diagnosed with advanced stage HCC (8). The five-year relative survival rate in the United States remains is 16.6% (12).

Major risk factors for HCC are a function of geography. In East Asia and sub-Saharan Africa, both high HCC-risk areas, the predominant risk factors are chronic HBV infection and consumption of aflatoxin (3). The main route of HBV infection in Asia is through vertical transmission from mother-to-newborn, while the main route in Africa is through horizontal, sibling-to-sibling transmission at young ages and parenterally (e.g., intravenous drug use [IDU]). It is important to note that 90% of those exposed to risk factors at younger ages, such as through vertical transmission, develop chronic HBV infection (i.e., absence of immunity and longer duration of infection), while 90% of those exposed at older ages have infections that resolve spontaneously. The predominant risk factors in low HCC-risk areas, such as in North America

and Europe, include chronic HCV infection and heavy alcohol consumption (>50-70 grams per day). In the United States, common routes of exposure to HCV occur through IDU, high-risk sexual behavior, and formerly through blood transfusion (13, 14).

In the United States, the population attributable risk of all known risk factors for HCC, including chronic HBV infection, chronic HCV infection, heavy alcohol consumption, type 1 or 2 diabetes, obesity, and metabolic disorders (i.e., hemochromatosis, α-1 antitrypsin deficiency, porphyrias, tyrosinemia, and Wilson disease) after adjusting for age, sex, race is 64.5% (15). The attributable risks of HBV and HCV for HCC in the United States are 5.7% and 20.7%, respectively (16, 17). The population attributable risks of HBV and HCV for HCC in developed countries are 23.3% and 19.9%, respectively (58.8% and 33.4% in developing countries) (17, 18). It is posited that non-alcoholic fatty liver disease (NAFLD), its more severe form of non-alcoholic steatohepatitis (NASH), smoking, and oral contraceptive use are potentially linked to HCC development (3, 19). Most of these risk factors contribute to the formation and progression of cirrhosis (liver scarring) (20). Between 70 and 90% of all HCC cases occur within an established background of chronic liver disease and cirrhosis (3, 20). However, not all individuals with chronic HCV infection develop HCC, which occurs at a rate of 1 to 3% after 30 years (3). Although HCV, HBV, and heavy alcohol consumption are the major risk factors for cirrhosis among individuals with HCC in the United States (6), between 15 and 50% of HCC cases have no established risk factors (3).

1.2.1 Linking Pesticides to Hepatocellular Carcinoma

Previous studies suggest a potential association between pesticide exposure and HCC (21). Pesticides are ubiquitous chemicals used for the treatment of pests such as insects, mice, weeds,

fungi, and microorganisms (e.g., bacteria and viruses) (22). Pesticides are delineated into functional groups, such as insecticides, herbicides, and fungicides, according to the organisms they control (23). Further classifications are derived from chemical classes such as organochlorines. Exposure to pesticides occurs through direct routes (e.g., occupational), which pose higher levels of exposure, versus low-level and more frequent indirect routes (e.g., drinking water, food, air, and dust) (23). Residential proximity to agricultural pesticide applications is a particularly important source of ambient environmental exposure, where pesticides applied from the air and ground may drift from intended sites (24). Additionally, occupational exposure may occur such as from work in agriculture, manufacturing, pesticide application, and landscaping (25). Factors affecting the magnitude of exposure include application techniques and usage of personal protective equipment (25).

Three different pesticide chemical classes have been linked to HCC: (1) organochlorines/organochlorine pesticides (OCPs), (2) organophosphates (OPs), and (3) carbamates (23). OCPs, such as dichlorodiphenyltrichloroethane (DDT) and dieldrin, were used to combat insects (some fungi). Most have been banned in the United States due to environmental persistence and adverse human health effects (26). OPs, such as parathion and diazinon, as well as carbamate pesticides, such as carbofuran, are also used to combat insects (22, 23, 26). Biological plausibility linking pesticides to HCC has been previously demonstrated. Pesticides are hypothesized to contribute to liver carcinogenesis through mechanisms of genotoxicity, tumor promotion, immunotoxicity, and hormonal action (21, 27). Animal models have demonstrated that exposure to DDT and its metabolite, dichlorodiphenyldichloroethylene (DDE), precipitate development of HCC and other liver tumors (28, 29). Subsequent to OCP exposure, the majority of tumors in rodents occur in the liver (27).

Several human studies suggest a potential increase in HCC risk and/or HCC-related mortality due to pesticide exposure. Occupational exposure to DDT among pesticide applicators was significantly associated with higher liver cancer mortality in Italy (30, 31). In the United States, higher adipose DDE levels (32) and increased occupational exposure to the OCPs DDT, aldrin, and dieldrin were significantly associated with liver cancer mortality (33). However, some studies have yielded inconclusive and inconsistent results. Studies conducted in Egypt (34), Japan (35), and the United States (36) have shown that rural residence non-significantly elevated risk for HCC. Liver cancer-related mortality was initially reported to be elevated among males exposed to DDT by way of antimalarial campaigns in Italy (37), but was later observed to be non-significantly protective after further follow-up (38). A case-control study in France showed a non-significant protective effect of general farming occupation after adjusting for age, hospital, and alcohol consumption (39). Similar deficits in HCC risk from farming occupations were demonstrated in studies conducted in New Zealand (40), Japan (41), and the United States (42). Cohort studies and case-control studies conducted in the Nordic countries of Denmark, Finland, Norway, Sweden, and Iceland have demonstrated significant protective effects among individuals working in agriculture in the absence of adjustment for risk factors (43-53).

HCC poses a significant and growing public health burden and studies suggest a potential association between pesticide exposure and risk of developing HCC. The purpose of this literature review was to summarize and evaluate previous research regarding pesticide exposure and primary liver cancer, with a focus on HCC, in order to gain better insight into their potential association and to guide future directions to improve research in this area.

1.3 METHODS

A search of the PubMed database was conducted between 1966 and January 2013 (Figure 1), combining the following search terms: (Hepatocellular Carcinoma OR Liver Cancer) AND (Pesticides OR Environmental Exposure) AND Humans [MeSH] AND Risk Factors. The search yielded 365 results, whose titles and abstracts were scrutinized for the following inclusion criteria: (1) English language, (2) primary source (i.e., excluding reviews), (3) the outcome of interest as HCC or primary liver cancer, (4) an exposure of interest as pesticides, and (5) investigation of the risk of disease. The title and abstract review yielded 14 studies. A review of cited references was also performed. A total of 17 studies, 12 from the PubMed search and five from citation chaining, were included in the literature review.

1.4 RESULTS

1.4.1 Summary

Seventeen studies, 16 case-control studies (54-69) and one prospective cohort study (70), were included in the literature review (Tables 1 and 2). The majority of these studies focused on the outcome of hepatocellular carcinoma. However, some studies investigated a broader outcome of primary liver cancer. Most studies were conducted in geographic areas with high rates of HCC incidence driven by chronic HBV infection—five studies in China, three studies in Egypt, and one study in Vietnam. Although Africa, overall, is characterized by high rates of HBV infection, development of HCC in Egypt is primarily driven by chronic HCV infection (3). Seven of these

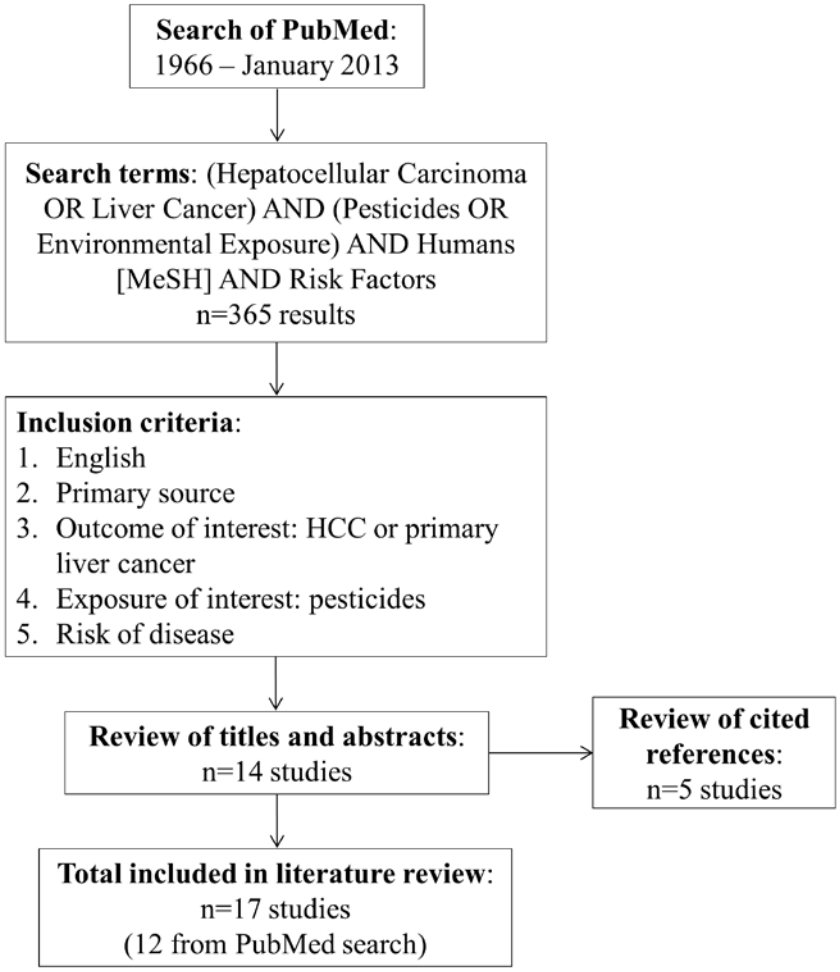


Figure 1. Literature Review Search

nine studies demonstrated that pesticide exposure was significantly associated with increased risk for HCC. The remainder of the studies were conducted in areas with low HCC incidence where chronic HCV infection is typically the primary risk factor for HCC—four studies in the United States, one study in Sweden, one study in Finland, one study in Italy, and one multinational study (Multicentre International Liver Tumour Study [MILTS]) spanning Europe. The majority of these eight studies demonstrated that pesticide exposure non-significantly elevated risk for HCC.

Table 1. Case-Control Studies Examining Pesticide Exposure and Hepatocellular Carcinoma

Authors	Geography	Outcome	Reference Population	Results	Exposure Methods	Major Limitation(s)
Persson et al. (2012)	China (Haimen City prospective cohort)	n=473 HCC cases (1993-2000)	n=492 controls, frequency-matched on age, sex, and area of residence	Highest DDT quintile (≥810 ng/g) significantly increased HCC risk (OR 2.96; 95% CI 1.19-7.40; trend p=0.04)—adjusted for age, sex, area of residence, HBsAg, family history of HCC, history of acute hepatitis, smoking, alcohol, occupation, and continuous DDE levels	Sera collected at enrollment (1992-1993)	Confounding from SES status and AFB_1
Zhao et al. (2011)	China	n=345 HCC cases (2007-2009)	n=961 hospital-based controls	Highest DDT quartile (≥43.09 ng/mL) significantly increased HCC risk (OR 4.07; 95% CI 2.72-6.10; trend p<0.001) Highest DDE quartile (≥10.56 ng/mL) significantly increased risk (OR 1.96; 95% CI 1.39-2.76; trend p<0.001) Highest β-HCH quartile (≥13.43 ng/mL) significantly increased risk (OR 3.67; 95% CI 2.45-5.53; trend p<0.001) All results adjusted for age, sex, education, alcohol consumption, smoking, AFB1-alb, HBsAg, and anti-HCV	Sera collected at enrollment (2007-2009)	Hospital-based, temporality
Soliman et al. (2010)	Egypt	n=150 HCC cases (2007-2009)	n=150 controls, healthy cancer center visitors or those accompanying patients (non-relatives), matched on age and sex	Farmer occupation compared to non-farmer significantly increased HCC risk among anti-HCV-positive individuals (OR 9.60; 95% CI 3.72-24.76); lower risk among HCV antibody-positive non-farmers (OR 4.54; 95% CI 1.82-11.15)—adjusted for age and sex Rural residence non-significantly protective (OR 0.63; 95% CI 0.22-1.83) compared to urban residence; agricultural occupation non-significantly increased risk (OR 2.13; 95% CI 0.89-5.09) compared to administrative occupation—adjusted for HBsAg positivity and/or anti-HCV positivity, schistosomiasis, water pipe use, and smoking	Self-reported agricultural occupation (longest duration) and residential history (longest duration)	Statistical imprecision, selection bias
McGlynn et al. (2006)	China	n=168 HCC cases (1984-2001)	n=385 controls, frequency-matched on sex and age	Highest DDT quintile (>787 ng/g) significantly increased HCC risk (OR 3.8; 95% CI 1.7-8.6; trend p=0.0024)—adjusted for age, sex, HBsAg, residential commune, and continuous DDE levels	Sera collected at enrollment (1984-1985)	Confounding from AFB_1

Table 1 continued

Authors	Geography	Outcome	Reference Population	Results	Exposure Methods	Major Limitation(s)
Ezzat et al. (2005)	Egypt	n=236 HCC cases	n=236 controls; frequency-matched on age, sex, and urban/rural residence	Among rural males and adjusted for HCV RNA, HBsAg, and age: Agricultural pesticides significantly increased HCC risk (OR 2.5; 95% 1.3-5.0) Carbamate pesticides significantly increased risk (OR 2.9; 95% 1.4-5.8) OPs significantly increased risk (OR 2.7; 95% 1.3-5.3)	Self-reported occupation and occupational exposure to pesticides, job-exposure matrix	Recall bias
Porru et al. (2001)	Italy	n=144 male liver cancer cases (n=94 HCC) (1997-1999)	n=283 hospital-based controls, frequency-matched on age, date of admission, and hospital of admission	Field-crop and vegetable farm workers non-significantly increased risk (OR 1.8; 95% CI 0.6-5.6) compared to those not in occupation—adjusted for age, residence, education, HBsAg, anti-HCV, and alcohol consumption Agricultural services industry non-significantly increased risk (OR 1.2; 95% CI 0.5-2.8) compared to those not in industry— adjusted for age, residence, education, HBsAg, anti-HCV, and alcohol consumption	Self-reported occupation, coded using ISCO and ISIC	Recall bias, statistical imprecision, selection bias
Heinemann et al. (2000)	Multicentre International Liver Tumour Study (MILTS): France, Germany, Greece, Italy, Spain, and United Kingdom	n=317 female HCC cases (1990-1996)	n=1,060 hospital-based controls and 719 population controls, frequency-matched on age	Farming occupation non-significantly increased HCC risk (OR 1.34; 95% CI 0.73-2.44) compared to non-farming occupation—adjusted for age, study center, smoking, alcohol, oral contraceptive use, and self-reported HBV or HCV infection Pesticide exposure non-significantly increased risk (OR 1.51; 95% CI 0.57-3.97) compared to no pesticide exposure—adjusted for age and study center Herbicide exposure non-significantly increased risk (OR 1.30; 95% CI 0.81-2.07) compared to no herbicide exposure—adjusted for age and study center	Self-reported occupation and occupational exposure to pesticides, job-exposure matrix	Recall bias, statistical imprecision, selection bias
Badawi et al. (1999)	Egypt	n=102 HCC cases	n=96 hospital-based controls	Pesticide exposure significantly increased HCC risk (OR 2.19; 95% 1.41-3.43) compared to none-- adjusted for sex, age, occupation, smoking, family history of cancer, schistosome infection, and HBV infection (presence of HBsAg, HBsAb, or HBcAb)	Self-reported exposure to pesticides	Confounding from HCV

Table 1 continued

Authors	Geography	Outcome	Reference Population	Results	Exposure Methods	Major Limitation(s)
London et al. (1995)	China (Haimen City prospective cohort)	n=183 deceased male HCC cases (1992-1995)	n=868 controls, matched on age, township of residence, and HBsAg status	Pesticide exposure in the past 5 years non-significantly protective (OR 0.90; 95% CI 0.47-1.40) compared to no pesticide exposure	Self-reported exposure to pesticides	Recall bias
Cordier et al. (1993)	Vietnam	n=152 male HCC cases (1989-1992)	n=241 hospital-based controls, frequency-matched on age, sex, hospital, and residence	Overall exposure to at least one pesticide type non-significantly increased HCC risk (OR 1.2; 95% CI 0.6-2.5) compared to none Exposure to ≥30 liters/year OPs significantly increased risk (OR 4.7; 95% CI 1.1-20.1) compared to none Exposure to ≥30 liters/year OCPs non-significantly increased risk (OR 4.8; 95% CI 0.9-25.1) compared to none Exposure to ≥30 liters/year other pesticides non-significantly increased risk (OR 4.0; 95% CI 0.3-47) compared to none All results adjusted for age, hospital, place of residence, HBsAg, and alcohol consumption	Self-reported occupational exposure to pesticides	Recall bias, statistical imprecision
Kauppinen et al. (1992)	Finland	n=344 deceased HCC cases (1976-1978, 1981)	n=861 controls, deceased stomach cancer or coronary infarction, frequency-matched on age and sex	Other agricultural work significantly increased HCC risk (OR 3.46; 95% CI 1.32-9.10) compared to those not in occupation—adjusted for alcohol consumption	Next-of-kin-reported occupational history, job-exposure matrix, coded using British classifications of occupations and industries	Recall bias, selection bias

Table 1 continued

Authors	Geography	Outcome	Reference Population	Results	Exposure Methods	Major Limitation(s)
Brownson et al. (1989)	United States (Missouri)	n=74 male HCC cases (1984-1988)	n=14,926 other cancer controls	Non-significant increased HCC risk among farmers compared to non-farmers (OR 1.19; 95% CI 0.58-2.37)—adjusted for age	Occupation reported on medical records, coded using U.S. Census Bureau Classified Index of Industries and Occupations (1980)	Exposure misclassification selection bias, confounding
Suarez et al. (1989)	United States (Texas)	n=1,742 deceased male HCC cases (1969-1980)	n=1,742 deceased controls, frequency-matched on age, race, and year of death	Non-significant increased HCC risk among farmworkers compared to those not employed in occupation (OR 1.35; 95% CI 0.82-2.23)—adjusted for age and race	Occupation reported on death certificates, coded using U.S. Census Bureau Classified Index of Industries and Occupations (1980)	Exposure misclassification, confounding from major risk factors
Austin et al. (1987)	United States (five cities)	n=80 HCC cases	n=146 hospital-based controls, matched on age, sex, race, and study center	Non-significant increased HCC risk in agriculture industry compared to no employment in industry (RR 1.1; 95% CI 0.6-2.3) Pesticide exposure non-significantly increased risk (RR 2.1; 95% CI 0.6-6.9) compared to no pesticide exposure—adjusted for herbicide and fertilizer exposure Non-significant trend in years of farming and HCC risk (p=0.22)	Self-reported occupation and occupational chemical exposure, coded using Standard Industrial Classification Manual and Standard Occupational Classification Manual (1977)	Recall bias, confounding from HCV and smoking, statistical imprecision
Hardell et al. (1984)	Sweden	n=102 deceased male primary liver cancer cases (n=78 HCC) (1974-1981)	n=200 deceased controls, matched on age, sex, year of death, and municipality	No significant difference in occupational DDT exposure (farming: 4.8% among HCC cases vs. 10.0% among controls; forestry: 6.0% among HCC cases vs. 4.0% among controls)	Next-of-kin-reported occupational and leisure time pesticide exposure	Recall bias, confounding from major risk factors

Table 1 continued

Authors	Geography	Outcome	Reference Population	Results	Exposure Methods	Major Limitation(s)
Stemhagen et al. (1983)	United States (New Jersey)	n=265 alive and deceased primary liver cancer cases (n=216 HCC) (1975-1980)	n=530 controls, matched on age, race, sex, and county of residence	Among males ever-employed for ≥6 months: significant HCC risk in an agricultural occupation (OR 1.72; 95% CI 1.06-2.79) compared to no employment in occupation Significant risk among farm laborers (OR 3.20; 95% CI 1.11-9.21) compared to no employment in occupation	Self- or next-of-kin-reported occupation and occupational pesticide exposure, coded using U.S. Census Bureau Index of Industries and Occupations (1970)	Confounding from major risk factors

Abbreviations: anti-HCV, antibody to hepatitis C virus; CI, confidence interval; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; HBcAb, hepatitis B core antibody; HBsAb, hepatitis B surface antibody; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCH, hexachlorocyclohexane; HCV, hepatitis C virus; ISCO, International Standard Classification of Occupations; ISIC, International Standard Industrial Classification of All Economic Activities; OCPs, organochlorine pesticides; OPs, organophosphate pesticides; OR, odds ratio; SES, socioeconomic status.

Table 2. Cohort Studies Examining Pesticide Exposure and Hepatocellular Carcinoma

Authors	Geography	Outcome	Reference Population	Results	Exposure Methods	Major Limitation(s)
Evans et al. (2002)	China	n=1,092 deceased HCC cases	n=83,794 (n=58,454 males) subjects from Haimen City prospective cohort	Pesticide exposure not significant and not reported Among males, significant increased HCC risk among peasants (RR 1.5; 95% CI 1.3-1.8) compared to non-peasants—adjusted for age, HBsAg, history of acute hepatitis, family history of HCC, alcohol use, tea drinking, and well water drinking in 1980s	Self-reported occupation and pesticide exposure	Recall bias

Abbreviations: CI, confidence interval; HBsAg, hepatitis B surface antigen; HCC, hepatocellular carcinoma; RR, relative risk.

Overall, the majority of the 17 studies reported an increased risk of HCC associated with exposure to pesticides—over half of which were statistically significant. Franklin and Worgan (2005) distinguished between two types of pesticide exposure methods: *quantitative* pesticide exposure methods include direct biologic measurement of pesticide levels from blood, urine, etc. (i.e., biomonitoring), versus *qualitative* pesticide exposure methods acquiring information using questionnaires, interviews, and/or experts (25). Both quantitative and qualitative methods of pesticide measurement were employed (25), with quantitative assessment being adopted in more recent studies. Most of the studies measured pesticide exposure as self-reported or next-of-kin-reported occupational (or overall) pesticide exposure (54, 59, 60, 62-64, 66, 70). Occupational measures used either job title (e.g., farmer, farmworker, farm laborer, and peasant) or industry (e.g., agriculture and forestry) (57, 59, 61, 64, 65, 67-70). One study used rural residence as an indicator of pesticide exposure (57), and three studies directly measured blood samples to determine biologic levels of pesticide exposure such as from dichlorodiphenyltrichloroethane (DDT) (55, 56, 58). Apart from the studies performing direct measurement, though Persson et al. (2012) additionally analyzed farmer occupation, most studies evaluated more than one measure of pesticide exposure (e.g., job title in addition to self-reported pesticide exposure).

The three studies quantitatively measuring pesticide exposure via blood samples using study populations in China consistently yielded significant results showing pesticides elevating HCC risk (55, 56, 58). On the other hand, results from qualitative methods, although mostly showing increased risk, were inconsistent in reaching statistical significance. The three studies that reported non-significant protective effects used rural residence, self-reported pesticide exposure, and next-of-kin-reported occupation to quantify pesticide exposure (57, 63, 66).

1.4.2 Quantitative Pesticide Exposure Methods

1.4.2.1 China

All three studies employing biomonitoring pesticide exposure methods were conducted in China, where chronic HBV infection and aflatoxin are the major risk factors (3). China has the highest incidence rates for HCC in the world (age-adjusted rate 25.7 per 100,000) (4), as 50% of all HCC cases occur in China (3). DDT, the primary pesticide of interest in these studies, was banned in 1983, although production continues for use in malarial control and production of dicofol, an insecticide (3, 55). Two of the three studies were nested case-control studies utilizing blood samples taken at baseline and measured serum levels of DDT and dichlorodiphenyldichloroethylene (DDE), a metabolite of DDT (55, 56). A third study was a hospital-based case-control that measured serum levels of the OCPs DDT, DDE, and hexachlorocyclohexane (HCH) (58).

McGlynn et al. (2006) conducted a nested case-control study using participants from two randomized controlled trials (RCTs), the Dysplasia Trial (n=3,318) and General Population Trial (n=29,584) in Linxian, China (55). Linxian is not a high risk HCC region. Three hundred eighty-five controls were frequency-matched according to age and sex to 168 HCC cases, diagnosed between 1984 and 2001. Diagnoses were based on pathologic review, biochemical assays, clinical examinations, ultrasonography (US), and computed tomography (CT). Higher serum DDT levels were significantly associated with increased risk for HCC (odds ratio [OR] 3.8 highest quintile [>787 ng/g] versus lowest [<265 ng/g]; 95% confidence interval [CI] 1.7-8.6; trend p=0.0024), adjusting for age, sex, hepatitis B surface antigen (HBsAg), residential commune, and continuous DDE levels. On the other hand, higher serum levels of DDE were non-significantly protective against HCC (OR 0.8 highest quintile [>5,458 ng/g] versus lowest

[<1,767 ng/g]; 95% CI 0.3-1.7; adjusted for aforementioned covariates and DDT level). Stratification according to median DDE levels showed a significantly higher risk of HCC with increasing serum DDT levels when DDE levels were below the median (<2,961 mg/g) (OR 3.55 highest tertile [≥581 ng/g] versus lowest [≤331 ng/g]; 95% CI 1.45-8.74). The authors posited that the presence of this interaction was a result of a higher ratio of DDT to DDE reflecting more recent exposure to DDT, where over time, DDT degrades to DDE. Additionally, this ratio may reflect genetic differences in the ability to metabolize DDT.

These RCTs only screened for esophagus and gastric cardia-related cancers. Therefore, prevalent HCC may have been included in the case group, who may have different pesticide exposure experiences compared to incident cases. All participants of the Dysplasia Trial included individuals with cytologic esophageal dysplasia. Twenty-four of the 168 HCCs originated from the Dysplasia Trial, potentially limiting the external validity of these results. HCV infection, measured using the antibody for HCV (anti-HCV), was not significantly associated with HCC risk in this study and was therefore not included the final multivariable models. Inadequate serum samples did not allow for the evaluation of aflatoxin exposure. Positive confounding may have biased results away from the null by way of higher aflatoxin exposure being associated with pesticide exposure (e.g., handling of contaminated foodstuffs). However, the authors cited previous research reporting low serum levels of aflatoxin in Linxian.

Zhao et al. (2011) conducted a hospital-based case-control study in Xiamen, China, a high-risk HCC area (58). Blood samples were acquired at recruitment. Three hundred forty-five HCC cases, diagnosed between 2007 and 2009, and 961 controls from three different hospitals participated in the study. Higher levels of all three measured OCPs significantly increased the risk of HCC. The highest risk was conferred by higher levels of serum DDT (OR 4.07 highest

quartile [≥43.09 mg/mL] versus lowest [<16.11 ng/mL]; 95% CI 2.72-6.10; trend p<0.0001), followed by serum HCH levels (OR 3.67 highest quartile [≥13.43 ng/mL] versus lowest [<4.20 ng/mL]; 95% CI 2.45-5.53; trend p<0.0001), and serum DDE levels (OR 1.96 highest quartile [≥10.56 ng/mL] versus lowest [<2.62 ng/mL]; 95% CI 1.39-2.76; trend p<0.0001), all adjusted for age, sex, education, alcohol consumption, smoking, aflatoxin (aflatoxin-albumin adducts [AFB_1-alb]), HBsAg, and anti-HCV. After adjustment for age, significant synergism was observed between DDT levels and HBsAg positivity, and polycyclic aromatic hydrocarbon albumin adducts (PAH-alb), between DDE levels and HBsAg positivity, and PAH-alb, between HCH levels and diabetes and AFB_1-alb, and between DDT and HCH. There was also an antagonistic effect between DDT and heavy alcohol consumption. Consistent with McGlynn et al. (2006), there was a significant interaction between DDT and DDE, where increasing DDT levels conferred higher risk of HCC at lower levels of DDE (OR 3.18; 95% CI 2.49-4.06) compared to higher DDE levels (OR 3.01; 95% CI 2.26-4.01).

Persson et al. (2012) conducted a nested case-control study using the Haimen City prospective cohort, which recruited 83,794 subjects from a high-risk HCC area between 1992 and 1993 (56). Four hundred ninety-two controls were frequency-matched according to age, sex, and area of residence to 473 HCC cases diagnosed between 1993 and 2000, ascertained via histology and/or liver imaging, elevated α-fetoprotein (AFP) levels (>400 ng/mL), clinical criteria, or death certificate. Serum levels of DDT were significantly associated with risk for HCC (OR 2.96 highest quintile [≥810 ng/g] versus lowest [≤261 ng/g]; 95% CI 1.19-7.40; trend p=0.04), adjusted for age, sex, area of residence, HBsAg, family history of HCC, history of acute hepatitis, smoking, alcohol, occupation (farmer versus non-farmer), and DDE level. On the other hand, serum DDE levels were non-significantly protective against HCC (OR 0.81 highest

quintile [≥32,222 ng/g] versus lowest [≤10,000 ng/g]; 95% CI 0.33-2.03; adjusted for the aforementioned covariates and DDT level). Most subjects were farmers (85% of cases and 73% of controls). There was also a non-significant antagonistic interaction between sex and DDT, with DDT exposure non-significantly elevating risk for HCC among males (OR 2.93 highest quintile [≥780 ng/g] versus lowest [≤256 ng/g]; 95% CI 0.98-8.75), while non-significantly decreasing risk among females (OR 0.77 highest quintile [≥1,091 ng/g] versus lowest [≤320 ng/g]; 95% CI 0.10-6.03). Similar to McGlynn et al. (2006) and Zhao et al. (2011), there was a borderline significant interaction with increasing DDT conferring higher risk for HCC with decreasing DDE (OR 2.53; 95% CI 1.03-6.24).

There was no information regarding the risk factors of socioeconomic status (SES) and aflatoxin exposure. Although positive confounding may have manifested itself as lower SES associated with greater pesticide exposure (e.g., agricultural occupation), or higher aflatoxin exposure associated with pesticide exposure, mentioned previously, biasing results away from the null, an artificial inflation of the association between pesticides and HCC is unlikely. The authors highlighted their adjustment for geographic area, which may be associated with SES. A previous analysis also did not show a significant association between corn consumption (i.e., main Chinese food staple associated with aflatoxin) and HCC risk. Interestingly, DDT levels were not significantly associated with farming occupation (p=0.88), suggesting that DDT does not explain the significant relationship between farming and HCC risk. However, this finding does lend support to other pesticides potentially underlying the association with HCC. McGlynn et al. (2006) and Persson et al. (2012) demonstrated non-significant protective effects of DDE on HCC risk, while Zhao et al. (2011) showed that DDE significantly elevated risk. These different

results may reflect variability in pesticide exposure across China, with more recent exposure in the Zhao et al. (2011) study population.

1.4.3 Qualitative Pesticide Exposure Methods

1.4.3.1 China

Two earlier nested case-control studies using the Haimen City prospective cohort utilized questionnaire-derived measures of pesticide exposure (66, 70). The case groups for both London et al. (1995) and Evans et al. (2002) were comprised of deceased individuals. Evans et al. (2002) noted that incidence and mortality are identical in the Haimen City area, as effective therapy is not available to most residents. Using a mortality endpoint also protected against differential reporting of incident HCC diagnoses biasing results towards the null, as these diagnoses would have represented non-peasants seeking care outside of Haimen City (70).

London et al. (1995) conducted a nested case-control study with 183 male deceased HCC cases, occurring between 1992 and 1995 (66). Using death certificate information, validated through contacting doctors and reviewing hospital charts, 37% of cases were diagnosed based on elevated AFP levels (>400 ng/mL) and US or CT, 43.2% with US only, 17.5% through clinical examination, and 2.2% with AFP levels >400 ng/mL only. Eight hundred sixty-eight living controls were frequency-matched according to age, HBsAg status, and township of residence. A baseline questionnaire collected information regarding occupation and pesticide exposure. Self-reported pesticide exposure in the past five years, compared to none, was non-significantly protective against HCC in the univariate analysis (OR 0.90; 95% CI 0.47-1.4). Peasant occupation, compared to non-peasant, was significantly associated with increased risk for HCC (OR 1.66; $p<0.0001$), adjusted for the matching factors of age, HBsAg status, and township of

residence, in addition to history of acute hepatitis, family history of HCC, corn consumption in the 1970s, and the interaction between corn consumption and HBsAg.

Peasant occupation is comparable to an agricultural farming occupation. As the Haimen area is largely agricultural, with 91% of cases and 75% of controls employed as peasants, heterogeneity in the exact pesticide-related activities likely exist among those in this occupation (e.g., some peasants do not work with pesticides). Furthermore, the protective effect of self-reported pesticide exposure is also likely affected by recall bias. This is bolstered by the significant association demonstrated in Persson et al. (2012), which also used participants of the Haimen City prospective cohort, but used an objective measure of serum measurement of DDT and DDE.

Evans et al. (2002) subsequently conducted a prospective cohort study using participants in Haimen City, including both males (n=58,545) and females (n=25,340) (70). One thousand ninety-two deceased HCC cases occurring between 1992 and 2000 were included, 8.8% based on histology, 35.8% based on elevated AFP levels (>400 ng/mL) and US, CT, or magnetic resonance imaging (MRI), 39.5% based on imaging, 1.5% based on elevated AFP, 1.7% based on clinical criteria, and 21.5% based on death certificate and/or family and doctor interviews. Among males, peasant occupation, compared to non-peasant, was significantly associated with increased HCC risk (RR 1.5; 95% CI 1.3-1.8), adjusted for HBsAg, history of acute hepatitis, family history of HCC, alcohol consumption, tea drinking, and drinking well water in the 1980s. Peasant occupation was not significant among females (RR 2.5; 95% CI 0.9-4.2). Past pesticide exposure was not significantly associated with HCC among males or females; measures of association were not reported.

A selection bias may have manifested itself in the screening programs for HBV carriers in Haimen City. Evans et al. (2002) excluded 7,024 individuals—91% of whom due to residence in a township with less than 1,000 residents, where mortality reporting is believed to be inadequate. Rural areas are associated with agricultural pesticide use. Although the Haimen City area is predominantly agricultural, exclusion of the most rural areas (i.e., least populated areas) may have biased some of the results towards the null.

1.4.3.2 Egypt

Egypt is characterized by high HCC incidence rates and a high prevalence of HCV infection. The age-adjusted incidence rate of HCC is 14.6 per 100,000 among males and 4.2 per 100,000 among females (4). Three case-control studies demonstrated largely non-significant results—the majority of which showed elevated risk (54, 57, 60). Pesticide exposure measures included self-reported occupation, exposure, residential history, and a job-exposure matrix supplemented with agricultural experts.

Badawi et al. (1999) conducted a case-control study, recruiting 102 HCC cases from the National Cancer Institute outpatient clinic in Cairo, Egypt and 96 controls without any signs of hepatopathology (60). No confirmation of HCC versus non-disease was performed, although elevated AFP levels among cases (224 IU ± 203 among cases; 3.4 IU ±1.5 among controls) and nondifferential carcinoembryonic antigen (CEA) values (3.3 ng/mL ± 8.7 among cases; 2.1 ng/mL ± 4.7 among controls) lent confidence to accurate disease classification. Ninety-six percent of HCC cases had normal CEA values, indicative of primary and not metastatic liver cancer. Pesticide exposure was measured as self-reported occupational history and chemical exposure. Although farming occupation, compared to no farming occupation, was not significantly associated with HCC risk (OR 1.39; 95% CI 0.95-2.04), pesticide exposure,

compared to no pesticide exposure, significantly elevated HCC risk (OR 2.19; 95% CI 1.41-3.43), adjusted for age, sex, smoking, family history of cancer, schistosome infection, presence of HBsAg, hepatitis B surface antibody (HBsAb), or hepatitis b core antibody (HBcAb), and farming occupation or pesticide exposure. Schistosomiasis, also referred to as bilharzia, is an infection due to a parasitic worm and is a HCC risk factor in Egypt (54, 57, 60).

Ezzat et al. (2005) conducted a case-control study in Egypt, which recruited cases from the National Cancer Institute of Cairo University and controls from the Cairo University orthopedic department (54). Pesticide exposure was captured through a job-exposure matrix taking into account self-reported year of agricultural activity, crop types grown, and crop-specific pests that were controlled. Agricultural experts further incorporated information from pesticide registration manuals regarding pests associated with crops and pesticides used for each pest type. Two hundred thirty-six controls were frequency-matched according to age, sex, and rural/urban residence to 236 HCC cases. Included cases were characterized as definite (i.e., pathological or cytological confirmation, or AFP >1,000 ng/mL, or AFP >300 ng/mL and US or CT) or probable (i.e., AFP > 300 ng/mL, US or CT evidence, or treatment). Approximately 47% of HCC cases were pathologically confirmed. Among rural males, exposure to agricultural work-related pesticides significantly increased HCC risk (OR 2.5; 95% CI 1.3-5.0), adjusted for age, HCV RNA, and HBsAg. Specific pesticide chemical classes associated with increased HCC risk were carbamates (OR 2.9; 95% CI 1.4-5.8) and organophosphates (OR 2.7; 95% CI 1.3-5.3). Interestingly, the authors noted that a separate analysis of rural males engaging in agricultural work without the use of pesticides yielded a non-significant risk (OR 1.2; 95% CI 0.53-2.82), which was interpreted as pesticides, and not farming practices, being associated with HCC risk.

Three hundred twenty-two of 866 eligible cases did not participate due to late-stage disease and an additional 77 refused. The participation rate among controls was 76.7%; refusal was predominantly related to refusal to supply a blood sample. However, controls were likely a non-biased sample of the population, as HCV infection rates among participating controls were similar to the general population (54). As advanced HCC cases are likely not represented in the case group, participating cases may have been healthier and were more likely to be employed compared to non-participants. The case group may have been exposed to less pesticides during their employment compared to those excluded, potentially underestimating results. The authors noted recall bias was likely nondifferential, as there were no significant differences in alcohol consumption and tobacco use between cases and controls. However, although alcohol consumption is forbidden according to Muslim tradition, tobacco use may be expected to be lower among cases. Previous research has shown farmers are less likely to smoke compared to the general population (71). If the participating cases who were exposed to farming occupation in this study were more likely to smoke, the results may have been biased towards the null.

Soliman et al. (2010) were unable to replicate the results of Ezzat et al. (2005), which is a potential consequence of both studies utilizing self-reported pesticide exposure metrics (57). Soliman et al. (2010) conducted a hospital-based case-control study in Tanta, Gharbiah, a largely rural region of Egypt characterized by high rates of HCV infection (22% prevalence) and high HCC incidence. The authors used an interviewer-administered questionnaire to collect residential and occupational histories. One hundred fifty controls were matched according to age and sex to 150 HCC cases diagnosed between 2007 and 2009, 27.3% of whom were histologically confirmed. Results derived from occupational measures demonstrated increased, non-significant HCC risks. Agricultural occupation, as the longest occupation, compared to administrative

occupation conferred an increased risk for HCC (OR 2.13; 95% CI 0.89-5.09). Less prominent results were associated with farming (i.e., ever engaged in farming activities) compared to non-farming (OR 1.58; 95% CI 0.79-3.16) and farming exposure (i.e., ever exposed to farming including housewife living in rural area) compared to no farming exposure (OR 1.40; 95% CI 0.71-2.74). However, rural residence (longest residence) compared to urban residence was non-significantly protective against HCC (OR 0.63; 95% CI 0.22-1.83). All results were adjusted for HBsAg positivity and/or anti-HCV positivity, schistosomiasis, water pipe use, and smoking. There was also a significant qualitative interaction, where having a farming occupation is protective in the absence of HCV infection (OR 0.92; 95% CI 0.29-2.93), but elevates HCC risk in the presence of HCV infection (OR 9.60; 95% CI 3.72-24.76), adjusted for age and sex.

All of the pesticide metrics in Soliman et al. (2010) failed to reach statistical significance. Twelve cases and one control were excluded due to refusal to provide a blood sample. If, for example, refusal among cases was associated with poorer health by virtue of residing in areas with greater agricultural pesticide use and poor access to healthcare, this differential selection bias would have biased results towards the null. Information regarding aflatoxin and alcohol consumption was unavailable. However, the authors stated that the Egypt's climate is not ideal for fungal growth and alcohol consumption is not permitted in Muslim cultures. Measures derived from occupational histories were non-significantly elevated, while rural residence was protective. Soliman et al. (2010) stated that the effect of rural residence was explained by medical conditions and personal behaviors. However, all measures, whether derived from occupational or residential history, are non-specific with respect to actual pesticide exposure. Farming-related work spans a wide range of activities that may or may not involve pesticide exposure. Furthermore, the study area of Gharbiah was largely rural (36.4%), and the only

information used to derive this measure of rural residence was the participants' living histories. Pesticide use likely varies over the entire region, and the particular HCC cases recruited for this study may live in truly rural areas that more frequently participate in livestock-related rather than in agricultural activities. The protective effect of rural residence was non-significant and is a potential artifact of the residential classification.

1.4.3.3 Vietnam

Vietnam has a high incidence rate of HCC, with an age-adjusted rate of 42.3 per 100,000 among males and 18.5 per 100,000 among females (4). Cordier et al. (1993) conducted a hospital-based case-control study, recruiting 152 male HCC cases diagnosed between 1989 and 1992 from two different hospitals in Hanoi (62). Over 85% of cases were not histologically confirmed, but diagnosed clinically or with echographic suspicion, and elevated AFP levels (≥500 ng/mL). Male controls were admitted to surgery at the hospitals for reasons unrelated to HCC (64% gastroduodenal ulcer, 13% urinary lithiasis, 9% prostatic adenoma, and 8% biliary lithiasis). Controls were frequency-matched to cases according to age and place of residence. Interviewer-administered questionnaires asked participants to self-report occupational history and occupational pesticide exposure (i.e., name of products, quantity used per year, number of days exposed, and number of years exposed). Organophosphates, compared to no pesticide exposure, significantly elevated risk of HCC (OR 4.7; 95% CI 1.1-20.1), while organochlorine pesticides non-significantly increased HCC risk (OR 4.8; 95% CI 0.9-25.1), adjusted for age, hospital, place of residence, HBsAg, and alcohol consumption.

Although the results point to an elevated risk, the wide confidence intervals of the results allude to statistical imprecision of the findings. The external validity is limited, which was by design as authors wanted to recruit males old enough to have been exposed to Agent Orange

during the Vietnam War. The authors stated that excluded cases (history of cancer and/or no histological confirmation and AFP <500 ng/mL) were slightly younger and reported less pesticide use compared to included cases, which indicates a potential bias away from the null. This selection bias also potentially limits the external generalizability of the findings, where included cases may represent advanced-stage HCC, indicating the study results do not represent the pesticide exposure experiences among all HCC cases. Selection bias is also manifest in the control group, with over 60% admitted to surgery for gastroduodenal ulcers, which occurs among individuals with high tobacco use and lower alcohol consumption. The effects of this selection bias depend on these risk factors' association with pesticide use. Farmers are less likely to use tobacco, which indicates the control group may have been exposed to less pesticides compared to the general population, which would have biased results away from the null.

1.4.3.4 Europe

There is significant variation in HCC incidence rates across Europe. Overall, the age-adjusted incidence rate of HCC is 6.7 per 100,000 among males and 2.3 per 100,000 among females (4). Rates are heterogeneous across different countries, 13.4 per 100,000 among males in Italy (4.4 per 100,000 among females) and 3.2 per 100,000 among males in Sweden (1.8 per 100,000 among females). The results from these studies were largely non-significant; most demonstrated elevated risk, though some showed deficits in risk (63-65, 67). The exposure measures were all qualitative, most of which were based on self-reported occupation. Two studies used the British job-exposure matrix.

Heinemann et al. (2000) conducted a hospital-based case-control study using the Multicentre International Liver Tumour Study (MILTS), originally designed to examine the relationship between sex steroid hormones (e.g., oral contraceptives [OCs]) and sex steroid drugs

containing chlormadinone acetate (CMA) (64). MILTS spanned six countries, with seven clinics in Germany, one in the United Kingdom, one in France, one in Italy, one in Greece, and one in Spain. Three hundred seventeen female HCC cases were included in the study—prevalent cases diagnosed between 1990 and 1994 and incident cases diagnosed between 1994 and 1996. Definite (histologically confirmed) and probable (US, CT, or MRI, and elevated AFP levels >500 ng/mL) were included. On average, four controls were frequency-matched to cases according to age (n=1,798). The majority of controls were derived from hospitals and the remainder from the general population through citizen registers. Exposure was ascertained as self-reported lifetime history of occupations, self-reported exposure to chemicals, and lifetime exposure calculated using a British job-exposure matrix (JEM). The British JEM required jobs to be coded using the British Registrar Generals' Classification of Occupations and the 1968 Classification of Industries. Compared to non-farming occupation, farming (overall) non-significantly elevated risk for HCC (OR 1.34; 95% CI 0.73-2.44), while specific farming occupations showed higher risks (farm owner OR 1.96, 95% CI 0.81-4.77; farm laborer OR 1.43, 95% CI 0.19-18.30), adjusted for age, study center, smoking, alcohol consumption, OC use, HBV or HCV infection. Self-reported occupational exposure to pesticides non-significantly elevated risk (OR 1.51; 95% CI 0.57-3.97), adjusted for age and study center. Using the British JEM, exposure to herbicides non-significantly increased risk (OR 1.30; 95% CI 0.81-2.07). Overall, the results point to higher risk for HCC among farm owners, who may have had more contact with handling pesticides.

This study was limited in statistical power to detect an association between pesticides and HCC. As 136 of the 317 HCC cases were too ill to participate or deceased, next-of-kin interviews were conducted. Although the authors noted that inclusion of hospital-based controls

would minimize differential recall bias, the presence of recall bias is not completely mitigated, as it is likely to have non-differentially biased results towards the null. Blood samples were not collected for population controls; therefore, HBV and HCV information was missing. In addition, 57 cases were missing serology information. As the proportion of cases and population-based controls truly infected with HBV and/or HCV was underreported, this differential information bias could have potentially biased results towards the null. Including prevalent cases introduces a potential incidence-prevalence bias, as prevalent cases may have different pesticide exposure experiences compared to incident cases.

1.4.3.5 Sweden

Hardell et al. (1984) conducted a case-control study of 102 deceased, histologically confirmed, male liver cancer cases that were diagnosed between 1974 and 1981 and reported to the Swedish Cancer Registry (63). Eighty-three cases were HCC, 15 cases were intrahepatic cholangiocellular carcinoma, three cases were hemangiosarcoma, and one case was unspecified sarcoma of the liver. Five cases of mixed HCC and cholangiocellular carcinoma were included in the HCC group. Two hundred and six deceased controls from the National Population Register were matched to cases according to age, sex, year of death, and municipality. Deaths from cancer and suicide were excluded. Exposure was collected as self-reported occupation and occupational chemical exposure. Exposure to DDT from a farming occupation was lower among HCC cases compared to controls (4.8% among cases versus 10% among controls). Exposure to DDT from a forestry occupation was slightly higher among HCC cases compared to controls (6.0% among cases versus 4.0% among controls). Self-reported exposure to DDT was not significantly different between cases and controls, which did not warrant further investigation.

Recall bias, particularly from questionnaires conducted with relatives for the deceased study subjects, likely contributed to a misclassification of past occupational exposure, which may explain why a smaller proportion of cases were exposed to DDT in the farming occupation. No details were given regarding the composition of the control group; it is unknown whether or not any causes of death may potentially share risk factors with HCC such as pesticides. Selection bias may have contributed to a larger proportion of relatives for controls reporting higher DDT exposure from farming.

1.4.3.6 Finland

Kauppinen et al. (1992) conducted a case-control study in Finland (65). Three hundred forty-four deceased cases (International Classification of Diseases, Ninth Revision [ICD-9] code 155.0) reported to the Finnish Cancer Registry between 1976 and 1987, and 1981 were included. Confirmed cases were included, but the exact nature of confirmation was not detailed. The control group was derived from stomach cancer cases reported to the Finnish Cancer Registry in 1977 and deceased coronary infarction patients from the same hospitals as the cases. Four hundred seventy-six stomach cancer controls and 385 coronary infarction controls were frequency-matched to cases according to age and sex. Exposure was collected as self-reported occupational history and using a British job-exposure matrix (JEM). Occupations were coded using the British Classification of Occupations and Industries. Although results were significantly elevated for one occupational group, the majority of the results were not significant, likely due to lack of statistical power. Other agricultural workers compared to those not in the occupation significantly increased liver cancer risk (OR 3.46; 95% CI 1.32-9.10), adjusted for alcohol consumption. Among females, risk was non-significantly elevated (OR 2.01; 95% CI 0.69-5.82); seven of the ten liver cancer cases that were other agricultural workers were females.

The majority of females in this occupational class were milkmaids. Two of the three males were farm laborers and the other male was a garden worker. However, a protective effect was demonstrated among male farmers (OR 0.77; 95% CI 0.51-1.15) and among farmers' wives (OR 0.63; 95% CI 0.42-0.95). Using the JEM, specific exposure to herbicides was not elevated beyond an OR 1.5 to warrant further analysis.

Results were adjusted for alcohol consumption, but not for other established risk factors. Residual confounding due to lack of adjustment may have biased the results towards the null, though the authors did note that HBV infection is rare in Finland. Selection bias from utilizing cancer and deceased controls may have also biased results towards the null, as these controls may share a common risk factor in pesticides precipitating their adverse health outcomes. Furthermore, all information was acquired through the closest traced relatives, likely to not have accurately recalled the study subjects' occupational history.

1.4.3.7 Italy

Porru et al. (2001) conducted a hospital-based case-control study in the Brescia province of Italy (67). This region is highly industrialized (e.g., metal and agricultural industries) and is associated with a high rate of HCC. One hundred forty-four male liver cancer cases (ICD-9 codes 155.0 and 155.1) admitted to two different hospitals were diagnosed between 1997 and 1999. Sixty-eight percent of cases were confirmed through histology, 2.5% through cytology, 5.6% through elevated AFP levels (≥400 ng/mL), and 23.6% through US or CT imaging. Ninety-six percent of all liver cancer cases were confirmed as HCC via histology. Two hundred thirty-eight controls, admitted to various departments in the two hospitals, were frequency-matched to cases according to age, date of admission, and hospital of admission. Controls were admitted to hospitals for the following issues: genitourinary (22.6%), digestive (21.9%), circulatory (21.5%) and respiratory

(5.6%) systems, dermatologic (5.3%) and metabolic (4.9%) disorders, traumas (4.2%), ill-defined symptoms (12.4%), and other conditions (1.4%). Through questionnaires, subjects self-reported lifetime occupational history for each job lasting more than one year and occupational exposure to chemicals, particularly pesticides (i.e., OPs, OCPs, arsenicals, and herbicides). Job titles were coded using the International Standard Classification of Occupation (ISCO) and International Standard Industrial Classification of All Economic Activities (ISIC). Using information from the questionnaire, chemicals were further classified according to their reliability, route, intensity, and frequency. A blinded occupational physician utilized self-reported information, as well as other relevant occupational information, to convert job titles to specific chemical exposures. Field crop and vegetable farmworkers (ISCO 622) compared to those not within this occupation were at a non-significantly increased risk for liver cancer (OR 1.8; 95% CI 0.6-5.6). Agricultural services (ISIC 1120) compared to those not within this occupation were also at a non-significantly elevated risk for liver cancer (OR 1.2; 95% CI 0.5-2.8). All results were adjusted for age, education, residence (city of Brescia versus rest of province), HBsAg, HCVAb, smoking habits, and heavy alcohol consumption. Less specific occupational groupings that may or may not have included individuals exposed to pesticides indicated a protective effect (agricultural and livestock production [ISIC 1110]; OR 0.8; 95% CI 0.4-1.7). When analyzing time since first employment (TSFE) for field crop and vegetable farmworkers, the results were unstable, but showed a slightly higher risk for liver cancer with greater TSFE (OR 1.9 ≥30 years TSFE, 95% CI 0.5-6.5; OR 1.2 <30 years TSFE, 95% CI 0.1-21.5). Analysis according to duration of exposure for agricultural workers did not show higher risk for liver cancer with 20 years or more employment versus less than 20 years (data not

shown). Furthermore, as few subjects reported exposure to pesticides (2 cases versus 12 controls), the authors did not analyze specific occupational pesticide exposure.

This occupational, hospital-based case-control lacked the statistical power to detect any real association between pesticides and liver cancer. Although some results indicated an elevated risk, albeit non-significant, recall bias of occupation and specific exposure to chemicals may have non-differentially biased results towards the null. However, the authors did note that the number of reported occupations was similar between cases and controls (mean number of jobs 3.68 among cases versus 3.53 among controls). Selection bias may have occurred in utilizing a hospital-based control group. Although liver neoplasms were excluded, these subjects are potentially unhealthier compared to the general Brescia province population, potentially as a result of pesticide exposure—which would have further biased results towards the null. Although 70% of cases were histologically or cytologically confirmed, those with and without confirmation were similar with respect to HBV, HCV, alcohol consumption, and sociodemographic characteristics. Selection bias may have further biased results towards the null as 96% of the histologically confirmed cases were HCC (65.3% cases overall). Pesticides may be specifically linked to HCC, and inclusion of other liver cancer types may have diluted its effect.

1.4.3.8 United States

The United States is a low-risk area, with HCC occurring at a rate of 5.9 per 100,000 in 2010 (4). One of four studies conducted in the United States demonstrated a significantly elevated risk for HCC associated with agricultural occupation (68). The remainder of the studies showed non-significantly increased risk for HCC (59, 61, 69). Exposure metrics included self-reported occupation and exposure, in addition to occupation derived from medical records or death certificates.

Stemhagen et al. (1983) conducted a case-control study in New Jersey between 1975 and 1980 (68). Liver cancer cases were derived from three sources: New Jersey hospitals between 1975 and 1978, the New Jersey Cancer Registry between 1978 and 1980, and through death certificates between 1975 and 1979 (ICD codes 155.0 and 155.1). Death certificates with underlying causes of death as primary liver cancer were verified through medical records. Two controls were matched to each case according to age, race, sex, and county of residence. Living hospital controls were matched to live cases, and death certificate controls were matched to deceased cases. Controls with diagnoses or underlying causes of death due to liver cancer, hepatitis, cirrhosis, and other liver diseases were excluded. Exposure was collected through interviews as occupational histories and coded according to the U.S. Census Bureau Index of Industries and Occupations. Two hundred sixty-five cases were included in the study, 11 of which were alive for an interview and 254 of which were deceased (i.e., next-of-kin were interviewed). Approximately 82% of live and deceased cases were histologically confirmed as HCC. A total of 530 controls were included in the study, 22 of which were hospital controls and 508 from death certificates. Among males employed for more than six months, all agriculture (industries and occupation) compared to never employment in the industry significantly increased HCC risk (OR 1.72; 95% CI 1.06-2.79). Agricultural production and services (exclusive of horticulture), compared to never employment in the industry, also significantly increased risk for HCC (OR 1.76; 95% CI 1.09-2.86). Farm laborers, compared to those not in the industry/occupation, were at a non-significantly increased risk for HCC (OR 1.49; 95% CI 0.90-2.46). When evaluating duration of farming employment, more than half of the person-years contributed by cases were before 1930, while the person-years contributed by controls were at a later time period (1930-1939). The authors pointed out that cases may have been

exposed at earlier ages compared to controls. After adjusting for alcohol consumption, the OR for liver cancer associated with the farm laborer occupation increased to 2.14 (no confidence interval provided).

Utilizing hospital and death certificate controls potentially introduced a selection bias underestimating the true effect of pesticide exposure. The control group may have overrepresented smokers and heavy consumers of alcohol, as smoking and alcohol consumption did not significantly differ between cases and controls. Most controls died of ischemic heart disease, lung cancer, and cerebrovascular disease. The majority of cases and controls were deceased, which required use of next-of-kin interviews. Recall bias likely misclassified occupational information, biasing results towards the null. Lack of adjustment for known risk factors, such as HCV, also likely biased results towards the null.

Austin et al. (1987) conducted a case-control study of subjects recruited from one of five study centers across the United States: University of Alabama at Birmingham, University of Miami, Duke University, University of Pennsylvania, and Harvard School of Public Health (59). Of the 86 HCC cases recruited, 80 were histologically confirmed and six were clinically diagnosed. Two hospital controls were matched to each case according to age, sex, race, and study center. Controls were admitted to the hospitals for cancers of the lung, oral cavity, esophagus, larynx, bladder, or pancreas. One hundred forty controls of 161 eligible controls with employment information were included in the analysis. Exposure was collected as self-reported exposure to select chemicals (at least 3 hours/week for at least 6 months during work/leisure time), and any self-reported occupation held for at least 6 months over a lifetime. Occupations were coded according to the Standard Industrial Classification Manual (1972) for industries and the Standard Occupational Classification Manual (1977) for job titles. Occupation in the

agricultural industry compared to those never employed in this industry slightly, though non-significantly, increased risk of HCC (OR 1.1; 95% CI 0.6-2.3). Occupation as a farmer or farmworker also slightly increased HCC risk (OR 1.4; 95% CI 0.7-2.9). There was no significant association between years of farming and HCC risk (trend p=0.22). Exposure to pesticides compared to no exposure non-significantly increased HCC risk (OR 2.1; 95% CI 0.6-6.9), adjusted for fertilizers and herbicides.

Recall bias of occupations and occupational chemical exposure may have biased results towards the null. The small number of exposed subjects makes it difficult to derive inferences from the results. The authors did note that adjustment for HBsAg and alcohol consumption did not change the results. However, blood specimens were only collected for 49 of the 86 cases and 59 of the 161 controls (72). This information bias may have underestimated the effect of pesticides on HCC risk. Selection bias of hospital-based controls is problematic if the controls share risk factors HCC such as pesticides. However, controls with hospital admissions related to chronic bronchitis, emphysema, or primary liver diseases were excluded.

Brownson et al. (1989) conducted a case-control study of 15,000 white males reported to the Missouri Cancer Registry (61). Men at least 20 years old at diagnosis and reported beginning in 1972 were included in the study. Hospital records with occupational data were extracted and coded using the 1980 U.S. Census Bureau codes. A total of 1,720 were registered farmers or farmworkers between 1984 and 1988. There were 74 liver cancer cases (ICD for Oncology [ICD-O] code 155) overall, 11 of which occurred among farmers and farmworkers. The control group was comprised of all other cancers. Farmers and farmworkers, compared to non-farming occupations, were at a non-significantly increased risk for liver cancer (OR 1.19; 95% CI 0.58-

2.37), adjusted for age. When stratified by age group, older white males were at increased risk (OR 1.32; 95% CI 0.61-2.78) compared to younger males (OR 0.64; 95% CI 0.03-4.45).

Brownson et al. (1989) evaluated a large number of cancers and were thus not well-powered to assess the potential relationship between farming occupation and liver cancer. Using a cancer control group may have introduced selection bias, as some cancers may share common risk factors with liver cancer. For example, there were 2,480 prostate cancer cases; prostate cancer has been previously associated with pesticide exposure (73), potentially biasing results towards the null. Although the authors adjusted for age, residual confounding due to lack of adjustment for other major risk factors may exist. The authors also lacked information regarding type of farming and duration. Subjects may have held farming occupations before the 1984 to 1988 registration period, misclassifying exposure and diluting the effect of farming occupation.

Suarez et al. (1989) conducted a death certificate-based case-control study in Texas between 1969 and 1980 using the Texas Bureau of Vital Statistics (69). A total of 1,742 male liver cancer cases (ICD-8 and -9 code 155.0) were included. Controls (n=1,724) were frequency-matched to cases according to age, race, ethnicity, and year of death. Causes of death in the control group included ischemic heart disease (40.4%), cerebrovascular diseases (12.1%), and external injuries (11.7%). Causes of death related to any cancers, liver diseases, gallbladder diseases, infectious hepatitis, and alcoholism were excluded from the control group. Exposure was acquired from the occupation recorded on the death certificates and coded according to the U.S. Census Bureau Classified Index of Industries and Occupations (1980). The overall effect of farming occupation, compared to those not in the occupation, non-significantly increased risk of liver cancer (OR 1.02; 95% CI 0.82-1.28), adjusted for age and race. Farmers were at a

decreased risk (OR 0.96; 95% CI 0.75-1.23), while farmworkers were at an increased risk for liver cancer (OR 1.35; 95% CI 0.82-2.23).

Reliance on death certificates for occupational information is subject to misclassification, biasing results towards the null. As the case group was derived from death certificates, the external validity of the results shifts towards the impact of farming occupation on dying from liver cancer/advanced-stage liver cancer, reflecting different pesticide exposure experiences compared to other HCC cases. Furthermore, underreporting of liver cancer deaths may have misclassified some liver cancers into the control group. However, exclusion of causes of death due to liver-related diseases minimized this occurrence. Lack of adjustment, apart from age and race, for major HCC risk factors is also a limitation of the study.

1.5 DISCUSSION

There is evidence to suggest an association between exposure to pesticides and an increased risk of developing hepatocellular carcinoma. Of the 17 studies included in the literature review, three studies conducted in China using biomonitoring to quantify pesticide exposure demonstrated the strongest evidence for a potential association, particularly with DDT (55, 56, 58). The remaining studies used at least one qualitative method of exposure ascertainment such as self-reported occupation and/or exposure. These studies were largely inconsistent in their results. Although most of the studies demonstrated an elevated risk for HCC, many failed to reach statistical significance, and three studies showed non-significant deficits in risk.

Three specific pesticide classes were associated with increased risk of HCC. The organochlorine pesticides (OCPs) DDT, its metabolite of DDE, and HCH, are potentially

associated with HCC (55, 56, 58, 62). OCPs were frequently used in the United States, but many, such as chlordane, have been removed from the U.S. market due to environmental persistence and adverse health effects (74). Broad groupings of carbamates and organophosphates (OPs) were also associated with HCC risk (54, 62). Both carbamates and OPs are harmful to the human nervous system. However, OPs are not environmentally persistent.

Despite evidence suggesting a potential association between pesticides and many human health outcomes, such as HCC, pesticides remain pervasive in use (23). Global pesticide expenditures exceeded $35 billion in 2006 and $39 billion in 2007, equated with approximately 5.2 billion pounds of used pesticides (75). United States usage encompassed 22% of the world market. Usage remains high, though current regulation is in place in the United States through the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) of 1979, which charged the Environmental Protection Agency (EPA), Food and Drug Administration (FDA), and the U.S. Department of Agriculture (USDA) with the regulation of production and use of pesticides (75). In both 2006 and 2007, a total of 5.1 billion pounds of pesticides were used in the United States (17% conventional use, e.g., treat insects). The agricultural sector has played a prominent role in pesticide usage over time. In 2007, the agricultural sector accounted for 80% of conventional pesticide usage (684 million pounds), followed by non-agricultural sectors (i.e., industry, commercial, or government; 12% or 107 million pounds), and home and garden (8% or 66 million pounds).

Overall, the quality of the three studies using biomonitoring, in terms of adjustment for established risk factors and method of exposure ascertainment, truly bolster confidence in their findings that link pesticides to HCC. Biomonitoring allows for the examination of specific pesticides, rather than heterogeneous groupings of pesticides, that may not directly affect the risk

of developing HCC. Although Persson et al. (2012) did not adjust for HCV infection, McGlynn et al. (2006) showed that, although conducted in a low-HCC risk region, HCV was not a significant risk factor. HBV, compared to HCV, typically drives HCC risk in East Asia (3). McGlynn et al. (2006) also posited that undernourishment in the source population in Linxian, which has been associated with elevating carcinogenic risk subsequent to DDT exposure, might have led to the significant results. However, these significant findings were later replicated in Zhao et al. (2011) and Persson et al. (2012). McGlynn et al. (2006) and Persson et al. (2012) measured serum DDT and DDE levels at study baseline, demonstrating temporality. All three studies also showed a significant dose-response relationship between DDT and HCC risk. Furthermore, the evidence of a biodegradation ratio of DDT to DDE, manifest as a significant interaction of higher risk of HCC with increasing serum DDT levels as DDE levels decreased, were consistent across studies (borderline significant in Persson et al. (2012)). The magnitude of the measures of association for DDT ranged between ORs of 2.96 and 4.07, which likely reflect heterogeneity in the distribution of risk factors and differences in pesticide application practices across China. Results from these three studies do not likely represent an artificially inflated association between pesticides and HCC. Neither McGlynn et al. (2006) nor Persson et al. (2012) were able to determine aflatoxin exposure. Positive confounding from higher aflatoxin exposure associated with pesticide exposure, such as through pesticide-handling peasants also handing aflatoxin-contaminated foodstuffs, may have biased results away from the null. However, McGlynn et al. (2006) cited previous research reporting low serum levels of aflatoxin in the study population, and Persson et al. (2012) cited a previous analysis showing no association between corn consumption (commonly associated with aflatoxin) and HCC risk.

On the other hand, a limitation across all three studies involves the accuracy of case ascertainment. McGlynn et al. (2006) and Persson et al. (2012) were unable to histologically confirm all cases of HCC. Zhao et al. (2011) did not offer any details regarding case confirmation, although they were all derived from hospital settings. Inclusion of cases not truly diseased could potentially bias results towards the null, especially if pesticides are specifically associated with HCC-type primary liver cancer. Another consideration is the external validity of the findings. In the United States, the 95th percentile of serum DDT levels was 28 ng/g between 1999 and 2000 and 19.5 ng/g between 2003 and 2004 (26). This is considerably less than levels showcased in the studies conducted in China. The study population in Persson et al. (2012) was characterized by DDT levels with a geometric mean 468 ng/g ± 18 among cases and 478 ng/g ± 18 among controls. This likely reflects China's continuing use of DDT for non-agricultural purposes (56).

The remainder of the 17 studies were inconsistent in their findings. The majority of the studies demonstrating a statistically significant association between pesticide exposure and increased HCC risk were conducted in East Asia and Africa (Egypt), while the majority of studies with non-significant findings were conducted in Europe and the United States. Three studies, using either rural/urban residence or self-reported pesticide exposure and occupation, demonstrated a non-significant protective effect. Five primary issues likely contributed to these inconsistent results: selection bias in sampling of controls, confounding resulting from lack of adjustment of HCC risk factors, lack of histological confirmation of cases, potential overadjustment, and most importantly, inadequate measures used to indicate pesticide exposure.

Six case-control studies used hospital-based controls (58-60, 62, 64, 67). Usage of hospital controls, who likely represent less healthy individuals compared to the general

population, limits the external validity of the findings. Soliman et al. (2010) conducted a case-control study in Egypt using healthy visitors and individuals accompanying patients to a cancer center. Brownson et al. (1989) conducted a case-control study in the United States (Missouri) using other cancer controls, the majority of which were lung cancer (27.6%), followed by prostate cancer (16.6%). However, there is evidence of an association between pesticide exposure and prostate cancer (73). Usage of hospital controls, visitor controls, and other cancer controls likely biased results towards the null as a result of potential shared common risk factors with HCC such as pesticide exposure.

Four studies utilized deceased controls as comparative groups for deceased cases (63, 65, 68, 69). These studies cited ethical reasons (e.g., HCC being associated with poor survival) and data limitations (e.g., usage of death certificates) as motivating factors. Deceased controls died from stomach cancer or coronary infarction (65), ischemic heart disease, cerebrovascular diseases, and external injuries (69), excluding causes of death related to liver cancer, hepatitis, cirrhosis, and other liver diseases (68), or from unspecified causes (63). Inferences derived from these studies must consider the generalizability of the findings reflecting the relationship between pesticide exposure and dying from HCC, or developing advanced-stage HCC associated with higher mortality risk. These results may also be biased towards the null when the underlying causes of death among controls share pesticide exposure as a risk factor. The most important consideration is reliance on interview-based measures of pesticide exposure, particularly next-of-kin-reported information, subject to recall bias. However, London et al. (1995) and Evans et al. (2002) justified their usage of deceased HCC cases and live controls in their studies in Haimen City, China by stating that the incidence rate reflects the mortality rate in this region (i.e., extremely poor survival and poor availability of treatment options).

Many studies limited in available data did not adjust for established risk factors of HCC. Badawi et al. (1999) did not adjust for HCV, which is a major risk factor of HCC in Egypt. Among studies conducted in the United States, Brownson et al. (1989) presented age-adjusted results, Suarez et al. (1989) presented results adjusted for age and race, Austin et al. (1987) presented unadjusted results (matched controls according to age, sex, race, and study center), and Stemhagen et al. (1983) presented unadjusted results (matched controls according to age, sex, race, and county of residence). Depending on the geographic area, the impact of residual confounding from lack of adjustment varies for different risk factors. For example, lack of adjustment for HCV infection for studies conducted in the United States may have resulted in negative confounding. HCV, by virtue of its main routes of transmission through intravenous drug use and high-risk sexual behavior (13, 14), may be a more frequent urban phenomenon. Pesticide exposure occurs more frequently in rural areas (25), therefore lack of adjustment for HCV potentially biased results towards the null—underestimating the true effect of pesticide exposure.

The majority of studies were able to identify HCC cases through histological confirmation, clinical examinations, imaging (e.g., ultrasonography, magnetic resonance imaging, and computed tomography), and elevated alpha-fetoprotein (AFP) levels, or a combination of different methods. Histological confirmation through biopsy or surgical resection is ideal (76). Therefore, other methods of confirmation introduce the potential misclassification cases and controls. Additionally, usage of cases derived from the underlying cause from death certificates is problematic in underreporting, whether if due to the absence of an autopsy, coding errors, absence of medical records for the certifying physician, etc. (77, 78). Selection bias may manifest itself as inclusion of metastatic liver cancer in the case group, biasing results away or

towards the null depending on its association with pesticide exposure. A related issue is inclusion of prevalent HCC cases, which would bias results if they differ from incident cases in their pesticide exposure experiences. Perhaps individuals with prevalent HCC have survived due to minimal pesticide exposure relative to the typical HCC case.

Another consideration in interpreting the results from these studies is the potential presence of overadjustment, or inclusion of variables highly correlated with the exposure of interest (79). Specifically, pesticide exposure is a largely rural phenomenon, as agricultural activities are more common in less densely populated areas (25). Adjustment for variables that are inherently geographic, such as rural or urban residence, may produce comparable cases and controls, but may also adjust away the effect of measures designed to capture pesticide exposure such as agricultural employment. This effect dilutes the association between pesticides and HCC, biasing results towards the null. Similarly, some studies, by design, matched on geography-related variables such as area of residence (54, 56, 62, 63, 66, 68). For example, Porru et al. (2001) adjusted for geographic area (residence in the city of Brescia versus outside of the city) when evaluating the association between self-reported occupation and HCC. A significantly larger proportion of cases (52.8%) lived outside of the city of Brescia compared to controls (31.1%). Results for field crop workers and vegetable farmworkers and the agricultural services industry were not significant. If the geographic areas outside of the city of Brescia are predominantly rural and used for agriculture, controlling for residence may have resulted in overadjustment, biasing results towards the null. On the other hand, several studies were able to achieve statistically significant results despite adjustment for geographic-related variables. Persson et al. (2012) matched on and adjusted for area of residence, Cordier et al. (1993) matched on and adjusted for place of residence, McGlynn et al. (2006) adjusted for residential

commune, and Badawi et al. (1999) adjusted for farming occupation and pesticide exposure in the same multivariable logistic model. This potentially highlights the variable effect on results depending on how pesticide exposure was measured and the study area. For example, delineations for place of residence in Vietnam may more accurately reflect variation in healthcare rather than rural versus urban demarcations. Furthermore, usage of more accurate, quantitative methods of pesticide exposure estimation is less subject to misclassification compared to a qualitative method such as self-reported occupation. The effect of overadjustment would be expected to be more pronounced when an exposure metric more accurately reflects pesticide exposure. Closer examination of overadjustment is warranted.

The major limitation of the majority of the evaluated studies was how pesticide exposure was measured. Apart from the three studies measuring blood samples, all of the studies utilized qualitative methods of pesticide exposure ascertainment that contributed to the inconsistent results, in terms of significance and direction of effect. These methods included self-reported pesticide exposure, self-reported occupation, job-exposure matrices, occupational experts, and rural residence. If the study subject was deceased, next-of-kin were interviewed. Occupations were often coded using established standards such as the International Standard Classification of Occupations (ISCO) for occupations and the International Standard Industrial Classification of All Economic Activities (ISIC) for industries.

The inherent limitation of all of these methods is the potential for recall bias. Inaccuracies in recalling past occupations and exposure to particular pesticides were likely nondifferential between cases and controls, biasing results towards the null. However, it cannot be ruled out that misclassification was differential in favor of better recall among cases that artificially inflated results. Reliance on next-of-kin reported information could have engendered even more

pronounced exposure misclassification. Furthermore, the exposure definitions lend themselves to misclassification as being heterogeneous groupings that do not truly reflect the pesticide exposure experienced by the study subjects. Specifically, focusing on occupations of farmers, which may include farm laborers and farm owners, may be problematic in that farm laborers may directly apply pesticides, but may not accurately recall the exact name or quantity of the chemical to which they were exposed, while farm owners may purchase the pesticide product. Using a definition of rural residence masks the heterogeneity from residing in areas that are proximate to urban areas or vice versa, and rural areas not near agricultural lands applying pesticides. Most studies using qualitative measures, such as ever-employment, did not attempt to examine historical pesticide exposure. Pesticide exposure may have spanned a relatively long time period during which an individual was employed, or where an individual resided if examining ambient exposures, and recent exposure may be irrelevant to hepatocarcinogenesis.

These limitations reflect the pervasive inadequacies present in the majority of studies attempting to ascertain pesticide exposure (23). Most pesticide exposure-focused case-control studies have utilized questionnaires, and sometimes medical records, with questions regarding occupational histories, tasks, residences, and/or past pesticide exposure (25). Accurately quantifying pesticide exposure, particularly when investigating their role in chronic diseases, such as cancer, must consider historical exposures to take into account latency periods, or the time between from the initial exposure to clinical disease (23, 80). Pesticide exposure is ideally collected before disease onset in order to minimize recall bias from self-reported data and demonstrate a temporal relationship (25). When evaluating agricultural pesticide use, many pesticides are applied throughout a growing season, often in combination, or tank mixes, which is also dependent on crop type (25). Common routes of exposure should also be considered such

as directly through occupation, and indirectly such as through food consumption (23). Specific types of pesticides may also be associated with disease, which are not adequately captured with dichotomous (yes/no) classifications.

Yet methods of ascertaining pesticide exposure have progressively improved over time. The evolution of measuring pesticide exposure began with crude surrogates, such as rural residence and farming occupation, to sophisticated techniques involving direct measurement of biological samples. Pesticide exposure methods are categorized as either qualitative or quantitative (25). Qualitative methods include occupational history, job exposure matrices, self-reported exposure, and expert-reviewed estimation. Quantitative methods include usage of exposure databases, empirical modeling of exposure determinants, and biological measurements. This progressive improvement is evident in the 17 studies included in the literature review, with studies favoring more sophisticated and accurate quantitative methods over time.

Qualitative methods, such as those involving questionnaires, are more reliable among farmers compared to farmworkers, who are often personally involved in the purchase and application of pesticides (23, 25). A weakness of occupational history is the lack of identification of specific chemicals (25). Self-reported history of pesticide exposure may overcome this weakness, but is also associated with potential recall bias (25). Another method is to incorporate a review of self-reported pesticide exposure by experts, such as occupational hygienists, as a way to evaluate self-reported exposures in relation to other study subjects. There are also integrated exposure metrics, such as job-exposure matrices (JEMs) and exposure intensity algorithms, which incorporate data from different sources to improve the estimation of pesticide intensity. JEMs utilize data regarding job title, tasks, and industry, but their reliability and validity vary according to the available supplementary data such as regional job data. Exposure intensity

algorithms derive estimates of intensity and dose by weighting cumulative pesticide exposure by chemical- and applicator-specific information. On the other hand, quantitative methods include empirical modeling, which uses information on determinants of pesticide exposure (e.g., weather conditions, work practices, and time since application) to estimate potential dermal exposure (25). Biomonitoring/biological samples, such as OCP levels in adipose tissue, can be used to assess long-term exposure—particularly those with long biological half-lives and whose concentrations may not be affected by disease when studying chronic health outcomes (23, 25). However, direct measurement poses challenges in terms of time and cost.

A promising quantitative method is using geographic information systems (GIS) to objectively ascertain past exposure (23, 81, 82). A GIS is a system of hardware and software that can be used to manage, visualize, and analyze spatial (i.e., with location information) data (83). For example, a GIS-based method developed by Rull and Ritz (2003), validated using serum levels of DDE among study subjects in California's Central Valley, incorporates multiple years of pesticide application records and land use data to ascertain pesticide exposure within a 500-meter buffer around a residence. In conjunction with information regarding other routes of potential exposure, such as from occupations, GIS-based methods offer an accurate, objective, and cost-effective alternative to measuring exposure. The general population may be unaware of their ambient pesticide exposure. Given available data, a GIS can reconstruct historical pesticide exposure relevant to chronic diseases associated with long latency periods, otherwise difficult to perform with other methods. The effects of specific pesticides can be examined. Furthermore, usage of a GIS-based method would address the paucity of public health research using geospatial methods (84).

Given the inconsistent results and limitations of the existing literature investigating the relationship between pesticide exposure and HCC, future research should focus on quantitative measures of exposure ascertainment, specifically focusing on plausible pesticides and/or chemical classes. Recall bias associated with self-reported measures of occupation and exposure can be avoided through use of biomonitoring and/or geospatial methods. Quantitative methods should also consider multiple routes of exposure such as through residential proximity to agricultural pesticide applications and occupation. Usage of histologically confirmed cases of HCC minimizes misclassification of disease. Adjustment for established risk factors HCC, which are a function of the geographic region in which the study is conducted, is important. Confounding due to lack of adjustment for HCC risk factors that are inherently geographic is particularly problematic such as HCV infection in the United States. Overadjustment from including variables closely correlated with pesticide exposure is a potential issue that also deserves further attention.

Future research should also allow for the examination of the relationship between pesticides and HCC among both males and females. Many of the reviewed studies utilized only male study subjects (61-63, 66, 67, 69), citing few cancer cases developing among females. Results from these studies are limited in their external validity. An exception was the Heinemann et al. (2000) study, whose subjects were derived from a study originally designed to evaluate the effect of oral contraceptive use on HCC among females. The majority of studies were also conducted in high-risk areas in Asia such as China. HCC has been increasing in traditionally lower rate countries, tripling in incidence in the United States since 1975 (8, 9). HCC is the ninth leading cause of cancer-related death in the United States (16). Given between 15 and 50% of individuals developing HCC in the United States do not have any established risk factors (3), it is

important to elucidate the role of potential risk factors of HCC--particularly in countries with rising incidence rates such as the United States.

1.6 CONCLUSIONS

Pesticides remain a ubiquitous environmental exposure, and overall, there is evidence supporting a potential association between pesticide exposure and HCC. The most convincing evidence includes studies conducted with study populations in China measuring pesticide levels in blood samples. Future research should focus on accurate and sophisticated methods of pesticide exposure ascertainment, such as biomonitoring and geospatial-based tools, while considering historical exposure, multiple exposure routes, and the impact of specific chemicals.

2.0 A LANDSAT REMOTE SENSING METHOD TO ESTIMATE AGRICULTURAL PESTICIDE EXPOSURE IN CALIFORNIA

2.1 ABSTRACT

Accurate pesticide exposure estimation is integral to studying pesticides and human health. Standard GIS methodology to estimate agricultural pesticide exposure in California matches pesticide applications to the most current land use survey (LUS) in Public Land Survey System sections. LUS's are intermittently updated and concurrent Landsat images may better capture land use changes. The results of using 1985 Landsat images, classified via maximum likelihood and per-field, and the 1990 LUS to estimate pesticide exposure in Kern County, CA in 1985 were compared. The Landsat and LUS methods separately matched pesticide applications to a crop (tier 1), all crops in a section (tier 2), or a section (tier 3). Pesticide application rates for residential parcels were calculated. The Landsat method achieved significantly more tier 1 matches compared to the LUS method, notably among temporary crops. Our novel Landsat method can improve crop identification for use in studies to reconstruct pesticide exposure.

2.2 INTRODUCTION

Pesticide exposure has been associated with adverse human health outcomes such as cancers (23). Pesticides are chemicals designed to treat pests such as insects and herbs (85). One source through which pesticide exposure may impact human health is via residential proximity to agricultural applications of pesticides (24, 86). Applied pesticides may drift through the air and the ground and through post-application volatilization (87). Gunier et al. (88) demonstrated that pesticides measured in carpet dust from 89 residences in California were significantly correlated with residential proximity to agricultural pesticide applications quantified using a geographic information system (GIS) (Spearman correlation coefficients 0.23 to 0.50; $p<0.05$). Humans are subsequently affected by pesticides through dermal contact and ingestion, especially as pesticides are less likely to degrade within houses (89).

Elucidating the exact role pesticide exposure may play in the risk of developing adverse health outcomes is impacted by the methods used to quantify exposure. GIS metrics can combine multiple data sources to reconstruct historical exposure to specific pesticides (90). The California Department of Pesticide Regulation (CDPR) has collected Pesticide Use Report (PUR) data pertaining to agricultural use pesticide applications since 1974, including pounds of pesticides used to treat specified crop types on specified dates within Public Land Survey System (PLSS) sections (91). However, PUR data alone cannot be used to match pesticide applications to specific geographic locations at a scale finer than the 1 mi^2 PLSS section level. This limitation has motivated attempts to combine PURs with land use data, notably the California Department of Water Resources (CDWR) land use surveys (LUS's). Rull and Ritz (24) developed the standard validated GIS method of estimating agricultural pesticide exposure in California via a three-tier methodology that assigns PUR pounds of applied pesticides to LUS crop fields (92).

The notion of “tiers” refers to the level of certainty with which a PUR pesticide application can be assigned to a particular LUS crop field (93). Combining PURs with a LUS enables pesticide application rate calculations at geographic scales finer than the PLSS level. However, CDWR LUS’s are infrequently conducted on a county basis once every seven to 10 years, during which time significant land use changes can occur (94).

Although vector data have typically dominated this research, raster data provide a valuable way to incorporate temporally current land use information in pesticide exposure estimation. Ward et al. (86) pioneered the integration of Landsat remote sensing, which has continuously captured satellite imagery of the Earth since 1972 (95), in estimating pesticide exposure. Supervised classification of a Landsat image of Nebraska from 1984 was implemented to classify agricultural land cover types, which were subsequently assigned crop-specific pesticide use probabilities. Maxwell et al. (96) demonstrated how Landsat imagery of Fresno County, CA could be used to downscale the identification of PUR pesticide-treated crop fields below the LUS level (minimum mapping unit 0.003 mi^2) (94). Normalized Difference Vegetation Index (NDVI) values, a measure of vegetative density, were used to classify imagery into crop fields via a minimum distance method, and when used in conjunction with PLSS sections, can identify probable crops treated with pesticides (97).

However, minimum distance classification is not widely used in practice as it cannot take into account the spectral variability present within land use classes (98). Alternative approaches include implementing per-pixel maximum likelihood classification (MLC) using NDVI values (99, 100) and/or per-field classification, which is useful in addressing within-field spectral heterogeneity (101). For example, Turker and Ozdarici (102) implemented per-field classification, where ML-classified pixels of SPOT, IKONOS, and QuickBird imagery of Turkey

in 2004 were used to classify vector fields according to the most frequently occurring land use class pixel.

To estimate exposure to agricultural pesticides in a year without a concurrent LUS, we developed and evaluated a method to link PUR pesticide application data to concurrent Landsat images classified via a maximum likelihood and per-field classification approach using NDVI land use signatures. Our first research objective was to execute an accuracy assessment comparing classified Landsat images in 1990 to the 1990 LUS gold standard. As part of this first objective, we determined the accuracy of 1990 agricultural pesticide exposure estimates using classified Landsat images from 1990 vs. the 1990 LUS. Our second research objective was to evaluate the crop specificity of 1985 pesticide applications matching to classified Landsat images. As part of this second objective, we compared pesticide exposure estimates derived from 1985 pesticide application data matched to classified Landsat images from 1985 vs. the 1990 LUS.

2.3 METHODS

2.3.1 Study Area

Kern County, CA is one of 19 counties in California's agriculturally intensive Central Valley, ranked third in agricultural sales in the state (Figure 2) (103, 104). The population in Kern County in 2012 was an estimated 856,158, 10% of whom resided in rural areas (105, 106). The largest city is Bakersfield, with a population of 358,597 (106). Kern County is 8,131.92 mi^2 in area. Agricultural croplands are predominantly found in the central and northwestern portions of

the county (Figure 2). The Sierra Nevada mountain range is found towards the east. In 2011, over 28 million pounds of pesticide active ingredients (AIs) were used in Kern County, the second highest total among counties in California (107).

According to the U.S. Department of Agriculture (USDA) Census of Agriculture, between 1982 and 1992, both the number of farms with cropland (1,685 to 1,522) and the associated farm acreage (1,002,811 to 963,761 ac) decreased in Kern County (108). The majority of this acreage was associated with harvested cropland (76.6-86.7%), which was consistently dominated by cotton (34.6-36.9%). According to the Kern County Agricultural Crop Report, the majority of the harvested cropland (823,974 ac) in 1985 was comprised of cotton (33.4%), alfalfa hay (10.0%), grapes (9.5%), almonds (9.4%), wheat (6.4%), barley (3.4%), potatoes (3.2%), and navel oranges (2.1%) (109). Twenty-six percent (213,140 ac) of the harvested cropland in 1985 was associated with permanent crop acres (does not require seeding after each harvest (110)) such as grapes. In addition, 2,292,000 ac were devoted to pasture. The majority of the harvested cropland (914,893 ac) in 1990 was comprised of cotton (34.1%), alfalfa hay (11.9%), grapes (8.2%), almonds (8.0%), wheat (3.3%), carrots (3.2%), potatoes (2.6%), navel oranges (2.2%), and barley (2.0%) (109). Twenty-four percent (217,684 ac) of the harvested cropland in 1990 was associated with permanent crop acres and a total of 2,096,713 ac were devoted to pasture.

Among the prevalent crops in Kern County, the following planting and harvesting dates are typically observed: cotton (plant Mar 20 to May 1; harvest Sep 15 to Nov 15); alfalfa hay (Sep 1 to Oct 31; Mar 20 to Oct 31); raisin grapes (Jan to Feb; Aug 15 to Sep 20); wine grapes (Mar 15 to Apr 15; Aug 15 to Oct 25); table grapes (May 1 to Jun 15; Jun 25 to Nov 24); almonds (Jan to Feb; Aug 4 to Oct 15); wheat (Nov 15 to Jan 15; Jun 1 to Jul 10); spring/summer-harvested carrots (Nov 15 to Mar 20; Apr 1 to Jul 15); fall/winter-harvested

carrots (Jul 1 to Sep 1; Oct 15 to Mar 15); fall-harvested potatoes (Aug 1 to Aug 30; Nov 25 to Mar 10); summer-harvested potatoes (Nov 20 to Mar 15; Apr 15 to Jul 15); oranges (Mar to Jun; Oct 15 to Sep 30); and barley (Nov 15 to Jan 15; May 20 to Jun 25) (111).

2.3.2 Data Sources

The CDPR PUR database provided pesticide application data, which has collected information regarding agricultural pesticide use in California since 1974 (91). Until 1989, commercial pest control operators were required to report all pesticide use and farmers were required to report restricted pesticide use (federally and/or state-designated as posing potential public health harm). A full-use reporting system was adopted in 1990. PUR data include the name, pounds, date, crop, and PLSS section associated with reported pesticide applications. The PLSS divides portions of the country into 1 mi^2 sections for surveying purposes (112). Sections are identified by the county, principal meridian, township, range, and section identifier.

Landsat images were requested from the U.S. Geological Survey (USGS) Global Visualization (GloVis) Viewer. The Landsat program was initiated in 1972 by the USGS and the National Aeronautics and Space Administration (NASA) to collect Earth imagery (95). The Thematic Mapper (TM) sensor onboard Landsat 4 and 5 captured seven spectral bands with at least 30 m spatial resolution. Each Landsat scene, defined by a Path-Row designation, spans 185 km and is captured every 16 days. Bands 3 (red; 0.63-0.69 μm) and 4 (near-infrared; 0.76-0.90 μm) were used in this analysis (98). The CDWR conducts LUS's of agricultural lands to monitor land use changes in California on a county basis focusing on over 70 crop types (113). Aerial and satellite images, Global Positioning System (GPS) devices, and ground verification are used to classify land uses. Residential parcels were selected from the 2012 Kern County Assessor file via

use codes (e.g., 0100=single family residence) (114). U.S. Census Bureau TIGER/Line® files provided administrative boundaries used in creating the figures (115).

2.3.3 Pesticide Data Processing

PUR data in 1990 (used for accuracy assessment) and in 1985 (used for pesticide exposure estimation in a year without a concurrent LUS) were processed using CDPR logic checks (116) such as duplicate removal (117). Outlier application rates in 1990 were defined using CDPR-created flags (118). Outliers in 1985 were defined as pesticide application rates >200 lb/ac (>1,000 lb/ac if fumigant) or pesticide application rates greater than 50 times the median rate for all uses of a given pesticide product, crop, unit type, and record type. Outlier rates were replaced with the statewide median rate for the pesticide AI in that year, and pounds of AI were recalculated using the number of treated acres (24). For this analysis, a PUR application was defined as each unique instance of an organophosphate AI being applied given a particular date, crop, and PLSS section.

A database of pesticides belonging to the organophosphate chemical class was created using agricultural references (23, 27, 89, 119-123). Organophosphates are pesticides pervasively used following the ban of organochlorine pesticides in the 1970s (124). A crosswalk between PUR commodity codes and CDWR LUS crop codes was created to facilitate data linkage. The following PUR extractions were made for inclusion into the analysis: agricultural records, associated with an organophosphate, and applied in Kern County in 1990 or 1985.

2.3.4 Accuracy Assessment in 1990

2.3.4.1 NDVI Signatures From 1990

The entire methodological workflow is shown in Figure 3. A time series of Landsat 4 and 5 TM images captured between January and October 1990 (10 monthly images from Jan 22 to Oct 28; no images available in November and December) was used to create NDVI land use signatures (Appendix A). The earliest available LUS in Kern County, conducted in 1990, served as the ground truth for land use types. Entire or portions of images with excessive cloud cover were excluded. Images from Paths 41 and 42 and Rows 35 and 36 were requested, which cover the geographic extent of Kern County (Figure 4).

Using IDRISI Selva, TM images for the red (R) and near-infrared (NIR) bands were corrected to at-sensor reflectance using published radiometric calibration coefficients and image metadata (126). Atmospheric correction was implemented via the Chavez cosine estimation of atmospheric transmittance (COST) model to address atmospheric effects associated with using multitemporal images to calculate NDVI values (127, 128). Path-Row images were mosaicked. Negative reflectance values were recoded to 0 (129). A median spatial filter (3x3 kernel) was applied to each mosaicked image to reduce random noise (102, 130, 131). NDVI values were calculated using the following equation: (NIR-R)/(NIR+R). NDVI values harness information from wavelengths of electromagnetic radiation absorbed and reflected by green plants and how reflectance patterns change throughout the growing season (132). NDVI values range from -1 (no or sparse vegetation) to 1 (dense vegetation). NDVI images were cropped to the 1990 NDVI signatures extent, which was defined by a region unaffected by clouds and/or shadows and within the 1990 Kern County LUS extent (Figure 5). NDVI images were re-projected to the

California Teale Albers (NAD83 datum; meter) coordinate system (30 m spatial resolution; nearest neighbor resampling to not alter pixels).

From the 1990 Kern County LUS, polygons of single-use (i.e., excluding double- or triple-cropped, intercropped, or mixed), representing classified areas (i.e., excluding not surveyed [NS], entry denied [E], or outside study area [Z]), and within the 1990 NDVI signatures extent were selected (15,997 polygons). A negative buffer (-30 m; spatial resolution of Landsat images) was created around each selected LUS polygon to exclude potential mixed pixels (e.g., raster cells along LUS crop field boundaries potentially associated with multiple crop types, roads, etc.) (133). After collapsing urban LUS polygons into a single category, LUS polygons with valid geometries (15,565 polygons) were intersected with the NDVI images. Land uses represented by fewer than 100 pixels in each month in 1990 were excluded from consideration for training data (134), resulting in 57 distinct land use classes.

2.3.4.2 Classification of 1990 Landsat Images

Site-specific error matrices were created to quantify the accuracy of classified 1990 Landsat images compared to the 1990 Kern County LUS gold standard. Stratified random sampling (SRS) selected 60% of the buffered polygons from the 1990 LUS within the 1990 NDVI signatures extent (133). Strata were defined by the land use classes. Processed NDVI images from January to October 1990 were overlain with the SRS-selected polygons to be used as training data. The remaining 40% of the 1990 NDVI signatures extent was segmented and classified via a maximum likelihood and per-field classification approach. Using ArcGIS 10.1, the NDVI signatures training data were used to execute per-pixel maximum likelihood classification of the monthly 1990 NDVI images within the 40% classification extent. The sample option was selected to assign a priori probabilities to land use classes in proportion to the

number of cells represented in the NDVI signatures training data (135). Using IDRISI Selva, segmentation was performed on the monthly 1990 NDVI images within the 40% classification extent using the following parameters: window of 3, tolerance of 0.01, weight mean factor of 0.5, and weight variance factor of 0.5. Using the ML-classified pixels, per-field classification (using the segments) was implemented based on the modal class or a majority rule (102, 136).

2.3.4.3 Error Matrices

Classified segments and the LUS were intersected and compared by segment and by total acreage using error matrices. Using the mean NDVI value for each land use class for each month, the SAS 9.3 Proc Cluster centroid method grouped together land use classes based on the distance between two clusters, or the squared Euclidean distance between their centroids or means, to form phenological groups (99, 137, 138). Agreement, kappa and 95% confidence intervals (CIs), producer's accuracy, user's accuracy, omission error, and commission error were calculated according to CDWR land use (i.e., specific crop), CDWR broad land use group (e.g., field crops), and phenological group.

2.3.4.4 Pesticide Exposure Estimation in 1990

The three-tier method (24) was implemented to estimate pesticide application rates in 1990 using either classified 1990 Landsat imagery (referred to as the Landsat method) or the 1990 Kern County LUS (referred to as the LUS method). Segments classified as agricultural use (grain, rice, field, pasture, truck, deciduous, citrus, vineyard, or idle) were spatially joined to the 2,337 PLSS sections intersecting the 40% classification extent. Segments were dissolved according to crop type and section, the geographic level of reporting of the PUR database (91). Agricultural use

LUS polygons selected from the 40% classification extent were spatially joined to sections and dissolved according to crop type and section.

Separately for the Landsat method and the LUS method, PUR-derived pounds of applied organophosphates were assigned to crop fields or PLSS sections according to crop type and section of application. Tier 1: Pesticides were matched to a crop field according to crop type and section. Tier 2: Pesticides were matched to all other crop fields in a section. Tier 3: Pesticides were matched to the entire section. A key difference between the established three-tier method and the LUS method implemented in this study was that non-permanent crop fields were not collapsed into a single category to facilitate the examination of specific crop types.

Using the tier-matched organophosphates, pesticide application rates (lb/ac) were calculated for residential parcels separately using the Landsat and LUS methods. SRS selected at most three residential parcel centroids from each of the 2,337 sections (strata) within the 40% classification extent, and 500 m (radius) buffers were created around the centroids of sampled residential parcels. Acreage estimates were derived from Landsat, LUS, and section data. Pesticide application rates were weighted by the proportion of pesticide-treated crop fields and/or sections intersecting the buffer. Rural and urban parcels were identified using the 2000 U.S. Census Bureau Urbanized Areas (UAs) and Urban Clusters (UCs) (139). Parcel centroids intersecting a UA or UC were categorized as urban.

2.3.5 Comparing Pesticide Exposure Estimation in 1985 Using 1985 Landsat Images vs. 1990 LUS

2.3.5.1 Classification of 1985 Landsat Images

Path 41-42 and Row 35-36 Landsat 5 TM images were requested for January through October 1985 (10 monthly images from Jan 31 to Oct 14) to parallel the 1990 NDVI signature time points (Appendix A). Portions of the February 1985 image missing Path 42 (majority of Kern County agricultural fields) were imputed with the average of the January and March 1985 images (140). Images were processed using the same workflow as the 1990 images (Section 2.3.4.1). All training data from the 1990 NDVI signatures extent, as opposed to the 60% sample used during the accuracy assessment, were used to classify the 1985 Landsat images. Using an MLC and per-field (using segments) approach, the 1990 NDVI signatures classified monthly 1985 NDVI images cropped to the 1985 imagery extent, a cloud- and shadow-free area within the 1990 Kern County LUS extent (Figure 5). A priori probabilities from the entire 1990 NDVI signatures extent were utilized. Segments, derived from monthly 1985 NDVI images, were classified (per-field) according to a majority rule.

2.3.5.2 Pesticide Exposure Estimation in 1985

Using the three-tier method (Section 2.3.4.4), PUR data from 1985 were matched to classified 1985 Landsat images (Landsat method) or the 1990 Kern County LUS (LUS method). Agricultural use segments were spatially joined to the 2,491 PLSS sections within the 1985 imagery extent and dissolved according to crop type and section. The 1990 Kern County LUS was selected to calculate comparative pesticide application rates as it was the LUS conducted closest in time to the 1985 PUR data. Agricultural use LUS polygons selected from the 1990

LUS within the 1985 imagery extent were spatially joined to sections and dissolved according to crop type and section. Pesticide application rates were calculated for at most three residential parcels selected via SRS from each of the 2,491 sections (strata) within the 1985 imagery extent.

2.3.6 Statistical Analysis

Bowker's test of symmetry for paired data compared the proportion of tier 1, tier 2, and tier 3 matches when using the Landsat vs. the LUS method. McNemar's tests compared the proportion of tier 1 vs. tier 2 and 3 matches, and tier 1 and 2 vs. tier 3 matches, when using the Landsat vs. LUS method, as well as the proportion of tier 1 vs. tier 2 and 3 matches by crop type according to each method. Wilcoxon signed-rank tests (non-parametric version of paired t-test) compared pesticide application rates estimated using each method, and Spearman rank coefficients quantified the correlation between rates. Weighted kappa coefficients quantified the agreement in pesticide exposure categorizations according to each method. The data analysis was generated using the SAS System for Windows software, Version 9.3 (Cary, NC, USA).

2.4 RESULTS

2.4.1 Accuracy Assessment in 1990: Error Matrices

A total of 898,076.90 ac were used to create NDVI signatures in 1990 (60% SRS of 1990 NDVI signatures extent), which in turn classified 625,760.59 ac. The segments were on average 6.57 ± 4.99 ac (median 5.34) in size, compared to LUS polygons that were on average 84.32 ± 119.61

ac (median 49.30). There was an average of 29 ± 22 pixels (median 24) available to classify each segment.

When selecting the intersections between the segments and the single-use LUS polygons comprising the majority of each segment's original area, CDWR land use (i.e., crop-specific) kappa was 0.700 (95% CI 0.696, 0.703) (top row of Table 3). The highest producer's accuracy was observed for asparagus (15/15=100%). Asparagus-classified segments were on average 7.64 ± 3.30 ac (median 7.53). The highest user's accuracy was observed for cotton (19,059/20,605=92.5%). Cotton-classified segments were on average 5.42 ± 3.48 ac (median 4.67). Kappa statistics improved when aggregating segments and LUS polygons into CDWR broad land use groups (kappa 0.732 [95% CI 0.728, 0.735]) and into phenological groups (kappa 0.779 [95% CI 0.776, 0.782]). Select phenological groups out of a total 17 representing NDVI patterns over a calendar-year time period are shown in Figure 6. All kappa statistics demonstrated substantial agreement beyond chance in terms of the performance of Landsat imagery in classifying the agricultural landscape as indicated by the LUS gold standard (143). Comparable overall percent agreement and kappa statistics were observed when examining the entire acreage of the intersections (bottom row of Table 3).

A closer examination of the accuracy assessment aggregated to CDWR broad land use groups reveals satisfactory producer's and user's accuracy for the majority of the agricultural broad land use groups (Table 4). Producer's accuracy was upwards of 82.6% for pasture crops and user's accuracy was upwards of 88.6% for field crops. Among agricultural broad land use groups, the highest omission error (96%) and commission error (91.9%) was observed for idle (i.e., fallow) lands. Truly idle lands were often misclassified as native vegetation (76.7%), while idle-classified segments were truly field crops (35.6%) or native vegetation (19.8%) - all of

which belong to the same phenological group (Figure 6). It is important to note that among segments that were classified as agricultural use, a high proportion (73.9-98.9%) truly belongs to an agricultural broad land use group as opposed to a non-agricultural group (NV, NW, S, or U) according to the LUS (Table 4).

2.4.2 Accuracy Assessment in 1990: Pesticide Exposure Estimation

According to 7,495 PUR applications in 1990, LUS (mean 1.69 lb/ac; median 0.40 lb/ac) and Landsat (mean 2.31 lb/ac; median 0.46 lb/ac) pesticide application rates in 1990 were not significantly different for the 1,291 sampled residential parcels (Wilcoxon signed-rank $p=0.8513$). Rates were significantly correlated (Spearman correlation 0.83; $p<0.0001$). A similar average number of crop types were present within any given section when using the LUS (mean 1.3; median 1.0) and Landsat layers (mean 2.8; median 2.0). A similar average number of pesticide-treated crop types intersected any given 500 m buffer when using the LUS (mean 1.4; median 1.0) and Landsat layers (mean 1.8; median 1.0).

Using quartiles defined by the distribution of LUS pesticide application rates, agreement between LUS and Landsat pesticide exposure classifications was high (weighted kappa 0.766 [95% CI 0.739, 0.792]) (Table 5). A small proportion of truly non-exposed parcels was misclassified as highly exposed using the Landsat method (14/481=2.9%). A small proportion of truly highly exposed parcels was misclassified as not exposed using the Landsat method (14/323=4.3%). Pesticide application rates were also comparable when stratified by rural/urban status. Among 463 urban parcels, LUS rates (mean 0.78 lb/ac; median 0 lb/ac) were similar to Landsat rates (mean 0.87 lb/ac; median 0 lb/ac) ($p=0.1934$) (weighted kappa 0.742 [95% CI 0.689, 0.796]). Among 828 rural parcels, LUS rates (mean 2.20 lb/ac; median 0.81 lb/ac) were

similar to Landsat rates (mean 3.11 lb/ac; median 0.88 lb/ac) (p=0.3896) (weighted kappa 0.742 [95% CI 0.706, 0.777]).

2.4.3 Land Use Classification and Pesticide Exposure Estimation in 1985

A total of 50,441 segments (mean 34.09 ± 28.30 ac; median 26.69 ac) derived from 1985 Landsat imagery were classified (Figure 7). There was an average of 153 ± 127 pixels (median 120) available to classify each segment. The majority of segments were classified as alfalfa (19.5%), followed by cotton (19.3%), field crop (18.7%), and native vegetation (8.6%).

According to 3,909 PUR applications in 1985, the proportion of tier matches were significantly different when using the Landsat method vs. the LUS method (Bowker's p<0.0001; Table 6). The Landsat method achieved a significantly higher proportion of tier 1 matches (60.3%) compared to the LUS method (57.4%) (McNemar's p=0.0002). The Landsat method (99.2%) achieved significantly more combined tier 1 and 2 matches vs. tier 3 matches compared to the LUS method (96.6%; p<0.0001). Among the 2,466 PUR applications associated with temporary crops (i.e., sown after each harvest; e.g., cotton), the Landsat method (65.4%) achieved a significantly higher proportion of tier 1 matches compared to the LUS method (52.4%; p<0.0001). Among the 1,443 PUR applications associated with permanent crops (e.g., oranges), the LUS method (66.0%) achieved a significantly higher proportion of tier 1 matches compared to the Landsat method (51.6%; p<0.0001).

A larger proportion of PUR applications associated with the following temporary crops were matched at tier 1 to Landsat compared to the LUS: alfalfa (N=468; Landsat 98% vs. LUS 76%; McNemar's p<0.0001), dry beans (N=75; 67% vs. 7%; p<0.0001), cotton (N=792; 97% vs. 83%;p<0.0001), and potatoes (N=300; 65% vs. 51%; p=0.0001). Assuming PUR data are

accurate, a larger proportion of tier 1 matches among temporary crops is indicative of the capacity of Landsat imagery to delineate agricultural use lands not otherwise present in an outdated LUS. For example, Figure 8 shows a residential parcel that was sampled in the 1985 imagery extent and located in section 15M29S25E10. PUR data indicated one organophosphate application of 27.9619 lb occurred in PLSS section 15M29S25E15 (section south of parcel but intersecting parcel buffer) on alfalfa on October 5, 1985. No alfalfa fields were present in this section using the 1990 LUS (left), resulting in a tier 2 match and an estimated rate of 0.82 lb/ac for the selected residential parcel. However, 1985 Landsat imagery (right) identified alfalfa-classified segments in this section, achieving a tier 1 match and an estimated rate of 1.15 lb/ac. The crop types present in section 15M29S25E15 according to the LUS and Landsat methods - alfalfa, cotton, and sugar beet - belong to three different phenological groups, providing support that the alfalfa-classified segments are not a result of phenological misclassification (i.e., misclassification of a segment as another land use belonging to the same phenological group).

PUR applications associated with the following permanent crops achieved more LUS vs. Landsat tier 1 matches: almonds (N=588; LUS 85% vs. Landsat 75%; $p<0.0001$), oranges (N=322; 91% vs. 75%; $p<0.0001$), and peaches/nectarines (N=89; 85% vs. 46%; $p<0.0001$). Further examination of the crops associated with a greater number of LUS tier 1 matches revealed potential phenological misclassification. Among the 58 PLSS sections associated with LUS tier 1 almond matches, but no Landsat tier 1 matches, 95% contained segments classified as alfalfa, 26% with mixed pasture, and 12% with apples - all three of which belong to the same phenological group as almonds (Figure 6). These occurrences demonstrate the difficulty associated with classifying crop types with similar phenological patterns.

LUS rates were significantly different from Landsat rates (Wilcoxon signed-rank p=0.0448; Table 7). Assuming Landsat rates are accurate, LUS rates overestimated exposure for 35.7% of parcels and underestimated exposure for 32.5% of parcels - although differences were typically within 1 lb/ac. A similar number of crops intersected any given section using the LUS (mean 3.1; median 3.0) and Landsat layers (mean 4.0; median 4.0). A similar number of pesticide-treated crops also intersected any given buffer using the LUS (mean 2.2; median 2.0) and Landsat layers (mean 2.7; median 2.0). Pesticide exposure classification according to LUS quartiles (none: 0 lb/ac; low: >0-0.18 lb/ac; moderate: 0.18-0.60 lb/ac; high: >0.60 lb/ac) demonstrated good agreement between both methods (weighted kappa 0.711 [95% CI 0.682, 0.740]).

Stratification according to rural/urban status revealed some differences. LUS (mean 0.29 lb/ac; median 0 lb/ac) and Landsat (mean 0.24 lb/ac; median 0 lb/ac) rates among 452 urban parcels were significantly different (p=0.0363) (weighted kappa 0.649 [95% CI 0.589, 0.708]). As shown in the Bland-Altman plots (Figure 9), LUS rates typically overestimated exposure among urban parcels compared to Landsat rates (mean difference 0.05 lb/ac). LUS (mean 0.64 lb/ac; median 0.31 lb/ac) and Landsat (mean 0.74 lb/ac; median 0.27 lb/ac) rates among 841 rural parcels were comparable (p=0.2553) (weighted kappa 0.694 [95% CI 0.656, 0.732]). Differences were typically characterized by LUS rates underestimating exposure (mean difference -0.10 lb/ac).

Two discrepant parcel pairs associated with large differences in LUS and Landsat rates are presented in Figure 10. The top row of images shows a rural parcel with an LUS rate of 3.24 lb/ac and Landsat rate of 50.73 lb/ac. Oranges were treated with 1,565.47 lb in the section in which the parcel is located (15M29S29E33). The discrepancy in estimated rates is due to the

parcel's buffer intersecting large orange orchards using the LUS (i.e., pesticides distributed over larger area), while a relatively smaller orange orchard intersected the buffer using Landsat. The bottom row of images shows an urban parcel (within the Bakersfield Urbanized Area) with an LUS rate of 21.03 lb/ac and Landsat rate of 2.38 lb/ac. Discrepant rates resulted from almonds treated with 1,160.59 lb in the section in which the parcel is located (15M29S26E27), and the buffer intersecting the entire (and only) almond crop field present in the LUS in this section. The Landsat method classified multiple almond crop fields in section 15M29S26E27.

2.5 DISCUSSION

GIS-based metrics are powerful tools in examining the relationship between pesticide exposure and human health outcomes. Important underlying issues when examining chronic diseases include long latency periods, or the time between initial exposure and the clinical diagnosis of disease, which can be 20 years or more among cancers (85, 144). Historical reconstruction of past exposure is integral in capturing the potential effect of a latency period. Multiple routes of exposure exist such as dermal, inhalational, and oral, and humans are potentially exposed to a wide variety of pesticides at different points in time (90). Recall bias regarding past pesticide exposure may under- or overestimate true exposure. GIS-based pesticide exposure metrics address all of these issues through the capacity to incorporate multiple data sources with locational, dated information and specific chemicals. For example, the standard GIS method used to estimate agricultural pesticide exposure in California combines residential locations with

CDPR PURs and CDWR LUS's (24). Additional information such as biologic samples can be acquired to quantify human pesticide exposure occurring through other routes such as ingestion (90).

However, dynamic agricultural landscapes, as a result of crop rotation and land use conversion (e.g., urbanization) (145), contribute to relevant changes that may impact GIS-based methods of estimating pesticide exposure. CDWR LUS's are intermittently updated every seven to 10 years on a county basis. Pesticide exposure estimation during a year lacking a current LUS will be affected as the utilized LUS may not adequately capture agricultural lands during that particular time period. Methods of incorporating remote sensing such as Landsat, which provide multispectral and multitemporal imagery capable of distinguishing landscape features (95, 146), allow for a useful approach to improving pesticide exposure estimation. The primary strengths of this research include the implementation of an improved MLC and per-field approach to classify Landsat imagery (compared to minimum distance methods) and the demonstration of a linkage between PUR data and agricultural crops derived from Landsat to estimate agricultural pesticide exposure in California in a year without a concurrent LUS. Our presented Landsat GIS method can be used in epidemiologic studies to reconstruct individual-level historical agricultural pesticide exposure using residential locations, especially as both PUR and Landsat data date back to the 1970s.

2.5.1 Accuracy Assessment in 1990

An accuracy assessment comparing classified 1990 Landsat images to the 1990 LUS gold standard demonstrated substantial agreement, with kappa statistics ranging from 0.700 to 0.789. Manifest in the varying producer's and user's accuracies across broad land use groups, some

agricultural classes were associated with greater misclassification compared to others (e.g., idle lands). Despite any land use misclassification, a large proportion of agricultural use segments (using Landsat images) truly belonged to an agricultural broad land use group (according to the LUS), which may be beneficial to pesticide exposure estimation. Specifically, if a PUR application was unable to achieve a tier 1 Landsat match, it would be likely matched at tier 2 to the other classified crops in the section.

The non-significant difference in pesticide application rates in 1990 estimated using the LUS and Landsat methods bolsters the potentially negligible impact of any land use misclassification on pesticide exposure estimates. This is further supported by a weighted kappa of 0.766 demonstrating substantial agreement in pesticide exposure classification. Given the importance of specificity in low prevalence exposures (<10%) regarding attenuating study results (144), such as an approximately 2.2% pesticide exposure prevalence in the California population (104), encouraging results were observed in the high probability of parcels being classified as not exposed given no exposure (90.9%, few false positives; Table 5).

2.5.2 Pesticide Exposure Estimation in 1985

Pesticide exposure was estimated in a year without a concurrent LUS - 1985. A lower number of pesticide applications in 1985 vs. 1990 is a result of requiring farmers to change from reporting restricted pesticide usage to reporting all pesticide usage, the documented increases in pesticide usage in California in the early to mid-1990s, pest outbreaks (e.g., cotton aphids), and inclusion of different PLSS sections associated with the 40% classification extent vs. the 1985 imagery extent (118, 147, 148). LUS and Landsat rates were significantly different, where the Landsat method achieved a significantly higher proportion of tier 1 matches compared to the LUS

method. Assuming Landsat rates are accurate, LUS rates typically overestimated pesticide exposure among urban residential parcels. Such differences may be a reflection of land use conversion occurring between 1985 and 1990. Ancillary data beyond Landsat imagery (e.g., aerial photographs) would help clarify observed differences in LUS and Landsat rates, for example, through showing the presence or absence of farmlands in 1985 vs. 1990. On the other hand, the observed difference between LUS and Landsat rates in 1985 may not be generalizable to comparisons between the LUS and Landsat methods during other years. Furthermore, the difference in median pesticide application rates was 0.04 lb/ac (LUS median 0.18 lb/ac vs. Landsat median 0.14 lb/ac), which may not be meaningful in terms of affecting pesticide exposure categorizations (e.g., change from low to high exposure). This is supported by a weighted kappa of 0.711 demonstrating good agreement in pesticide exposure classification between the LUS and Landsat methods.

Assuming segments in 1985 were accurately classified, the linkage between PUR data and Landsat-derived agricultural crop fields was most beneficial to PUR applications associated with several truck, field, and pasture crops. Agricultural growing practices are not limited to monocultures (i.e., repetitive growing of the same crop on the same land), but may include multiple cropping systems (149). Multiple cropping, also known as mixed cropping or polyculture, intensifies agricultural production through maximizing the efficiency of space and time, which serves to bio-diversify and stabilize the land, fertilize the crops in sequence, and promote pest control (149, 150). For example, crop rotation consists of the repetitive growing of different crops in a systematic and recurring sequence on the same field (151). Crop rotation is characterized by annual crop changes. For example, a legume crop (e.g., pea) may be planted to serve as a nitrogen source, followed by a sod crop to maintain organic matter and a cereal or root

crop to complete the rotation (151). Crop rotation is widely practiced across the U.S., evidenced by the infrequency of farmers continuously growing the same crop each year on the same field (152). Between 1996 and 2010, 84 to 94% of corn, soybean, and wheat acreage in the U.S. was associated with crop rotation.

The important theme underlying multiple cropping systems is the dynamic nature of agriculture, where any given year or growing season does not remain static. Promising results that serve to support the utility of a Landsat and PUR database linkage were demonstrated among the crops associated with Landsat tier 1 matches. A significantly higher proportion of 1985 organophosphate PUR applications associated with alfalfa, beans, cotton, and potatoes were able to be matched to classified segments using the Landsat method as opposed to the 1990-dated LUS method. These specific crops have been associated with documented crop rotation cycles in California. Alfalfa is often rotated with crops that are not hosts to pests that typically damage alfalfa populations (e.g., nematodes) such as cotton and beans (153). Potatoes can be rotated every three or more years with alfalfa to control wireworms (154). These results were bolstered by Landsat achieving a significantly higher proportion of tier 1 matches among temporary crops. Temporary crops are sown/seeded and harvested during the same crop growing season (e.g., cotton), as opposed to permanent crops (trees [e.g., apples] and vines [e.g., grapes]), which are sown or planted once and do not require replanting following harvests as they occupy the land for a long period of time (110). Results may differ when examining different pesticide chemical classes used more frequently on other crops, especially those that may or may not be adequately captured using the presented classification methodology.

The established three-tier method of linking PUR data with LUS's collapses several crops due to the uncertainty of their rotation such as cotton and tomatoes (24, 122). In other words,

PUR data are distributed to all crop fields belonging to this non-permanent class, which may include crops to which pesticides were not applied - potentially problematic in under- or overestimating pesticide exposure. Neither the Landsat nor the LUS method in this analysis collapsed crops to allow for a direct comparison of the crops benefiting from using current Landsat imagery vs. an outdated LUS to match PUR applications.

Conservative NDVI signature sample size constraints were implemented, requiring land use classes to have at least 100 pixels per land use class per month to ensure the inclusion of representative classes (134). A sensitivity analysis requiring at least 10 pixels revealed similar accuracy assessment results (data not shown). The classification of Landsat imagery can be implemented in epidemiologic studies estimating historical pesticide exposure from 1972 to the present (95). However, Landsat usage must be reconciled against the availability of ground truth data to derive NDVI land use signatures, as well as differences in the sensors (e.g., Multispectral Scanner [MSS]) that capture the images used to create signatures vs. the images used for classification - e.g., spatial resolution.

2.5.3 Combined LUS and Landsat Pesticide Three-Tier Matching Methodology

Harnessing the observed results regarding phenological misclassification in addition to differential capabilities of capturing temporary vs. permanent crops, a refined three-tier methodology could be developed that addresses the demonstrated strengths and limitations of both the LUS and Landsat methods (Appendix B). For example, a tier 1 match could be defined as when both LUS and Landsat crop fields match a PUR-reported crop. However, when LUS and Landsat data disagree in terms of crop types present in a section, among temporary crop PUR pesticide applications, Landsat data could be given more weight. A pesticide application would

be matched to a Landsat crop field if the Landsat and PUR crop match and the LUS crop field(s) present in that portion of the section belong to a different phenological group from that of the Landsat crop. Among permanent crop PUR pesticide applications, LUS data could be given more weight. A pesticide application would be matched to a LUS crop field if the LUS and PUR crop match, assuming that the PUR crop occupies the land for a long period of time. This alternative three-tier approach would have to reconcile differences between LUS vs. Landsat crop field boundaries and multiple crop fields belonging to the same crop type in a section.

2.5.4 Strengths

Strengths included the classification method, which was an improvement over previous approaches to classifying NDVI imagery into agricultural land use classes. Maxwell 2011 (97) implemented a minimum distance method, classifying each segment using the median NDVI value of each CDWR LUS land use class via a sum of squared differences measure across multiple time points in a year. Each LUS polygon in the training data was represented by a single pixel value at the label point or that was seemingly representative of the polygon, which does not take into account the variability within a LUS polygon and across polygons within the same class. The supervised classification method in this analysis was comprised of per-pixel MLC followed by per-field classification using segments - offering several advantages compared to minimum distance methods. MLC is a parametric method based on Bayes' Theorem, classifying a pixel through maximizing the probability of correct classification (98). Training data from the 1990 LUS and 1990 NDVI images were used to estimate a mean vector and covariance matrix for each land use class. The sample a priori weighting logic was selected, which assigned greater weight to the more frequently occurring pixels (i.e., land use classes) (135). MLC is useful when

land use classes overlap in spectral space, which is evident in the phenological groups comprised of crops with similar temporal NDVI patterns (98). For example, this Bayesian methodology was able to incorporate information regarding particular land use classes (e.g., cotton) having consistently dominated the agricultural landscape in Kern County throughout the study time period (108).

Per-field classification integrates raster and vector data in order to account for within-field spectral variability (102). Pixels are impacted by variability due to soil moisture conditions, pests, and disease, ultimately altering the captured spectral signature (99, 100). Through implementing a majority rule, the spatial autocorrelation of agricultural crop fields is addressed, as pixels close in proximity are likely of the same agricultural crop field, essentially averaging out the noise caused by the typical salt-and-pepper effects of per-pixel classifiers (101). Furthermore, fields were created using a local behavior-based image segmentation procedure in IDRISI, which implemented a watershed delineation, or region extraction, method of merging/growing pixels across input bands (monthly NDVI images) exhibiting minimal variance, or spectrally similar pixels that are likely of the same land use (155). Although the vector fields used in per-field classification are typically parcels dividing the landscape (i.e., segments used may not parallel crop field boundaries) (101, 156), classified segments were dissolved according to crop type and PLSS section, which was meaningful in terms of linking the PUR database to agricultural crop fields.

2.5.5 Limitations

Limitations of the classification approach included MLC per-pixel classifiers ignoring the mixed pixel problem, which is present with moderate spatial resolution imagery such as Landsat (101).

Mixed pixels occur when a pixel is not completely occupied by a single homogeneous category such as at the edge of large discrete objects (e.g., agricultural fields) and due to linear features (e.g., roads) (98). Sub-pixel classifiers can be used to address the mixed pixel problem such as the fuzzy set technique, where a soft classification method assigns a pixel to multiple memberships. Mixed pixels could have translated to issues regarding classification if features (e.g., streams) were within a segment's boundaries and were not otherwise made distinct from the other spectral signatures used to classify the segment (100). De Wit and Clevers (100) recommended segmenting within the parcel boundaries of a large-scale topographic map to account for non-agricultural features such as roads. Although per-field classification is affected by field size and shape, sensitivity analyses of the accuracy assessment using different segment sizes demonstrated similar results (phenological group kappas ranging from 0.70 to 0.75).

Misclassification of crops exhibiting similar phenological patterns was evident in some high omission and commission errors, and potentially contributed to a higher proportion of LUS tier 1 matches compared to Landsat among some crops. Almonds, oranges, and peaches/nectarines are permanent crops that were better captured using an outdated LUS compared to the classified Landsat images. Almond trees require four years after planting to mature before the nuts can be harvested (157). Given a five-year difference between the LUS and Landsat imagery, such longstanding crops could be expected to be present in both datasets. Their absence may be a result of phenology-related misclassification. A soft classification approach could be adopted to incorporate the uncertainty posed by phenological similarities. Another alternative includes combining phenologically similar crops into a single category (e.g., tomatoes and peppers) (99, 102), although such combinations may result in misclassification of pesticide exposure. Another limitation is associated with the hard classification used to assign each

segment to one land use class. Although precautions were taken at the land use NDVI signature creation stage (single-use LUS, negative buffers, ≥100 pixels per month required), the classification method could not account for intra-annual crop rotation, for example, when two or more crops are successively grown on the same field each year (149). Focusing the analysis on acquiring imagery to classify specific crop types with known planting and harvesting dates may minimize this issue.

Implementing sophisticated classification techniques that incorporate ancillary information to further enhance classification accuracy, and thus pesticide exposure estimation, should be explored. For example, Yu et al. (155) classified Digital Airborne Imagery System (DAIS) images of Point Reyes National Seashore, CA into land use classes in conjunction with ancillary data to enhance class discrimination - e.g., elevation and slope (using a USGS digital elevation model [DEM]), distance to watercourses, intensity, hue, and saturation. A Classification and Regression Tree (CART) algorithm selected features most important for classification and a k-nearest neighbor classifier assigned segments to vegetative classes. Conrad et al. (158) segmented SPOT 5 imagery of Uzbekistan in 2006 for use in per-field classification. Using Tasseled Cap vegetative indices derived from bi-temporal Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images from 2007 and expert knowledge regarding vegetative density and soil wetness, a rule-based algorithm determined final land use classifications.

Other limitations include the absence of post-stratification weighting to derive representative residential parcel pesticide application rates and a Kern County population prevalence of organophosphate exposure (159). Although weighting would have been useful in an epidemiologic context to provide estimates of residential pesticide exposure experienced by

this population and to take into account unequal probabilities of selecting residential parcels into the sample, the purpose of this study was to compare and contrast two methods of estimating agricultural pesticide exposure. The temporal resolution of the NDVI signatures used to classify imagery, acquired for 10 months in 1990, may have limited land use classification. Landsat sensors capture images every 16 to 18 days (95). Using additional images may highlight intra-month NDVI variability to further enhance land use discrimination. The utility of the land use signatures is also affected by the extent to which NDVI values may change from year-to-year. NDVI values are affected by remote sensing system characteristics (e.g., sensor resolution), meteorological conditions (e.g., clouds and snow), ecosystem disturbances (e.g., fires), and seasonality (e.g., prolonged warm temperatures extending the growing season; prolonged droughts) (160). Maxwell and Sylvester (161) showed high between-year variability in maximum annual NDVI values for cropped lands between 1984 and 2010 derived from Landsat images of southwestern Kansas. Usage of signatures closer in time to the imagery to be classified and under similar physical conditions may provide more accurate agricultural land use classifications - although this is limited by ground truth data availability. The external validity of using NDVI signatures to classify images from other geographic areas around California and at different points in time should be explored. More sophisticated approaches to identify clouds could be implemented to minimize the inclusion of affected pixels such as harnessing spectral reflectance information (162).

2.6 CONCLUSIONS

GIS-based methods provide a powerful way to quantify agricultural pesticide exposure, addressing important facets of exposure estimation in the context of epidemiologic studies (e.g., minimizing recall bias, identifying specific pesticides, and reconstructing historical exposure). A methodology to classify Landsat imagery using NDVI signatures via a maximum likelihood and per-field classification approach was presented. An accuracy assessment demonstrated substantial agreement with the LUS gold standard and minimal pesticide exposure misclassification. Use of classified Landsat imagery to estimate agricultural pesticide exposure in 1985, a year lacking a current LUS, was demonstrated through a linkage with the California PUR database using crop type and PLSS section. The Landsat method achieved significantly more tier 1 PUR matches compared to the LUS method, particularly among temporary crops associated with annual crop rotation. Estimated pesticide application rates were significantly different. Future research should explore the combined use of the LUS and Landsat methods to estimate pesticide exposure, in addition to image classification methods using ancillary data to improve the accuracy of both land use classification and pesticide exposure estimation.

2.7 TABLES

Table 3. Accuracy Assessment of Classified Landsat Imagery vs. LUS in 1990

	CDWR land use[a]		CDWR broad land use group[a]		Phenological group[a]	
N	**Agreement**	**Kappa (95% CI)**	**Agreement**	**Kappa (95% CI)**	**Agreement**	**Kappa (95% CI)**
87,197 segments[b]	75.1	0.700 (0.696, 0.703)	78.5	0.732 (0.728, 0.735)	82.7	0.779 (0.776, 0.782)
559,908 ac[c]	76.4	0.701 (0.699, 0.702)	79.5	0.731 (0.730, 0.733)	84.4	0.789 (0.788, 0.790)

[a] The accuracy assessment was performed on classified imagery from the 40% classification extent that was derived from the 1990 NDVI signatures extent and not used for training data. Results are presented for the intersections between segments and single-use LUS polygons.
[b] N=86,060 segments for phenological groups.
[c] N=554,312 ac for phenological groups.

Table 4. Accuracy Assessment of Landsat Imagery vs. LUS in 1990 Using Number of Segments: CDWR Broad Land Use Groups

		Land use survey (gold standard)															
		C	**D**	**F**	**G**	**I**	**NV**	**NW**	**P**	**R**	**S**	**T**	**U**	**V**	**Total**	**User's**	**Agr[a]**
Landsat	**C**	2,058	44	17	9	5	76	1	8	0	19	23	76	42	2,378	86.5	92.8
	D	27	4,977	42	7	33	291	22	190	0	30	23	203	456	6,301	79.0	91.3
	F	68	201	20,884	312	111	517	24	499	0	35	628	50	242	23,571	88.6	97.3
	G	10	17	466	2,292	65	364	3	32	0	2	228	4	0	3,483	65.8	89.3
	I	13	63	406	91	92	226	6	58	0	33	97	33	24	1,142	8.1	73.9
	NV	242	418	1,887	148	1,774	22,770	20	93	0	121	78	1,262	43	28,856	78.9	16.2
	NW	0	2	16	1	2	53	38	0	0	0	2	10	0	124	30.6	18.5
	P	48	136	775	47	89	29	0	6,133	0	10	81	45	176	7,569	81.0	98.9
	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	--	--
	S	1	2	1	0	1	9	0	3	0	2	0	1	3	23	8.7	47.8
	T	6	20	985	139	52	133	0	195	0	1	2,130	7	31	3,699	57.6	96.2
	U	95	203	262	28	63	571	10	56	0	174	30	3,208	94	4,794	66.9	17.3
	V	33	583	164	7	27	230	6	161	0	25	91	40	3,890	5,257	74.0	94.3
	Total	2,601	6,666	25,905	3,081	2,314	25,269	130	7,428	0	452	3,411	4,939	5,001	87,197		
	Producer's	79.1	74.7	80.6	74.4	4.0	90.1	29.2	82.6	--	0.4	62.4	65.0	77.8			
	Agr[a]	87.0	90.6	91.6	94.3	20.5	7.4	47.7	98.0	--	34.3	96.8	9.3	97.2			

Abbreviations: C = citrus and subtropical; D = deciduous fruits and nuts; F = field; G = grain and hay; I = idle; NV = native vegetation; NW = water surface; P = pasture; R = rice; S = semi-agricultural; T = truck, nursery, and berry; U = urban; V = vineyard.

[b] Agr = agricultural; proportion of land use class classified as agricultural use (all classes except NV, NW, S, and U).

Table 5. Classification of Residential Parcels According to 1990 Pesticide Exposure Quartiles: Landsat vs. LUS Methods

		LUS (gold standard)[a]					
		None	Low exposure	Moderate exposure	High exposure	Total	Weighted kappa
Landsat[a]	None	437	30	18	14	499	0.766 (0.739, 0.792)
	Low exposure	14	91	19	8	132	
	Moderate exposure	16	30	222	57	325	
	High exposure	14	14	63	244	335	
	Total	481	165	322	323	1,291	

[a] None: 0 lb/ac; low: >0-0.40 lb/ac; moderate: 0.40-1.74 lb/ac; high: >1.74 lb/ac.

Table 6. Pesticide Application Tier Matching In 1985: LUS vs. Landsat Methods

		LUS[a,b]				
		Tier 1	Tier 2	Tier 3	Total	p[c]
Landsat[a,b]	Tier 1	1,864	452	40	2,356	<0.0001
	Tier 2	381	1,081	58	1,520	
	Tier 3	0	0	33	33	
	Total	2,245	1,533	131	3,909	

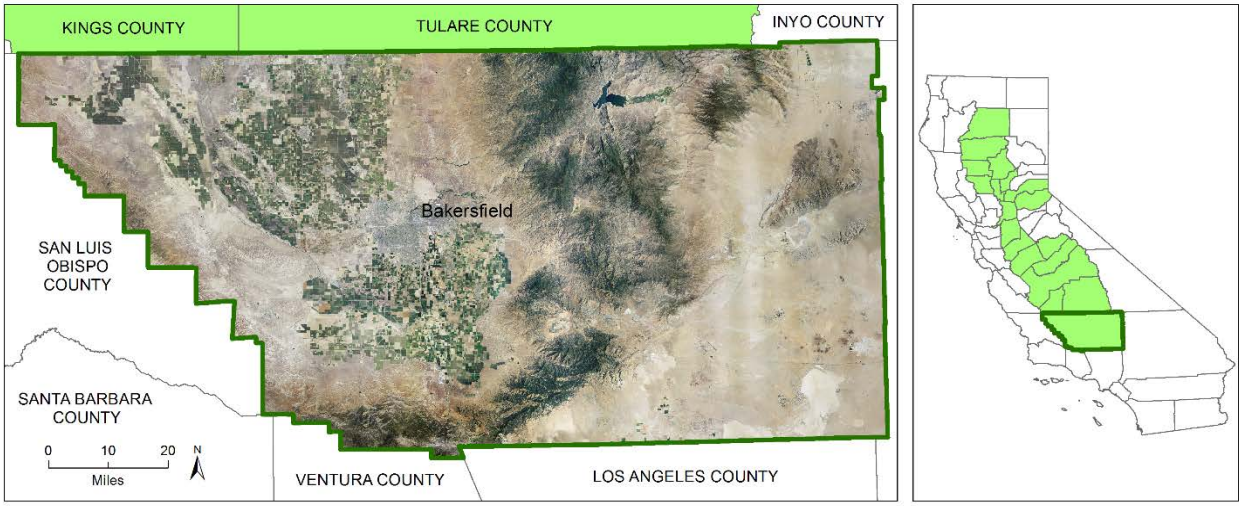
[a] 3,909 organophosphate applications (549,158.76 lb) in 1985 imagery extent.
[b] Landsat achieved more tier 1 vs. tier 2 and 3 matches (McNemar's p=0.0002) and more tier 1 and 2 vs. tier 3 matches compared to LUS (p<0.0001).
[c] Bowker's test of symmetry.

Table 7. Pesticide Application Rates in 1985: LUS vs. Landsat Methods

	Pesticide application rates (lb/ac)[a]			
Method	Min	Median; Mean ± SD	Max	p[b]
LUS	0	0.18; 0.52 ± 1.13	21.03	0.0448
Landsat	0	0.14; 0.56 ± 2.14	50.73	

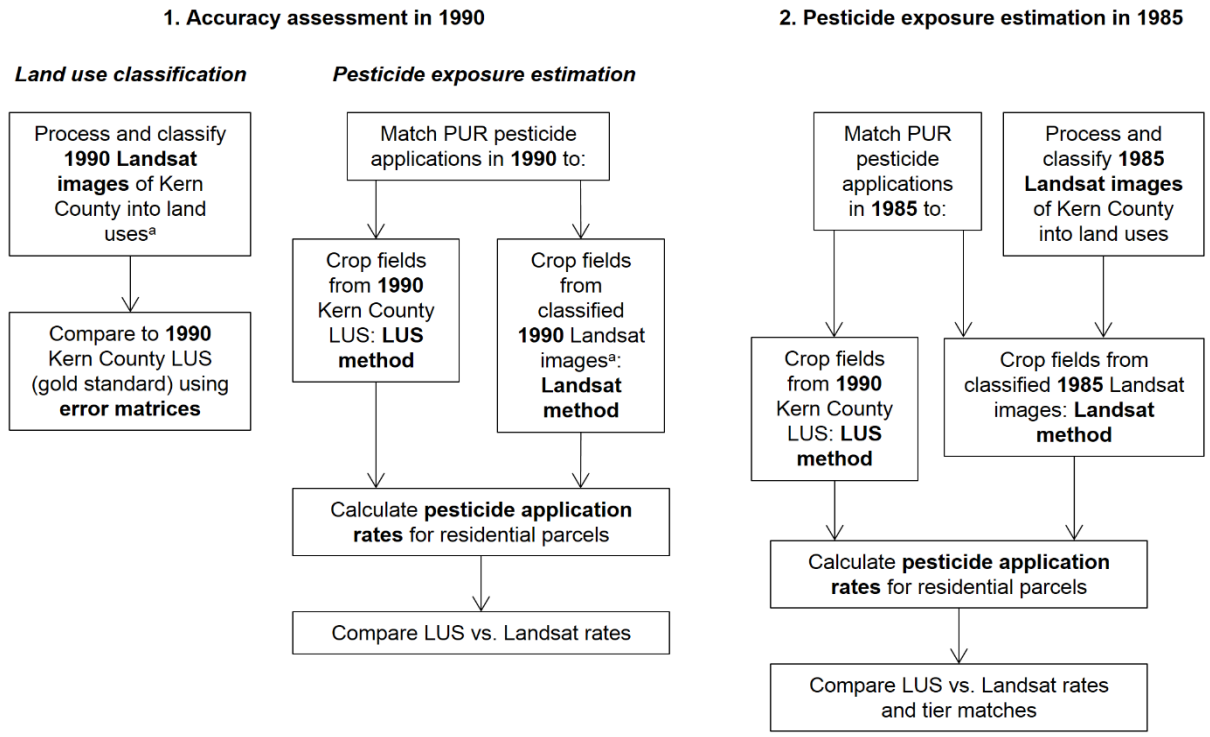
[a] 1,293 sampled residential parcels in 1985 imagery extent.
[b] Wilcoxon signed-rank test.

2.8 FIGURES



(Data Source: USDA FSA Aerial Photography, 2012)

Figure 2. Kern County, CA Study Area: National Agriculture Imagery Program (NAIP) Compressed County Mosaic (CCM) of Kern County From August 2012 (Left); Kern County Within California's Central Valley Agricultural Region (Right)



[a] Classified Landsat images from 1990 used to create error matrices were also used to estimate pesticide exposure.

Figure 3. GIS Workflow for Accuracy Assessment and 1985 Pesticide Exposure Estimation

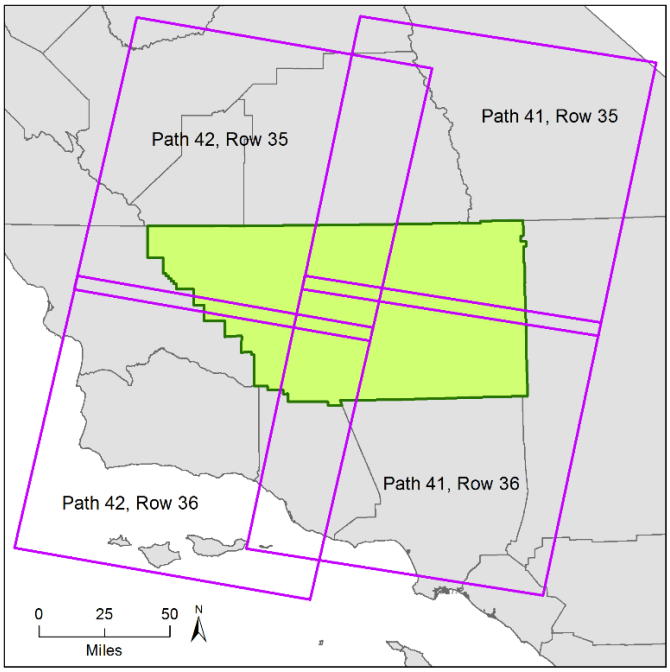


Figure 4. Landsat Paths 41 and 42 and Rows 35 And 36 Intersecting Kern County, CA

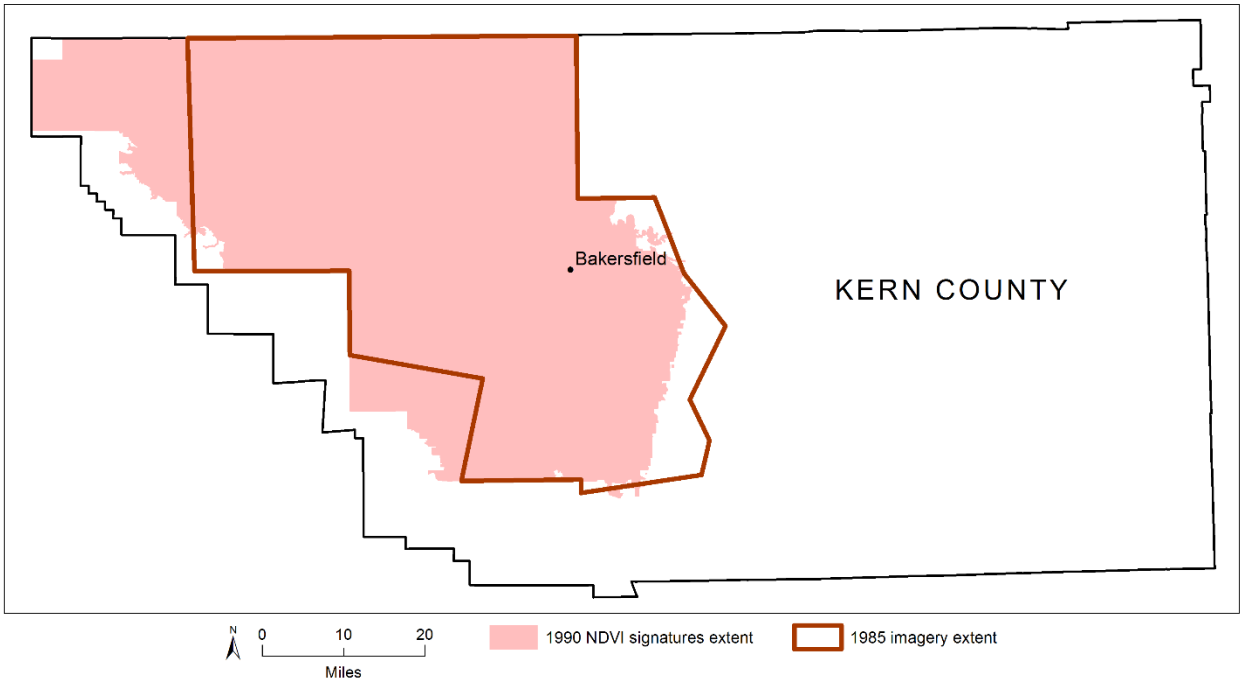


Figure 5. Geographic Extent of NDVI Training Data and Classification Data: NDVI Signatures for Land Use Classes Derived From Images Within the 1990 NDVI Signatures Extent (Pink Region) Used for Accuracy Assessment in 1990 and as Training Data for Maximum Likelihood Classification of 1985 Images Within the 1985 Imagery Extent (Red Region)

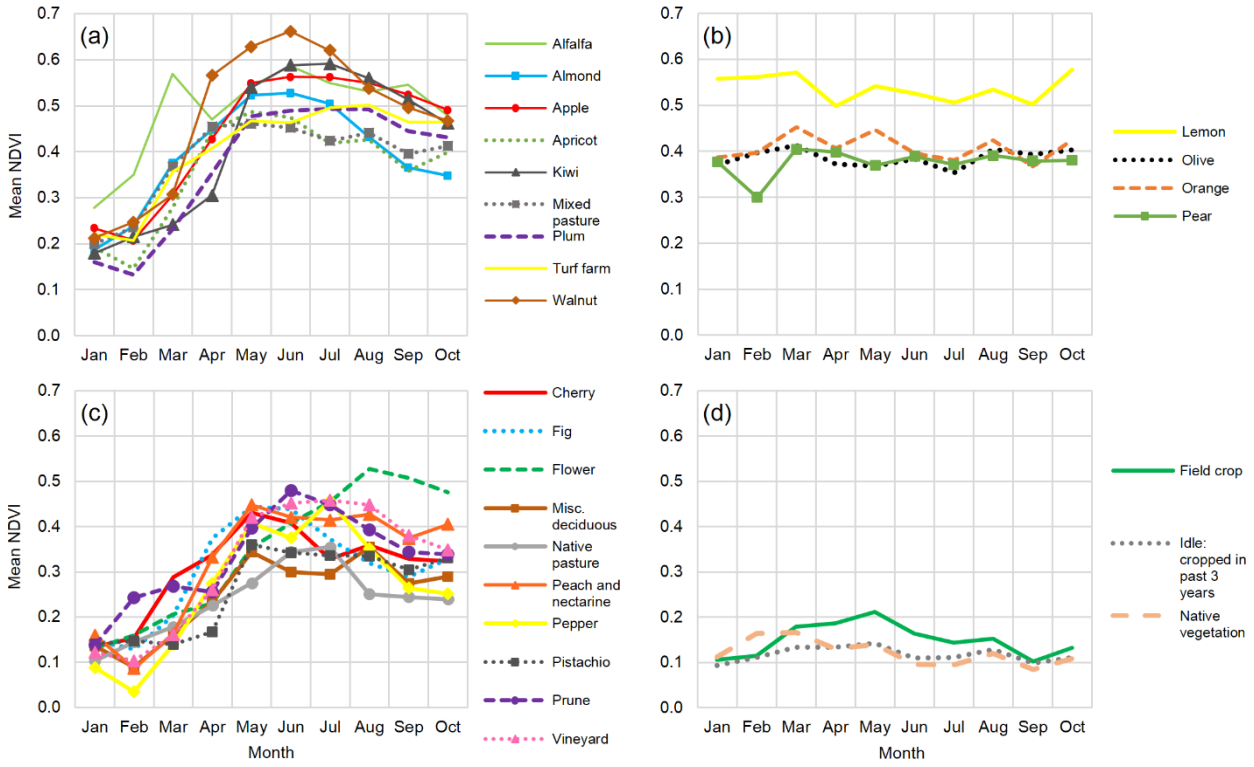


Figure 6. Phenological Groups Comprised of Land Uses Sharing Similar Annual NDVI Patterns Derived From Cluster Analysis: Land Uses Exhibiting (a) Gradual Summer NDVI Peak, (b) Stable NDVI Pattern, (c) Moderate Vegetative Density Peak, and (d) Low NDVI Pattern

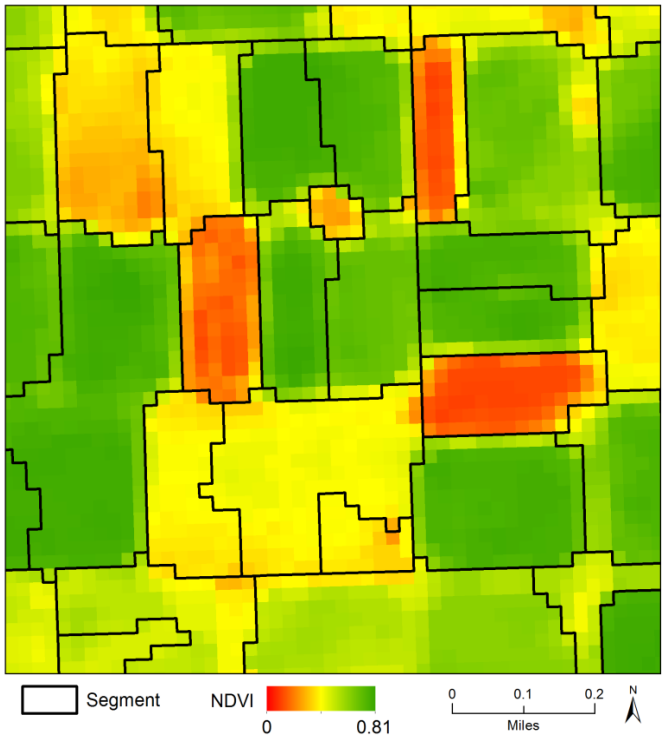


Figure 7. Segments Overlaying an August 1985 NDVI Image

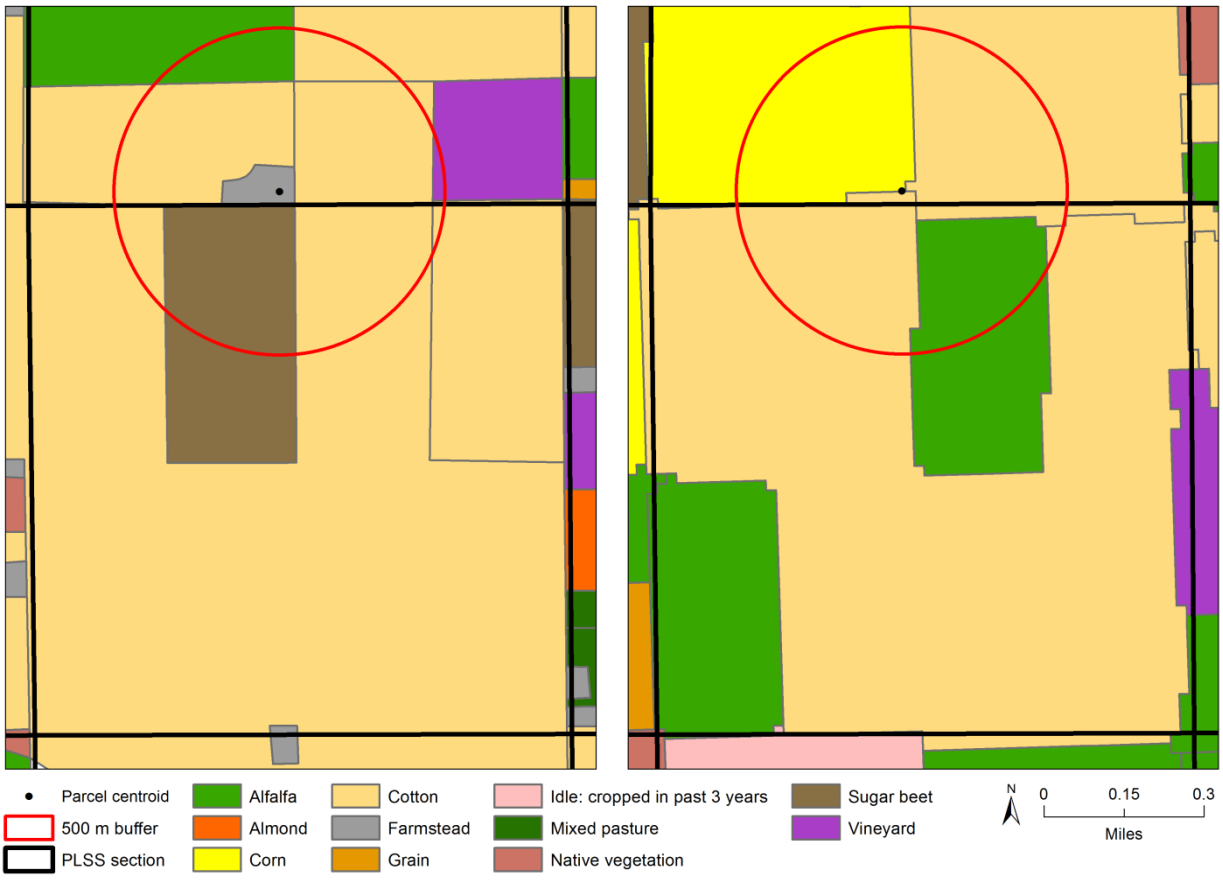


Figure 8. LUS vs. Landsat Pesticide Exposure Estimation in 1985: One Organophosphate Alfalfa Application (27.96 lb) in Section 15M29S25E15 With Tier 2 LUS Match (LUS Pesticide Application Rate 0.82 lb/ac; Left) and Tier 1 Landsat Match (Rate 1.15 lb/ac; Right)

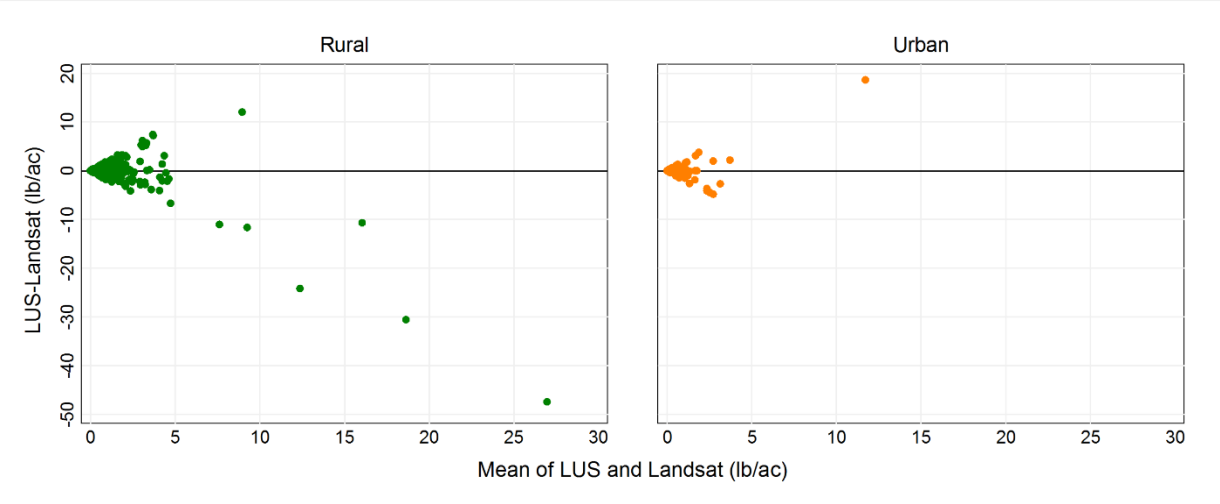


Figure 9. Bland-Altman Plots of LUS vs. Landsat Rates in 1985: Mean vs. Difference Between LUS and Landsat Pesticide Application Rates for Sampled Residential Parcels Stratified by Rural (Left) and Urban (Right) Location

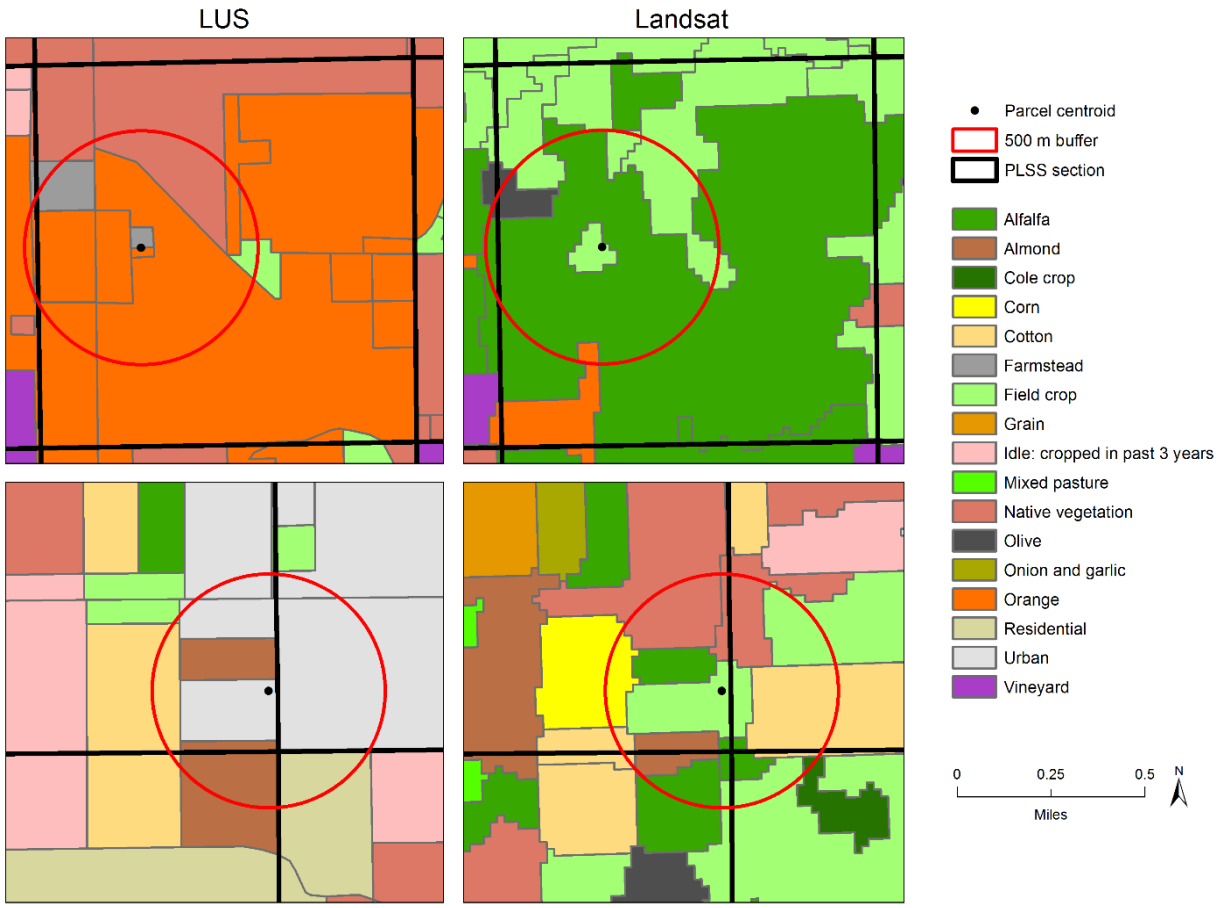


Figure 10. Discrepant Pesticide Application Rates According to LUS and Landsat Methods and Rural/Urban Status: Rural Parcel With 3.24 lb/ac LUS Rate and 50.73 lb/ac Landsat Rate due to Fewer Landsat Orange Fields in Parcel's Section (Top Row); Urban Parcel With 21.03 lb/ac LUS Rate and 2.38 lb/ac Landsat Rate Due To More Landsat Almond Fields in Parcel's Section (Bottom Row)

3.0 ZIP CODE-LEVEL GIS PESTICIDE EXPOSURE METRICS: ACCURACY OF AREA, POPULATION, AND NETWORK-ATTRIBUTED WEIGHTING METHODS

3.1 ABSTRACT

Spatial data are often available at an aggregated, ecologic scale (analysis scale) that may not reflect the finer scale at which the feature of interest operates (operational scale). When the source unit (unit of data in its current form) is smaller in size relative to the target unit (unit at which data are available), spatial aggregation can be applied to scale up the data from the source unit to the target unit. Usage of relevant ancillary data in a weighting method, as part of spatial aggregation, can address the discrepancy between the analysis and operational scales. We explored three different weighting methods while scaling up California agricultural pesticide exposure data from the Public Land Survey System (PLSS) 1 mi^2 section level to the ZIP Code level: area, population, and network-attributed/road. The research goal was to develop a GIS-based ZIP Code pesticide exposure metric that most accurately estimates exposure occurring among the population residing in California. California Pesticide Use Report pounds of applied pesticides were matched to PLSS sections and pesticide application rates calculated using the three ZIP Code weighting methods in a GIS were compared to rates calculated using the census block-level gold standard method. Compared to population and road weighting, area weighting achieved the most accurate pesticide application rates when compared to the gold standard. Area-

weighted rates were in moderate agreement with the gold standard (weighted kappa 0.55; 95% confidence interval [CI] 0.52, 0.58) and in substantial agreement in rural locations (0.63; 95% CI 0.57, 0.69). Area weighting is the most accurate approach to estimate agricultural pesticide exposure if scaling up California pesticide data from PLSS sections to ZIP Codes. The proposed weighting and aggregation approach can be applied in epidemiologic studies to study environmental exposures that affect human populations in the absence of large scale spatial data.

3.2 INTRODUCTION

The use of spatial data, or data with locational information, and geographic information systems (GIS) provide a powerful approach to studying the effects of environmental exposures on human health outcomes. In the context of epidemiologic studies, environmental exposure data available in a spatial format can be linked with georeferenced health outcome data, ranging from small-scale ecologic units (e.g., ZIP Codes) to large-scale individual-level units (e.g., geocoded residential locations) (164). However, a fundamental issue underlying the use of spatial data in epidemiology involves reconciling the analysis scale vs. the operational/phenomenon scale. The scale at which a feature of interest is measured or aggregated (i.e., analysis scale) may not correspond to the scale at which that feature of interest operates (i.e., operational scale) (165, 166).

One example surrounds pesticide exposure, which has been associated with the risk of developing many adverse human health outcomes such as cancers (23). Pesticides are chemicals designed to treat pests such as insects, potentially affecting humans through a variety of different routes (e.g., drinking water, food, air, and dust) (167). One important route of exposure affecting

rural populations occurs via residential proximity to agricultural pesticide applications (86). Previous epidemiologic studies recognized that meaningfully quantifying pesticide exposure occurring through this route requires high resolution data sources at the level of an individual's place of residence (i.e., operational scale). For example, the standard GIS method to estimate agricultural pesticide exposure in California links California Department of Pesticide Regulation Pesticide Use Report (PUR) data with California Department of Water Resources land use surveys (LUS's) (24). Specifically, PUR pounds of applied pesticides, reported according to crop type and 1 mi^2 Public Land Survey System (PLSS) section, are matched to LUS vector crop fields. Pesticide exposure is estimated using the pesticide-treated crops within a 500 m (radius) buffer around an individual's geocoded residence.

However, geocoded residential locations may not be available due to aggregation to protect patient confidentiality and lack of data availability (164, 168). As a consequence, epidemiologic studies can be constrained to smaller scale spatial units such as ZIP Codes (164). When the source units, or the unit of the data in its current form (e.g., PUR data reported by PLSS sections) are nested within or smaller in size than the target units, or the unit at which the data are available (e.g., ZIP Codes), scaling-up or spatial aggregation methods must be employed to aggregate the source units to the target units (169, 170). For example, Clary and Ritz (171) implemented a summation approach when examining the association between ZIP Code-level organochlorine pesticide exposure and pancreatic cancer mortality in California. As only ZIP Code data was available on death certificates, PUR pounds of pesticides were matched to PLSS sections and aggregated to the ZIP Code level.

Given a particular spatial aggregation (e.g., PLSS section level to ZIP Code level) and a particular phenomenon of interest (e.g., human agricultural pesticide exposure), a specific

weighting method that uses ancillary data to convert information from the source unit to target unit level may be more appropriate. For example, a simple summation of PLSS section-level applied pounds of pesticides to an aggregated areal unit does not take into account individuals who may or may not reside near the PLSS sections in which pesticides were applied. The weighting component of a spatial aggregation method represents one way to address the discrepancy between the analysis and operational scales by incorporating ancillary data to craft an aggregated metric that best reflects the finer scale at which pesticide exposure impacts human health.

We conducted a validity study to explore three different methods of weighting when spatially aggregating PLSS sections to the ZIP Code level to determine which method best reflects, or is the most accurate indicator of, agricultural pesticide exposure at the higher resolution census block level. Although this study specifically addresses spatial scale in the context of agricultural pesticide exposure at the ZIP Code level, the proposed weighting methods can be adapted to study other environmental exposures and other geographic units of analysis.

3.3 METHODS

3.3.1 Study Area and Data Sources

California is the most agriculturally productive state in the U.S. (172). It is comprised of 58 counties and is 158,706 mi^2 in area (173). The state population was 29,760,021 in 1990; 33,871,648 in 2000; and 37,253,959 in 2010 (174). U.S. Census Bureau TIGER/Line® files provided administrative boundaries used in creating the figures (115).

3.3.1.1 Pesticide Data

The California Department of Pesticide Regulation (CDPR) has collected Pesticide Use Report (PUR) data since 1974 (91). PURs contain information regarding agricultural use pesticide applications reported by farmers and commercial pest control operators, including the chemical name, pounds, crop type, date, and Public Land Survey System (PLSS) section of application. PLSS sections are vector data used for surveying purposes that are 1 mi^2 in area and span parts of the U.S. (112).

3.3.1.2 U.S. Census Bureau Data

The U.S. Census Bureau conducted the 2000 Census of Population and Housing starting on April 1, 2000 (175). As part of the 2000 Census, the Summary File 1 (SF1) contains information (e.g., age) asked of all persons and housing units in the U.S. (176). Census blocks are the smallest statistical units for which data are collected (177). Census blocks in urban areas are bound by streets and typically smaller than census blocks in rural areas, which are bound by roads and streams. ZIP Code Tabulation Areas (ZCTAs) were used to approximate U.S. Postal Service ZIP Code boundaries in California, which are used as the basis of mail delivery in the U.S. (178). ZCTAs are created by determining the most frequently occurring ZIP Code within each census block (179). Census blocks are hierarchically nested within ZCTAs. In this paper, the term ZIP Code will be used in place of ZCTA. U.S. Census Bureau roads include Federal Interstate highways, state highways, and local, neighborhood, and rural roads (180). TIGER/Line® vector files for census blocks, ZIP Codes, and roads and Census population data from 2000 were used as this year coincided with the time period during which PUR data was analyzed (1995 to 2005).

3.3.1.3 Dasymetric Data

The National Land Cover Dataset 1992 (NLCD1992) is a raster file of 21 land cover classes (e.g., mixed forest) spanning the 48 coterminous U.S. states (181, 182). The NLCD1992 is associated with 30 m spatial resolution and created via unsupervised classification of 1992 Landsat 5 Thematic Mapper images incorporating ancillary data such as elevation and population density. The California Protected Areas Database (CPAD) is a vector file of California lands owned in fee and protected for open space purposes (e.g., national and state forests) updated in March 2014 (183). The California Conservation Easement Database (CCED) is a vector file of California lands protected under conservation easements (e.g., legal agreement between land owner and government to conserve land) (183).

3.3.2 Pesticide Data Processing and Application Matching

Agricultural use PUR data from 1995 to 2005 were processed using CDPR-provided error files (errors with corrected values were retained) or logic checks (e.g., duplicate removal) if error files were unavailable (116). Outlier application rates were defined using three CDPR-created flags (e.g., >200 lb/ac if non-fumigant pesticide) (118) and replaced with the statewide median rate for the pesticide active ingredient in that year (24). Pounds of active ingredient were recalculated using the number of treated acres. Pesticide applications associated with the organochlorine, organophosphate, and carbamate chemical classes were extracted (23, 27, 89, 119-123, 184). Pounds of pesticides were matched and summed according to PLSS section, the geographic level of reporting of the PUR database (185).

3.3.3 Gold Standard Pesticide Exposure Metric: Dasymetric Mapping

Three ZIP Code metrics were compared to a gold standard pesticide exposure metric, which was calculated by first implementing the U.S. Geological Survey (USGS) dasymetric mapping method to identify potential locations where individuals in California likely reside (USGS, 186). California NLCD1992 raster files were reprojected to the California Teale Albers (NAD83 datum) coordinate system and mosaicked. A majority filter (3x3 kernel) was implemented and land cover classes were reclassified into four mutually exclusive classes: (1) high intensity residential (NLCD code 22), (2) low intensity residential (NLCD code 21), (3) non-urban (remaining 18 NLCD codes excluding water), and (4) water/excluded (NLCD code 11). High intensity residential, low intensity residential, and non-urban will be referred to as the three habitable classes. The water/excluded class includes geographic areas where humans are not expected to reside; thus, pesticide exposure was not estimated within these areas. CPAD and CCED data were combined via union into an open space layer, dissolved, and rasterized to 30 m pixels (spatial resolution of NLCD1992 layer). All open space pixels were recoded to a value of 10. Using map algebra, the NLCD and open space layers were summed and any pixels with values of 10 or greater were reclassified into the water/excluded class (i.e., all open space lands were excluded), creating the NLCD-open space layer.

The NLCD-open space layer was converted to a polygon vector layer. Census block-level total population estimates from the 2000 U.S. Census Bureau SF1 were joined with the 2000 California census blocks layer. Census blocks with >0 population were intersected with the NLCD-open space layer (habitable classes only) and dissolved according to census block and habitable class. Centroids for each unique census block-habitable class polygon were generated and 500 m (radius) buffers (meaningful distance regarding pesticide drift used in previous

studies) were created around each centroid (24, 86). Buffers were intersected with PLSS sections. For each centroid, annual pesticide application rates (lb/ac) were calculated by dividing the summed pounds of applied pesticides by the PLSS section acreage, weighting the rate by the proportion of the sections intersecting the buffer, and dividing by 11 years.

3.3.4 ZIP Code Pesticide Exposure Metrics: Area, Population, and Network-Attributed/Road Weighting

After intersecting PLSS sections with ZIP Codes, PLSS section pesticide application rates were spatially aggregated to the ZIP Code level according to three different ZIP Code weighting methods. *Area weighting*: PLSS section pesticide application rates were weighted by the proportion of the ZIP Code area comprised by that particular section. *Population weighting*: After intersecting PLSS sections with census blocks, areal interpolation was implemented to disaggregate the population of each census block according to the proportion of the census block area intersecting PLSS sections (187). For PLSS sections intersecting multiple census blocks, interpolated population totals were summed. PLSS section pesticide application rates were weighted by the proportion of the ZIP Code population assigned to that particular PLSS section via areal interpolation. *Network-attributed/road weighting*: TIGER/Line® roads with Census Feature Class Codes (CFCCs) A21 to A48 were selected, which include local roads and state highways (180). Selected roads were intersected with PLSS sections and ZIP Codes. PLSS section pesticide application rates were weighted by the proportion of the ZIP Code roads assigned to that particular PLSS section. For all three ZIP Code metrics, recalculated pounds and intersecting PLSS section acreage were summed by ZIP Code and divided to calculate ZIP Code pesticide application rates. Annual rates were calculated by dividing by 11 years.

3.3.5 Statistical Analysis

Wilcoxon signed-rank tests compared ZIP Code pesticide application rates estimated using each weighting method. For each centroid, gold standard and ZIP Code rates were categorized into five mutually exclusive groups defined by each ZIP Code metric: values below the 75th percentile and quartiles of the values above the 75th percentile. Using the ZIP Code in which each centroid is located, weighted kappas were calculated for each ZIP Code metric and 95% confidence intervals (CIs) were estimated using a bootstrap method taking into account clustering of centroids within ZIP Codes (188). After dichotomizing gold standard rates using the 75th percentile (pesticide-exposed if $\geq$75th percentile), likelihood ratios (LRs) quantified how likely a pesticide-exposed gold standard rate was classified into each ZIP Code-derived pesticide exposure category compared to unexposed rates. Results were stratified by rural/urban location using Version 2.0 Rural-Urban Commuting Area (RUCA) codes (189). Using Categorization C, all ZIP Codes with RUCA codes of 1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1, or 10.1 were categorized as urban; all other RUCA codes were categorized as rural.

3.4 RESULTS

From 1995 to 2005, approximately 117,380,168 lb (53,243 metric tons) of pesticides belonging to the organochlorine (3,720 metric tons), organophosphate (28,872 metric tons), and carbamate (20,651 metric tons) chemical classes were applied in California. Among the 163,812 PLSS sections in California, 22,871 sections (13.96%) were treated with pesticides between 1995 and 2005 with annual rates ranging between 0 and 100.27 lb/ac (Figure 11). The highest rates were

concentrated in the Central Valley agricultural region spanning north-central to south-central California (190). Seventy-eight percent (N=127,983) of PLSS sections intersect ZIP Codes and 98% (N=22,435) of treated sections intersect ZIP Codes, which represent the sections used to calculate gold standard pesticide application rates and that were scaled up or spatially aggregated to calculate ZIP Code rates.

3.4.1 Gold Standard Pesticide Application Rates

There was a total of 694,087 unique polygons included in the analysis representing census blocks with a population total >0 that intersected a habitable class (i.e., high intensity residential, low intensity residential, or non-urban) and intersected a ZIP Code. These polygons correspond to 343,449 census blocks.

Figure 12(a) depicts pesticide application rates for each unique aforementioned polygon (N=694,087) - each rate calculated using 500 m buffers created around that particular polygon's centroid (mean 0.08 lb/ac ± 0.41, median 0 lb/ac; Table 8). The majority of rates were 0 lb/ac (57.31%). Higher rates were observed along the Central Valley region. Sixty-one percent (N=422,688) of the buffers intersected one or two PLSS sections. Among the 296,276 buffers intersecting at least one pesticide-treated PLSS section, 82.7% (N=245,020) intersected one or two pesticide-treated PLSS sections. A large proportion of California was associated with geographic areas corresponding to an absence of census blocks with >0 population intersecting habitable areas, as depicted by the white areas in Figure 12(a). These areas accounted for 70.89% (112,363.61 mi^2) of California's total area. These geographic areas were not considered for gold standard pesticide application rate calculations as it was assumed that no individuals resided in these areas.

3.4.2 ZIP Code Pesticide Application Rates

Average ZIP Code pesticide application rates ranged between 0.13 to 0.16 lb/ac (Table 8). Between 27.18 and 29.03% of ZIP Codes were 0 lb/ac. The geographic distribution of area-weighted rates is similar to the distribution of gold standard rates - higher rates observed along the Central Valley region >1.01 lb/ac (Figure 12). Among the 1,678 ZIP Codes in California, four were uninhabited in 2000 (90261, 91608, 94128, and 96095) and one did not contain any roads (96095).

3.4.3 Accuracy Assessment of ZIP Code Rates vs. Gold Standard Rates

ZIP Code rates (N=1,674) were significantly different from each other (area vs. population signed-rank $p<0.0001$; area vs. road $p<0.0001$; population vs. road $p<0.0001$). There were between 1 and 3,007 census block-habitable class polygon centroids (median 317; 25^{th} pctl 81; 75^{th} pctl 650) located within each ZIP Code. Gold standard pesticide application rates were categorized into one of five mutually exclusive pesticide exposure classes defined by values of each ZIP Code metric: none, low, low-moderate, moderate, and high (Table 9). Area weighting achieved the highest weighted kappa (0.55; 95% CI 0.52, 0.58), demonstrating moderate agreement with the gold standard (143). After stratifying ZIP Codes according to rural/urban location (N=90,999 rural centroids within 475 rural ZIP Codes; N=603,067 urban centroids within 1,199 urban ZIP Codes), area weighting demonstrated substantial agreement with the gold standard in rural locations (weighted kappa 0.63; 95% CI 0.57, 0.69). Population (weighted kappa 0.40; 95% CI 0.37, 0.44) and road weighting (0.42; 95% CI 0.39, 0.46) demonstrated fair to moderate agreement with the gold standard.

Gold standard rates were dichotomized using the 75th percentile (exposed if ≥0.004 lb/ac) to calculate likelihood ratios (LRs) according to each ZIP Code-level pesticide exposure category (Table 10). Across all ZIP Code metrics, LRs >1 indicate a higher probability of exposed gold standard rates compared to unexposed gold standard rates classified into the specified ZIP Code-level pesticide exposure categories. LRs <1 indicate a higher probability of unexposed gold standard rates vs. exposed rates being classified into the specified ZIP Code-level pesticide exposure category. Overall, area weighting achieved the highest LR for the high exposure category, where a truly exposed gold standard rate was 30.91 times more likely to be classified as highly exposed compared to an unexposed gold standard rate (overall 30.91; rural 45.00; urban 27.49). Area weighting was also associated with the greatest spread between the none and high categories (i.e., none category with low LR close to 0 and high category with high LR >1; 30.59), indicating its capacity to discriminate between exposure and non-exposure to pesticides (191). The highest likelihood ratios for the high exposure category in rural areas were observed when using population and road weighting.

Figure 13 includes dot plots categorizing gold standard rates according to ZIP Code-defined categories. Gold standard rates distributed towards higher values in accordance with ZIP Code pesticide exposure class. However, gold standard rates of 0 lb/ac were present in all ZIP Code pesticide exposure categories, although occurring less frequently with increasing ZIP Code exposure.

The proportion of census blocks within each ZIP Code exposed to pesticides (>0.004 lb/ac; 75th percentile of gold standard rates) was examined according to rural/urban location. Overall, 44.74% (N=749) of the 1,674 ZIP Codes with >0 population contained no census blocks exceeding the 75th percentile rate. On average, a higher proportion of the census blocks within

the 475 rural ZIP Codes (26.61%) exceeded the 75th percentile rate compared to urban ZIP Codes (24.01%).

3.5 DISCUSSION

Beyond trying to capture spatial data at as fine a scale as possible, spatial data are ideally captured at the scale at which the phenomenon under study operates. In practice, many issues may arise that can prevent spatial data from being available at an operational scale, including protecting patient confidentiality. In the context of agricultural pesticide exposure estimation in California, geocoded residential data are ideal for estimating pesticide exposure. However, data availability can constrain the analysis scale to an aggregated areal unit such as the ZIP Code. In our study, different methods of aggregating PLSS section information to the ZIP Code level were explored and compared according to how well they were able to estimate the pesticide exposure of census blocks classified according to ZIP Code.

We compared ZIP Code pesticide application rates to a gold standard method, which considered pesticide-treated PLSS sections intersecting a 500 m buffer around census block-habitable class polygon centroids (24). We utilized population count data at the finest scale available via the U.S. Census Bureau - the census block - to determine where Californians likely reside and the pesticide exposure likely experienced by these individuals at these locations. USGS dasymetric mapping, a validated method that takes advantage of land cover information, typically to distribute population totals to habitable areas within areal units (192), was used to determine habitable areas in which to intersect with census blocks to calculate gold standard pesticide exposure.

The research goal was to design a ZIP Code pesticide exposure metric that best reflected the gold standard. As the source units (PLSS sections) are smaller in size relative to the target units (ZIP Codes), three different weighting methods were explored to scale up PLSS sections to the ZIP Code level: area, population, and network-attributed/road. Ancillary data was used to weight PLSS section-level rates. As PLSS sections are typically 1 mi^2, area weighting was the simplest method to apply while both population and road weighting were intuitive approaches that directly addressed the number of individuals living within each PLSS section. Population weighting used census block population totals areally interpolated into PLSS sections to weight rates, while network-attributed weighting used roads intersecting each PLSS section to weight rates. ZIP Codes are linear features representing mail delivery, which are directly associated with roads, and using roads to weight rates directly addressed human activity within the ZIP Codes as individuals live and work in and around roads (193).

In practice, ecologic metrics can be used in a variety of ways in epidemiologic studies. A ZIP Code pesticide exposure metric, for example, could be used as an independent variable in an ecologic epidemiologic study where the outcome is similarly aggregated to avoid a spatial mismatch (194). An alternative approach could be to use the ZIP Code metric as a predictor in a hierarchical (multilevel) model to study an individual-level outcome while taking into account clustering within the ZIP Code, as individuals within the same geographic unit will be more similar and ignoring such a lack of independence would lead to inaccurate standard errors (195).

3.5.1 Accuracy Assessment

Area weighting was superior to both population and road weighting in terms of the weighted kappa and some likelihood ratio results. Area weighting achieved moderate agreement with the

gold standard and substantial agreement among rural locations (143). Area-weighted LRs increased with increasing ZIP Code pesticide exposure (none to high) and were associated with the greatest spread. The LR observed in the high pesticide exposure category was 30.91, while an even higher LR of 45.00 was observed among rural locations, which are both greater than 10, demonstrating strong evidence of 'ruling in' pesticide exposure (196). In the same vein, LRs for the none pesticide exposure category (overall 0.32; rural 0.11; urban 0.37) were <1, demonstrating evidence of 'ruling out' pesticide exposure. Furthermore, improved accuracy results in rural locations lends support to the utility of area weighting approaches in a geographic setting where pesticide exposure is both more prevalent and important to accurately estimate. It is interesting to note that population and road weighting achieved significantly higher LRs in rural areas compared to area weighting, which potentially demonstrates their effectiveness in correctly classifying rates that are truly pesticide-exposed in these areas, especially in the high pesticide exposure category. Yet when considering weighted kappa and LR results in addition to being less time and resource intensive, area weighting represents an acceptable method to employ. Furthermore, area weighting was associated with the LR closest to 0 for the none pesticide exposure category in rural areas, demonstrating evidence of 'ruling out' pesticide exposure and reflecting high specificity in rural areas.

Area weighting is a straightforward technique that can be applied within any GIS to take into account the extent to which the geographic units under study intersect. Area weighting did not modify the source units (PLSS sections) when spatially aggregating to the ZIP Code level, which is similar to how the gold standard rates were calculated (i.e., pesticide-treated PLSS sections intersecting buffers were directly used). Although area weighting did not use any ancillary data beyond what was provided regarding the spatial intersection of the PLSS sections

and ZIP Codes, this method still incorporates all pesticide-treated and untreated PLSS sections within a ZIP Code when deriving a pesticide application rate to represent the exposure experience of those residing in that ZIP Code. On the other hand, population and road weighting were calculated in a way that provided a measure of how important pesticide exposure is with respect to how many individuals within a particular PLSS section are potentially affected. Population and road weighting then had the effect of altering the source units so that resultant weighted rates were in less agreement, although still in fair to moderate agreement, with the gold standard as compared to area-weighted rates.

As with all studies exploring the use of an aggregated spatial unit, the modifiable areal unit problem (MAUP) is a prominent issue, referring to observing different patterns and relationships as a result of how the data are aggregated (197). Two manifestations of the MAUP are the aggregation effect, or observing different results due to different hierarchical nesting of units (e.g., census block rates vs. census block group rates), and the zoning effect, or observing different results due to different methods of partitioning geographic space (e.g., census tracts vs. ZIP Codes). Therefore, it is important to note that different accuracy assessment results may be observed if examining pesticide exposure aggregated to a different spatial unit (e.g., census tract).

3.5.2 Strengths

To the best of our knowledge, this is the first study to explore the accuracy of an ecologic metric to estimate agricultural pesticide exposure. Research investigating optimal weighting methods to spatially aggregate data is important to address as larger scale geographic data are often unavailable. The framework used to execute spatial aggregation in this study can be applied to

study other environmental exposures and other geographic units, particularly when the goal is to develop a metric that estimates exposure for a specific population for use in an epidemiologic study. All of the data used to execute area weighting can be implemented in any basic GIS software.

3.5.3 Limitations

Attempting to reconcile the discrepancy between the operational and analysis scales via spatial aggregation and weighting inevitably manifests in some misclassification. In this study, attempting to use a ZIP Code to estimate large-scale census block resulted in ZIP Code rates typically overestimating exposure. For example, the dot plots show a large proportion of 0 lb/ac gold standard rates misclassified in all ZIP Code pesticide-exposed categories (low, low-moderate, moderate, and high). In other words, many false positives were observed, where 0 lb/ac gold standard rates were inaccurately classified as pesticide-exposed using ZIP Codes. It would useful to explore the extent to which using ZIP Code pesticide application rates biases results compared to using a gold standard method in an epidemiologic analysis, as the lb/ac differences between true and aggregated results may or may not substantially impact results.

The gold standard rates were calculated using the centroid of each populated census block-habitable class polygon. Conservation easements were incorporated into the dasymetric method as geographic areas in which individuals do not reside. Although easements are devoted to open space purposes, some California counties allow minimal development within easements for residential purposes. For example, a land parcel between 20 and 39 ac may have ≤5% of its area developed for residential use (198). Therefore, some easement areas may have been misclassified as uninhabitable when the area could have actually been considered when creating

census tract-habitable class polygons. However, a small proportion of California is associated with easements (2,716 mi^2; 1.71% of California's total area).

As PLSS sections intersecting a 500 m buffer around these centroids were used to calculate rates, it is conceivable that individuals may not reside at each centroid's location. Although great care was taken to determine habitable areas within each census block using dasymetric methods, more comprehensive approaches could be employed to locate residential locations to estimate pesticide exposure, including incorporating tax parcel data, elevation, or generating multiple points within habitable areas as opposed to one centroid. Furthermore, constraining the study area to the state of California may have underestimated exposure for individuals residing along the edges of the state, where pesticides applied in neighboring states (Oregon, Nevada, and Arizona) would impact exposure estimates. Such exposure would have affected both gold standard and ZIP Code rates. However, the majority of lands along California's perimeter were coded as water/excluded using the dasymetric method (data not shown). Therefore, only a small proportion of California's population likely resided along the state's periphery.

3.6 CONCLUSIONS

The aggregated spatial scale at which data is available may necessitate the use of a spatial aggregation method to scale up the data for use in epidemiologic settings. As a way to reconcile the differences between the scale at which a phenomenon of interest operates and the scale at which the data is available, weighting methods using ancillary data relevant to the feature under study can be explored. Area weighting of pesticide application rates from the PLSS section level

to the ZIP Code level, applied by weighting PLSS section rates according to its percentage area within the ZIP Code, provided the most accurate results compared to a finer scale census block gold standard, as compared to population and network-attributed/road weighting. Future research should formally compare the impact of using an aggregated ZIP Code pesticide metric vs. a larger scale gold standard metric in an epidemiologic study.

3.7 TABLES

Table 8. Annual Gold Standard and ZIP Code Pesticide Application Rates (lb/ac)

	N	Min	25th pctl	Mean ± SD	Median	75th pctl	Max
Gold standard	694,087	0	0	0.08 ± 0.41	0	0.004	24.77
ZIP Code: area weight[a]	1,674	0	0	0.13 ± 0.38	0.001	0.02	5.13
ZIP Code: population weight[a]	1,674	0	0	0.16 ± 3.22	0.00002	0.001	95.29
ZIP Code: road weight[a]	1,674	0	0	0.16 ± 3.08	0.00003	0.001	95.29

Abbreviations: min, minimum; pctl, percentile; SD, standard deviation; max, maximum.
[a] Rates are expressed in lb/ac. Rates do not include ZIP Codes with a zero population according to the SF1 (90261, 91608, 94128, and 96095).

Table 9. Weighted Kappa and Agreement: ZIP Code Weighting Methods vs. Gold Standard

ZIP Code weighting method		Weighted kappa (95% CI)	Agreement (%)
Area weight[a]		0.55 (0.52, 0.58)	75.08
	Rural	0.63 (0.57, 0.69)	68.15
	Urban	0.51 (0.47, 0.55)	76.12
Population weight[b]		0.40 (0.37, 0.44)	65.40
	Rural	0.38 (0.29, 0.46)	61.15
	Urban	0.41 (0.37, 0.45)	66.04
Road weight[c]		0.42 (0.39, 0.46)	65.95
	Rural	0.40 (0.33, 0.47)	60.48
	Urban	0.43 (0.39, 0.47)	66.78

Abbreviations: CI, confidence interval.
[a] Area weight pesticide exposure categories: none (≤0.02 lb/ac); low (>0.02 to 0.07 lb/ac); low-moderate (>0.07 to 0.29 lb/ac); moderate (>0.29 to 0.67 lb/ac); high (>0.67 lb/ac).
[b] Population weight pesticide exposure categories: none (≤0.001 lb/ac); low (>0.001 to 0.002 lb/ac); low-moderate (>0.002 to 0.005 lb/ac); moderate (>0.005 to 0.02 lb/ac); high (>0.02 lb/ac).
[c] Road weight pesticide exposure categories: none (≤0.001 lb/ac); low (>0.001 to 0.003 lb/ac); low-moderate (>0.003 to 0.01 lb/ac); moderate (>0.01 to 0.02 lb/ac); high (>0.02 lb/ac).

Table 10. Likelihood Ratios: ZIP Code Weighting Methods vs. Gold Standard

ZIP Code weighting method	N	Centroids[a,b]		Likelihood ratio		
		Pesticide-exposed	Pesticide-unexposed	Overall	Rural	Urban
Area weight						
None (≤0.02 lb/ac)	500,932	47,892	453,040	0.32	0.11	0.37
Low (>0.02 to 0.07 lb/ac)	55,201	22,071	33,130	2.00	2.09	2.11
Low-moderate (>0.07 to 0.29 lb/ac)	53,030	33,123	19,907	4.99	6.13	4.82
Moderate (>0.29 to 0.67 lb/ac)	50,103	38,695	11,408	10.18	8.59	9.68
High (>0.67 lb/ac)	34,821	31,740	3,081	30.91	45.00	27.49
Population weight						
None (≤0.001 lb/ac)	515,544	56,768	458,776	0.37	0.37	0.37
Low (>0.001 to 0.002 lb/ac)	61,948	30,618	31,330	2.93	8.67	2.51
Low-moderate (>0.002 to 0.005 lb/ac)	54,371	37,408	16,963	6.62	13.78	6.53
Moderate (>0.005 to 0.02 lb/ac)	43,204	31,457	11,747	8.03	42.64	7.68
High (>0.02 lb/ac)	19,020	17,270	1,750	29.61	200.32	26.62
Road weight						
None (≤0.001 lb/ac)	507,893	53,216	454,677	0.35	0.35	0.35
Low (>0.001 to 0.003 lb/ac)	66,640	32,701	33,939	2.89	7.13	2.72
Low-moderate (>0.003 to 0.01 lb/ac)	57,345	37,985	19,360	5.89	12.30	5.44
Moderate (>0.01 to 0.02 lb/ac)	40,758	30,714	10,044	9.17	48.18	8.42
High (>0.02 lb/ac)	21,451	18,905	2,546	22.28	187.64	21.53

[a] Centroids of the census block-habitable class polygons were used to calculate gold standard rates. The 75thpercentile of gold standard rates (0.004 lb/ac) was used to determine whether or not a centroid was exposed (≥0.004 lb/ac) or unexposed (<0.004 lb/ac).

[b] Pesticide-exposed and unexposed centroids, used to calculate gold standard rates, were tabulated by ZIP Code location and compared to ZIP Code rates calculated using three different methods, each applying a different weighting method to aggregate pesticide exposure from the PLSS section to ZIP Code level.

3.8 FIGURES

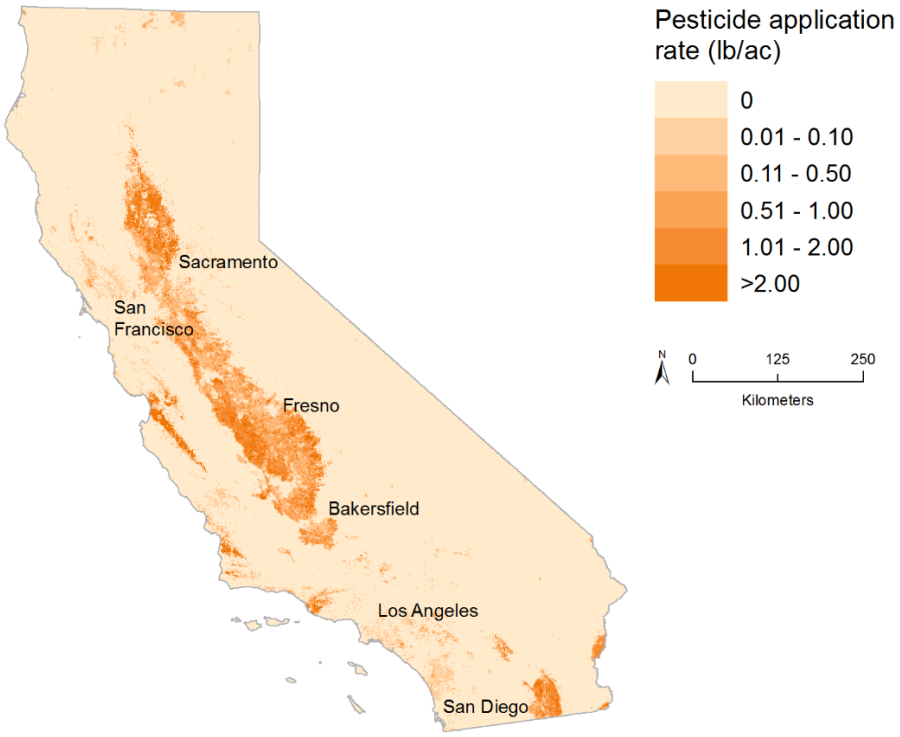


Figure 11. Annual PLSS Section Pesticide Application Rates (lb/ac) From 1995 to 2005 in California

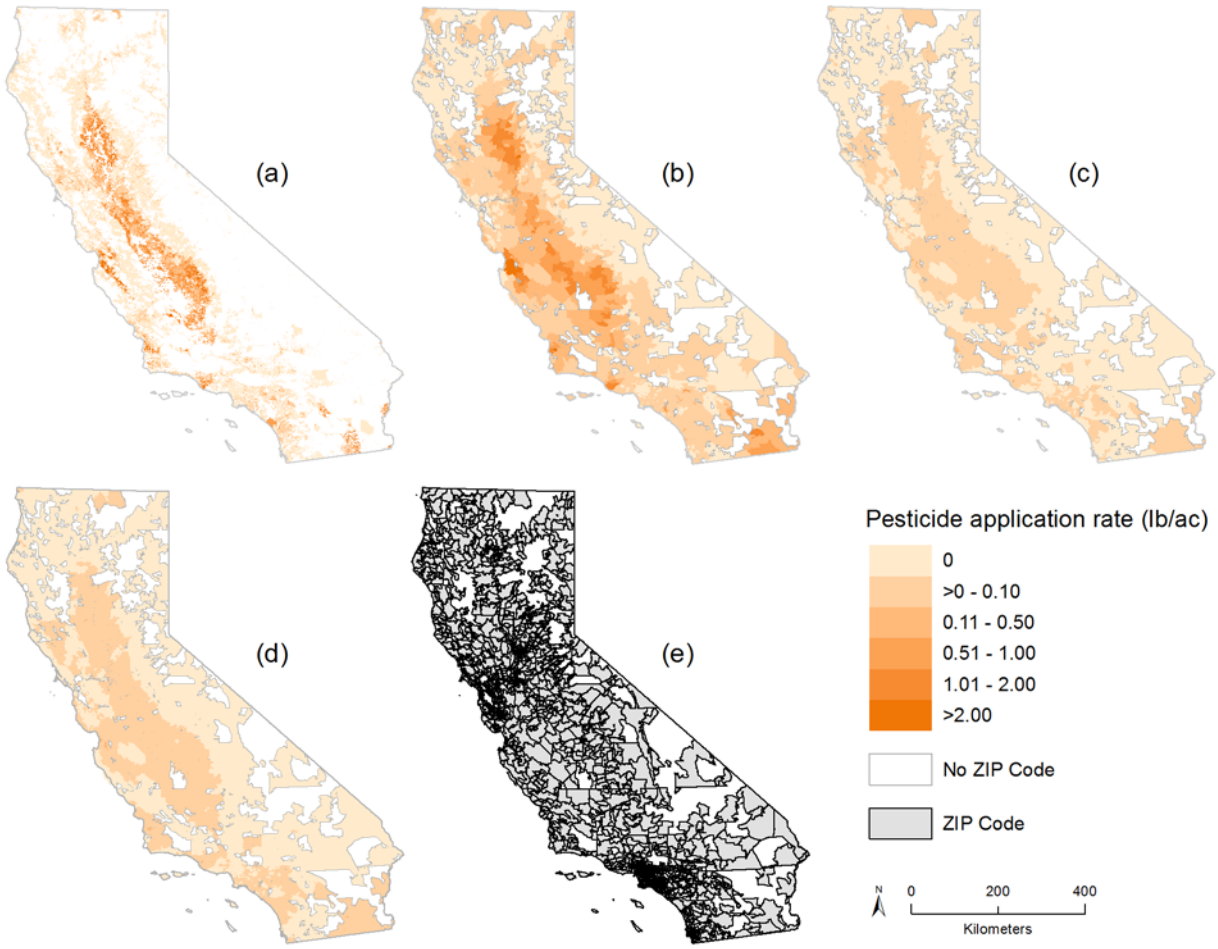


Figure 12. Annual Pesticide Application Rates (lb/ac) Calculated Using the (a) Gold Standard, (b) Area Weighting, (c) Population Weighting, and (d) Road Weighting; (e) ZCTA Boundaries Shown for Reference

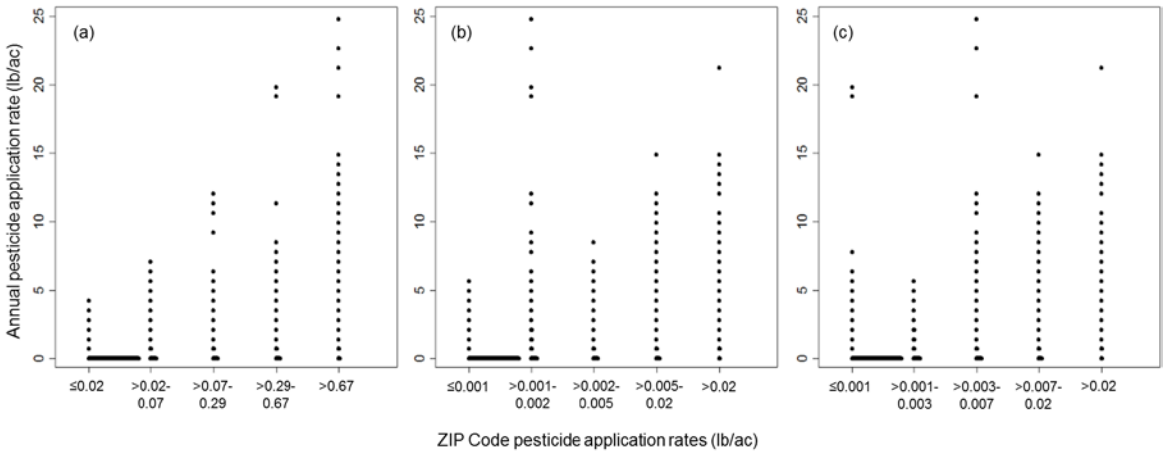


Figure 13. Dot Plots of Gold Standard Pesticide Application Rates (lb/ac) Categorized According to ZIP Code None, Low, Low-Moderate, Moderate, and High Categories: (a) Area Weight, (b) Population Weight, and (c) Road Weight

4.0 USING GIS TO LINK SEER-MEDICARE AND CALIFORNIA PESTICIDE DATA: A POPULATION-BASED CASE-CONTROL STUDY OF PESTICIDE EXPOSURE AND HEPATOCELLULAR CARCINOMA RISK

4.1 ABSTRACT

Background: Hepatocellular carcinoma (HCC), or primary liver cancer, is associated with low survival. U.S. studies examining pesticide exposure in relation to HCC have demonstrated inconclusive results, relying on self-reported exposure. Objective: We aimed to clarify the association between agricultural pesticide exposure and HCC by implementing a novel data linkage between Surveillance, Epidemiology, and End Results (SEER)-Medicare and California Pesticide Use Report (PUR) data using Medicare ZIP Codes in a geographic information system (GIS). Methods: HCC cases diagnosed between 2000 and 2009 in California were frequency-matched to controls by year, age, race, sex, and duration of residence in California. Potential confounders were extracted from Medicare claims. From 1974 to 2008, PUR pounds of applied organophosphate, organochlorine, and carbamate pesticides were summed according to 1-square-mile Public Land Survey System (PLSS) sections and aggregated to the ZIP Code level using area weighting in a GIS. ZIP Code estimates were linked to subjects using Medicare ZIP Codes to calculate pesticide exposure. Multivariable conditional logistic regression estimated the association between pesticide exposure and HCC. Results: Among rural California residents,

annual exposure to over 0.06 pounds/acre of organochlorine pesticides (median among controls) was associated with an increased risk of HCC after adjusting for liver disease and diabetes (adjusted odds ratio [OR] 1.52; 95% confidence interval [CI] 1.02, 2.28 P = 0.0415). Risk increased after accounting for a 20-year exposure lag (adjusted OR 1.81, 95% CI 1.19, 2.75; P = 0.0058). Conclusions: This is the first epidemiologic study using GIS to study pesticide exposure and HCC. Given potential evidence of organochlorine pesticides increasing HCC risk, future research should explore usage of finer spatial resolution data.

4.2 INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most common cancer in the world and the second leading cause of cancer-related death (199). Between 70 and 85% of primary liver cancer cases are HCC (200). HCC incidence is highest in East Asia and Africa but has been rising in the U.S. (1, 12). U.S. HCC incidence, adjusted to the 2000 U.S. Standard Population, significantly increased 29% from 4.4 per 100,000 (2000 to 2004) to 5.7 per 100,000 (2005 to 2009) (rate ratio 1.29, 95% confidence interval: 1.27, 1.32, $P < 0.0001$) (201). Overall, HCC incidence between 2000 and 2009 was 5.1 per 100,000. Rising incidence has been attributed to increasing rates of obesity and diabetes, hepatitis C virus (HCV) infection rates peaking in the 1960s and 70s and HCC clinically manifesting 20 years later, and improving survival among cirrhosis patients (3, 200).

In the U.S., HCC is more common among males and among individuals of Asian descent (3). The mean age at diagnosis is 64 years (median 63) (202). The majority of adult-onset HCC cases occur sporadically, or among individuals with no similarly affected first-degree relative

(3). Early stage HCC is treatable via liver transplantation, surgical resection, and ablation (203). Embolization (occlusion of tumor's blood supply) or chemoembolization (combined chemotherapeutic and embolization agents) may improve survival in patients with later stage HCC. However, many HCC cases are diagnosed at a regional or distant stage (49% between 2000 and 2009) (202), which contributes to the low five-year 16.6% relative survival rate in the U.S. (12).

Predominant HCC risk factors in high-risk areas, such as Asia, include chronic hepatitis B virus (HBV) infection and consumption of aflatoxin-contaminated foods (3). Predominant risk factors in low-risk areas, such as the U.S., include chronic HCV infection and heavy alcohol consumption (>50 to 70 g per day) (3). Approximately 64.5% (95% confidence interval: 63.3, 65.6) of all HCC cases occurring in the U.S. population aged 68 years and older are attributed to HCV, HBV, alcoholic liver disease (e.g., alcoholic cirrhosis of liver), rare metabolic disorders (e.g., hemochromatosis), and diabetes and/or obesity (15). Most of these risk factors contribute to the formation and progression of cirrhosis, or scarring of the liver (20). Between 70 and 90% of all HCC cases occur within an established background of chronic liver disease and cirrhosis (3, 20). Although HCV, HBV, and heavy alcohol consumption are the major risk factors for cirrhosis among HCC cases in the U.S., between 15 and 50% of all HCC cases have no established risk factors (3, 6).

Epidemiologic studies have shown that pesticide exposure may increase the risk of HCC. Pesticides are chemicals used frequently in agriculture to treat pests such as insects (27, 204). Pesticides are hypothesized to contribute to liver carcinogenesis through mechanisms of genotoxicity, tumor promotion, immunotoxicity, and hormonal action (21, 27). Several case-control studies conducted in China demonstrated statistically significant increased risks for HCC

(55, 56, 58). Persson et al. (56) showed that the highest quintile of serum dichlorodiphenyltrichloroethane (DDT) (≥810 ng/g), an organochlorine pesticide, compared to ≤261 ng/g significantly increased HCC risk after adjusting for risk factors including age, hepatitis B surface antigen (HBsAg), and alcohol consumption (adjusted odds ratio [OR] 2.96; 95% confidence interval: 1.19, 7.40). However, some studies have shown inconclusive results. In the U.S., three studies reported non-significant increased risks for HCC among those employed in farming (59, 61, 69). However, farming in New Jersey conferred significantly higher risk for HCC compared to no employment in this occupation (adjusted OR 3.20; 95% confidence interval: 1.11, 9.21) (68).

In the U.S., pesticide exposure occurs most frequently via diet (82). Additional routes of exposure include via occupation (e.g., pesticide application), drinking water, and very importantly, residential proximity to agricultural pesticide applications. Applied pesticides can drift from their intended sites through the air and ground via spray drift and post-application volatilization (24). Vulnerable populations include rural residents and farming families (86), as pesticides can enter homes through drift and from clothing (82). Gunier et al. (88) demonstrated that residential proximity within 1,250 m to pesticide-treated crops in California was significantly correlated with pesticide concentrations in sampled carpet dust. Pesticides are less likely to degrade within homes due to the absence of moisture, sunlight, and microorganisms (82, 88), and humans can be subsequently exposed via dermal contact and ingestion (89).

To clarify the relationship between pesticide exposure and HCC in the U.S., we conducted a population-based case-control study in California, the most agriculturally productive state in the U.S. (205). Cases and controls, in addition to claims used to identify comorbidities, were derived from the Surveillance, Epidemiology, and End Results (SEER)-Medicare database.

Pesticide exposure was estimated using California Pesticide Use Reports (PURs). We implemented a novel data linkage between SEER-Medicare and PUR data using Medicare ZIP Codes in a geographic information system (GIS).

4.3 METHODS

4.3.1 Study Population

SEER-Medicare represents a data linkage between SEER cancer data and Medicare claims (206, 207). SEER is a National Cancer Institute (NCI) program collecting information on cancer incidence and survival from 18 population-based cancer registries covering 28% of the U.S. Medicare is a U.S. federal health insurance program for qualifying individuals ≥65 years old, covering 97% of this age group, in addition to those <65 years with end-stage renal disease (ESRD) or medical disability. All Medicare beneficiaries are entitled to Part A (hospital insurance), approximately 96% enroll in Part B (medical insurance), 24% enroll in Medicare Advantage or a managed care plan (e.g., health maintenance organization [HMO]), and 38% enroll in Part D prescription drug coverage (206, 208). Part C does not process bills through Medicare. The SEER-Medicare data linkage includes all SEER cancer cases who are found in the Medicare Enrollment DataBase. Medicare claims are linked to cases via personal identifiers, e.g., Social Security number. The 2012 data linkage includes SEER cases from 1991 to 2009 and Medicare claims from 1991 to 2010, which is 94% successful among those 65 years and older (3% of elderly do not receive Medicare and 3% have insufficient linkage information). SEER cancer data for the entire state of California has been available since 2000 (209).

4.3.2 Case and Control Ascertainment

Eligible study participants were not of unknown race and ≥66-year-old California residents with at least 13 months of continuous Parts A and B, non-HMO coverage and at least one California ZIP Code by the time of diagnosis/selection. Cases were defined using the following criteria: International Classification of Diseases for Oncology, Third Edition (ICD-O-3) topography code C22.0 (primary liver cancer) and ICD-O-3 histology codes 8170 to 8175 (210); diagnostic confirmation (e.g., positive histology) excluding clinical diagnosis only (211); sequence number 00 or 01; reported to a California cancer registry; diagnosed between 2000 and 2009; and not reported via autopsy or death certificate only. Controls were selected from a 5% random sample of Medicare beneficiaries residing in SEER geographic areas who are not found in any SEER data (i.e., they are not diagnosed with any cancer according to SEER). Cases included in the 5% random sample were considered during control selection. For each year between 2000 and 2009, eligible controls who were not a case and alive as of July 1 of that year were enumerated. Eligible controls may have included cases diagnosed after July 1 in the year they were selected. Cases and controls were frequency-matched according to age, sex, race (white, black, Asian, other, Hispanic, Native American), and years of non-continuous California residence (using available Medicare ZIP Codes not carried back, categorized using tertiles among cases: 1-5, 6-10, ≥11). Controls were sampled with replacement.

4.3.3 Pesticide Exposure

Agricultural pesticide exposure was estimated by linking California Department of Pesticide Regulation (CDPR) Pesticide Use Report (PUR) data with available Medicare ZIP Codes from

1991 until the year before diagnosis/selection (last billing ZIP Code in that year) in a GIS (212). PUR data include pounds of applied pesticides, crop type, chemical name, date, and Public Land Survey System (PLSS) 1 mi^2 section of application (213). Agricultural use PURs from 1974 to 2008 were checked for errors (e.g., duplicates). Outlier application rates (pounds per acre; lb/ac), defined using CDPR flags (e.g., >200 lb/ac if non-fumigant pesticide) from 1990 to 2008 and as rates >200 lb/ac (>1,000 lb/ac if fumigant) or 50 times the median rate for all uses of a given pesticide product, crop, unit type, and record type from 1974 to 1989, were replaced with the statewide median rate for that pesticide in that year (24). Pounds of active ingredient were recalculated using PUR number of treated acres. Pesticide applications belonging to the organophosphate, organochlorine, and carbamate chemical classes, which have been previously associated with HCC, were extracted (54, 56, 62). For each year between 1974 and 2008, pounds were matched and summed according to PLSS section and divided by section acreage to calculate pesticide application rates (lb/ac) (185, 214). PLSS sections and California TIGER/Line® ZIP Code Tabulation Areas (used to approximate ZIP Code boundaries) were intersected (215). PLSS section rates were spatially aggregated to the ZIP Code level using area weighting, where section rates were weighted by the proportion of the ZIP Code area comprised by that section. For each study subject, using available California ZIP Codes from 1991 up until the year before diagnosis/selection and carrying back the earliest available ZIP Code to 1974, ZIP Code pesticide application rates were summed and divided by the number of years of California residence. This study examined two ZIP Code pesticide exposure metrics: annual pesticide application rates and average annual applied pounds of pesticides.

4.3.4 Covariates

The following were extracted from inpatient (Part A), outpatient (Part B), and carrier (e.g., physician) Medicare claims: HCV (ICD-9-CM [Ninth Revision, Clinical Modification] codes 070.41, 070.44, 070.51, 070.54, 070.70, V02.62), HBV (070.22, 070.23, 070.32, 070.33, V02.61), unspecified hepatitis (070.9, 070.59, 070.49, 571.4, 571.8, 571.9), diabetes (250), obesity (278.00, 278.01, 278.02, V77.8, 259.9), alcoholic liver disease (571.0, 571.1, 571.2, 571.3; 571.5 or 571.6 in the presence of 303, 291, 305.0, V11.3, or V79.1), non-specific cirrhosis (571.5 or 571.6 not in the presence of HCV, HBV, unspecified hepatitis, or alcoholic liver disease), rare genetic disorders (α1 antitrypsin deficiency 273.4, hemochromatosis 275.0, porphyria 277.1, tyrosinemia 270.2, Wilson disease 275.1), human immunodeficiency virus (HIV) (042, V08), and smoking (V15.82, 305.1, 989.84; ever-smoking as there is not enough information to identify former smokers) (15, 211, 216). Conditions were considered present if there was a single Part A diagnosis or two Part B or carrier claim diagnoses separated by at least 30 days (206). Due to differential availability of claims data depending on when cases were diagnosed (claims from 1991 to 2010 if diagnosed before 2003; from 1998 to 2010 if diagnosed between 2003 and 2005; from 2000 to 2010 if diagnosed between 2006 and 2007; and from 2002 to 2010 if diagnosed between 2008 and 2009) (207), Medicare claims within six years of diagnosis/selection were examined. Claims within one year of diagnosis/selection were excluded due to potential medical detection bias (206). As eligible study subjects were required to have at least 13 months of continuous Parts A and B, non-HMO enrollment prior to diagnosis/selection, no Medicare claim diagnosis codes identifying a particular health condition indicated the absence of that condition (i.e., there were no missing variables in our study). State buy-in, or Medicare Savings Programs where states pay for Medicare premiums, deductibles, and/or coinsurance due

to limited income, was used as an indicator for socioeconomic status (211, 217). State buy-in was considered present if a subject was enrolled in Parts A or B state buy-in at any time point beginning in the year before diagnosis/selection. The proportion of each ZIP Code's ≥16-year-old population employed in the agriculture industry was provided by the 2000 Census Summary File 3 (218). The Medicare ZIP Code in or closest to 2000 was matched to Census data.

4.3.5 Statistical Analysis

Pesticide exposure was examined using all combined pesticide chemical classes (organophosphates, organochlorines, and carbamates) and each class separately. Exposure defined as a rate and as applied pounds were modeled separately. We used random-intercept logistic regression to explore the extent to which cases and controls within the same ZIP Code might have similar risk for HCC and thus, pesticide exposure. After exploring a random intercept defined as the ZIP Code at diagnosis/selection and the ZIP Code occurring most frequently, a low intraclass correlation coefficient indicated little variability between clusters (ZIP Codes) and that a random intercept was not necessary. Univariable conditional logistic regression using robust variance estimation and taking into account the frequency matching factors of year, age, sex, race, and California residence was used to assess the association between each variable and case control status. Chi-square, one-way analysis of variance (ANOVA), and Kruskal-Wallis tests evaluated the association between each variable and pesticide exposure (i.e., independent variable of interest). Using variables significantly associated with HCC and/or pesticide exposure ($P < 0.05$), backward elimination methods ($P > 0.20$ removed), confirmed with forward selection ($P < 0.20$ to enter), were utilized to build final models in which significant predictors ($P < 0.05$) were included. Regression diagnostics were performed on final models. Odds ratios (ORs) and

95% confidence intervals (CIs) were estimated using the final multivariable conditional logistic regression models (robust variance estimation) taking into account frequency matching factors. The effect of 10-, 15-, and 20-year lags were examined, where pesticide exposure occurring outside of the lag window (before diagnosis/selection) was considered.

In a supplemental analysis, statistical analyses were limited to rural California residents, defined as individuals who resided in California ZIP Codes associated with ≥0.20 lb/ac the majority of the time from 1974 until the year before diagnosis/selection. This cutoff was determined by comparing ZIP Code pesticide application rates (all classes) for each year from 1974 to 2008 between rural and urban ZIP Codes (defined using Rural-Urban Commuting Area [RUCA] categorization C codes) (219). Rural ZIP Codes were typically characterized by rates ≥0.20 lb/ac. Usage of a pesticide application rate to identify rural residents addresses how rurality occurs on a continuum, where a geographic area defined as rural according to a given metric may be sparsely populated, but is associated with few agricultural land uses and thus minimal agricultural pesticide exposure (220). Categorical pesticide exposure was explored using cutoffs defined by the 50th percentile of rates and 75th percentile of applied pounds among rural controls. Interactions between pesticide exposure and each covariate and matching factor were examined. All reported *P* values are two-sided. Analyses were conducted in 2014 using SAS, version 9.4 (SAS Institute, Inc., Cary, North Carolina).

4.4 RESULTS

There were 3,034 hepatocellular carcinoma cases in California diagnosed between 2000 and 2009 and 14,991 frequency-matched controls included in the analysis. Among 10,408 individuals

diagnosed with HCC as a first cancer in California between 2000 and 2009 who were not of unknown race and in the Medicare Enrollment DataBase, 29% were included in our study. Figure 14 shows cases excluded from our study according to each eligibility criterion. Thirty-three percent of all considered cases were excluded due to age and 29% due to HMO coverage. A comparison of characteristics for included vs. excluded HCC cases is shown in Table 11. Excluded cases were more likely to be urban-dwelling younger white males of higher socioeconomic status having resided in California for a shorter period of time (reflecting younger age). Table 12 compares the SEER-Medicare HCC source population of the case group to all SEER HCC cases in California, or all HCC cases irrespective of having Medicare coverage. Similarly, Table 12 also compares the included case group to all HCC cases in California over age 65 years, which reflects the Medicare elderly in California (most individuals ≥65 years receive Medicare) to whom we would like to generalize our results. Our case series (n = 3,034) compared to SEER HCC cases aged 65 years and older (n = 7,185) was slightly older, comprised of fewer whites, and less urban. However, differences in age and race are affected by SEER not individually reporting diagnosis ages >85 years and not excluding Hispanics from any race categories, respectively.

Table 13 presents population characteristics of cases and controls included in our study. Cases were on average 75.1 years old (median 74.0), typically males, of white race, and residing in California for over 6 years. By design, matching factors did not differ between cases and controls. When considering the time period of six years before diagnosis/selection during which claims were examined for health conditions, most cases (75.1%) and controls (75.1%) contributed between 4.1 and 6.1 years of claims to the study. As expected, a higher proportion of cases were diagnosed with HCV, HBV, unspecified hepatitis, alcoholic liver disease, non-

specific cirrhosis, diabetes, obesity, rare genetic disorders, and smoking ($P < 0.0001$). Cases were more likely than controls to be of low socioeconomic status (i.e., enroll in state buy-in) and reside in an urban area at diagnosis ($P < 0.0001$). Controls typically resided in ZIP Codes with a slightly higher percentage of ≥16-year-olds employed in agriculture ($P = 0.0002$). A lower, or comparable, proportion of cases vs. controls were exposed to both moderate and high levels of ZIP Code-level pesticide exposure across all pesticide classes when examining both rates and applied pounds (Table 14). Among controls, high pesticide exposure to all classes was more frequent among younger whites with alcoholic liver disease living in California for a longer period of time, while low pesticide exposure occurred more frequently among those with HCV, HBV, state buy-in, and urban residence (Table 15). Table 16 presents results for the random-intercept logistic regression; a low ICC of 0.03 was observed when defining the random intercept as the ZIP Code at diagnosis/selection and the ZIP occurring most frequently. The majority of the total variance is due to within-cluster, or within-ZIP Code, variance and thus the random intercept was not included in subsequent models.

After adjusting for liver disease, diabetes, rare genetic disorders, and state buy-in, previous ZIP Code-level exposure to organophosphates, organochlorines, and carbamates was not associated with HCC (Table 17). Similar null effects were observed when separately examining pesticide chemical classes according to rates and applied pounds. Pesticide exposure was not significantly associated with HCC after taking into account exposure lags (data not shown).

Among study subjects identified as rural residents ($n = 306$ cases and $n = 1,758$ controls), liver disease, diabetes, obesity, rare genetic disorders, HIV, smoking, and state buy-in were more common among cases (Table 18). Organophosphate, organochlorine, and carbamate exposure

(rates) were univariably associated with HCC, where moderate and high vs. low exposure conferred higher risk for HCC (Table 19). Among rural controls, high pesticide exposure was significantly associated with Hispanic race, diabetes, state buy-in, and ZIP Code agricultural occupation (Table 20). Organochlorines were the only pesticide chemical class remaining significant after adjustment for other risk factors. When examining exposure lags of 15 and 20 years, a 1 lb per ac increase in previous annual ZIP Code-level organochlorine pesticide exposure was significantly associated with a 2.74-fold and 2.43-fold increase in HCC risk, respectively (Table 21). A 1 lb per ac increase may be interpreted as a large increase as the median organochlorine pesticide application rate for the 220 rural ZIP Codes in which rural residents resided ranged between 0 and 0.10 lb/ac from 1974 to 2008. Fifteen of these 220 rural ZIP codes were associated with organochlorine pesticide application rates exceeding 1 lb per ac in any given year. Significant results irrespective of exposure lag were observed when considering pesticide application rates dichotomized using the median among controls. For example, past ZIP Code-level exposure to ≥0.06 lb/ac of applied organochlorine pesticides was significantly associated with a 52% increase in HCC risk after adjustment for liver disease and diabetes. Organochlorine exposure (as categorized rates) conferred increasingly higher risk for HCC in the time period of 10, 15, and 20 years before diagnosis.

Organochlorine pesticide exposure defined as applied pounds (continuous) demonstrated statistically significant results, although confidence intervals were wide (Table 21). There were no significant interactions between pesticide exposure and liver disease, diabetes, or any matching factors (data not shown). Furthermore, among rural residents with no known risk factors for HCC (n = 106 cases, n = 1,412 controls), previous organochlorine pesticide exposure conferred significantly higher risk for HCC when examining categorized application rates,

particularly when examining a 15-year exposure lag (OR 2.15; 95% confidence interval: 1.29, 3.58) (Table 22). Sensitivity analyses excluding individuals originally entitled to Medicare due to disability and/or end-stage renal disease demonstrated both stronger and more significant positive associations between organochlorine pesticide exposure and HCC among rural residents (Table 23).

4.5 DISCUSSION

Among rural California residents in the SEER-Medicare population, previous ZIP Code-level exposure to organochlorine pesticides was significantly associated with an increased risk for HCC after adjusting for liver disease and diabetes. Implementing organochlorine pesticide exposure lags of 10, 15, and 20 years before diagnosis/selection was also significantly associated with increased HCC risk. HCC is a significant public health concern, rising in incidence and associated with low survival in the U.S. As HCC occurs among many individuals with no known risk factors (3, 6), it is important to explore the role of other exposures that might contribute to liver carcinogenesis. Several epidemiologic studies have demonstrated that pesticide exposure significantly increases the risk of HCC (55, 56, 58). Pesticides are pervasively used chemicals in the U.S. (23). In 2007, the agricultural sector accounted for approximately 80% of conventional U.S. pesticide usage (e.g., treating insects; 684 million pounds) (221). The three specific pesticide chemical classes explored in this study have been previously associated with HCC. Organophosphates and carbamates are mostly insecticides, widely used in the 1980s and 90s after many organochlorines were banned, but have since declined in usage in favor of more environmentally friendly chemicals (222). Some highly toxic organophosphates (e.g., parathion)

and carbamates (e.g., aldicarb) have been banned in the U.S. (223, 224). Organochlorines are also mostly insecticides, were widely used in the 1940s to 60s, but have largely been banned in the U.S. due to adverse health effects and environmental persistence (225, 226). Animal models have demonstrated that exposure to DDT, an organochlorine banned in the U.S. in 1972, and its metabolite, dichlorodiphenyldichloroethylene (DDE), lead to the development of HCC and other liver tumors (28, 29).

Three case-control studies conducted in China using serum DDT provide the most convincing evidence of an association between pesticides and HCC (55, 56, 58). In particular, these studies showed a link between HCC and DDT, which belongs to the organochlorine pesticide chemical class demonstrating a significant association with HCC among rural California residents in our study. These studies utilized biomonitoring to estimate pesticide exposure, considered the gold standard method of capturing exposure from specific chemicals via all routes of human exposure (227). However, these findings may not be generalizable to the U.S. population, as China continues to use DDT as an anti-malarial agent (56). In the U.S., the 95^{th} percentile of serum DDT levels was 28 ng/g between 1999 and 2000 and 19.5 ng/g between 2003 and 2004, compared to the Persson et al. (56) Chinese study population associated with a geometric mean of 468 ng/g (standard deviation 18) DDT among cases and 478 ng/g (18) among controls (sera collected in 1992-1993).

Other research examining pesticides in relation to HCC has demonstrated inconclusive results, ranging from significant increases to non-significant deficits in risk (54, 57, 59-70). Most studies have relied on self-reported pesticide exposure and occupation, job-exposure matrices, occupational experts, and rural residence. Recall bias may have obscured or inflated study results. Accurately quantifying pesticide exposure, particularly when investigating their role in

chronic diseases such as cancer, must consider historical exposures to take into account latency periods, or the time from initial exposure to clinical disease (23, 228). Specific pesticides may be associated with disease, which are not adequately captured with dichotomous (yes/no) classifications. Four case-control studies conducted in the U.S. accounted for few confounders (e.g., age, race, sex) during analysis and/or study design and potentially introduced a selection bias in using other cancer or hospital controls (59, 61, 68, 69).

We focused on the Medicare elderly population in California, which represented a unique opportunity to study HCC as California is the most agriculturally productive state in the U.S. (205) and is characterized by relatively high HCC incidence. California age-adjusted HCC incidence between 2000 and 2009 is 6.3 per 100,000, 24% higher than the overall U.S. rate (201). Furthermore, the California Pesticide Use Reports database is the world's most comprehensive pesticide reporting system, collecting agricultural pesticide use since 1974 (213). Prior to 1990, California farmers were only required to report restricted use pesticides. As full-use reporting began in 1990, we confirmed that pesticide application rates across all three chemical classes were not significantly different between 1974 to 1989 and 1990 to 2008 (data not shown). We sought to improve upon the limitations of previous HCC U.S. epidemiologic studies by using population-based data sources providing information on HCC cases, controls representative of the same reference population as the cases, comorbidities as potential confounders, and pesticide exposure. PURs allowed for examining specific pesticides to reconstruct historical exposure, addressing a potential latency period. Medicare ZIP Codes allowed us to link SEER-Medicare and PUR data in a GIS. GIS is a powerful method allowing for the overlay of multiple spatial data sources based on a common geographic frame of reference (9). Specifically, we were able to overlay PLSS sections, the geographic level of

reporting of PURs, with TIGER/Line® ZIP Code Tabulation Areas (to approximate ZIP Codes) to estimate pesticide exposure for each study subject.

When examining the entire sample of SEER-Medicare cases and controls, pesticide exposure was not associated with HCC after adjustment for confounders. This reflects a study design that sampled cases and controls without considering geographic area of residence, or their opportunity for pesticide exposure. Agricultural pesticide applications predominantly occur in rural, less densely populated geographic areas (86). The majority of sampled HCC cases (76.5%) resided in highly urbanized areas (counties with metropolitan areas of ≥1,000,000 population) at diagnosis, defined using SEER-Medicare-provided Rural/Urban Continuum Codes. Similarly, 72.1% of controls resided in highly urbanized areas at selection.

After restricting analyses to cases and controls with the opportunity for pesticide exposure (i.e., rural residents), previous exposure to organochlorines significantly increased risk for HCC after adjusting for liver disease and diabetes. Applied organochlorines in California between 1974 and 2008 included endosulfan, toxaphene, and dicofol. When defining pesticide exposure as an annual application rate (categorical predictor), previous exposure to organochlorines conferred between 52 and 85% greater risk for HCC compared to those not exposed after adjustment. In analyses, we considered exposure to pesticides as typically residing in ZIP Codes that applied over the median rate among rural controls, ranging between 0.06 and 0.08 lb/ac depending on the exposure lag. Risk for HCC increased when taking into account longer exposure lags of 10, 15, and 20 years. This potentially reflects temporal trends in organochlorine usage, which has dramatically declined since the 1970s. Thus, exposure occurring in the time period of at least 20 years before diagnosis between 2000 and 2009 is more relevant in elevating HCC risk, as opposed to considering exposure from all years before

diagnosis. Our results were further bolstered by demonstrating that organochlorine exposure increases HCC risk among rural residents absent of HCC risk factors, e.g., hepatitis.

We did not exclude individuals who were originally entitled to Medicare due to disability and/or end-stage renal disease. Although these individuals are not considered representative of Medicare beneficiaries in terms of their clinical and demographic characteristics (211, 229), the majority of both cases (89.2%) and controls (91.8%) in our study were entitled to Medicare due to attaining the age of 65 years. Furthermore, individuals with a disability, such as chronic liver disease and alcoholism, represent those with risk factors who would likely develop HCC (230). Sensitivity analyses excluding those entitled not due to age demonstrated stronger positive, significant associations between organochlorines and HCC among rural residents. This potentially demonstrates how this subgroup of Medicare beneficiaries, due to their health, may have been confined to their homes, minimizing agricultural pesticide exposure, or living in relatively less rural areas to facilitate their healthcare access (e.g., dialysis centers for ESRD).

4.5.1 Strengths

Our study included many strengths. We conducted the first epidemiologic study using GIS to study pesticides and HCC by implementing a novel data linkage between SEER-Medicare cancer outcomes and health conditions data with PUR pesticide applications using Medicare ZIP Codes. ZIP Codes have been used previously to study pesticides and cancer mortality in California (171). While all previous U.S. studies have relied on self-reported exposure, we utilized all agricultural use pesticide applications of specific chemicals reported to the California state government by farmers and commercial pest control operators. Our population-based study was able to sample from all HCC cases reported to California cancer registries between 2000 and

2009 as part of the SEER program who were aged 65 years and older enrolled in Medicare. Controls were sampled from the same population that gave rise to the cases, using a 5% random sample of Medicare beneficiaries residing in SEER areas. Great care was taken to include HCC cases with the greatest specificity, including diagnostically confirmed first cancer cases, minimizing inclusion of metastatic liver cancer. Using claims from Medicare, a federal health insurance program servicing 97% of the ≥65-year-old population (206), important HCC risk factors were included in the statistical analysis to address potential confounding. We carried back the earliest available Medicare ZIP Code to 1974 and were able to craft a comprehensive and historical pesticide exposure metric for all cases and controls. This addresses the documented latency period of some HCC risk factors, where 20 years can pass between initial exposure and clinical diagnosis of disease (3). Our study results are consistent with previous literature linking organochlorines with HCC. Although PURs began reporting pesticide use in 1974 and DDT was banned in 1972, statistically significant positive associations were observed in our study between HCC and organochlorines, the pesticide chemical class to which DDT belongs. Pesticides within a given chemical class have similar chemical structures and biological mechanisms of action (23), therefore, other organochlorines may pose similar risks for liver-related outcomes.

4.5.2 Limitations

Several limitations provide opportunities for future research. Our study was not able to take into account other routes of pesticide exposure such as individual-level occupation, diet, and residential pesticide use. However, the route of exposure addressed in this study, residential proximity, has been used as a surrogate for pesticide exposure occurring through a variety of routes, including dermal contact in crop fields (89). Our study was likely underpowered to detect

a difference between pesticides and HCC particularly in the rural setting, manifest in the wide confidence intervals when examining the continuous rate and applied pounds metrics. In sensitivity analyses, categorical pesticide exposure calculated using cutoffs from the rural control sample demonstrated significant increased risks. Usage of Medicare claims is associated with inherent limitations as only conditions diagnosed and recorded by a healthcare provider are captured. Claims lack sensitivity with some conditions, including HCV, which are underdiagnosed in the elderly population (206). In this sample, a high proportion of all sampled controls (10.6%) were tested for HBV and/or HCV at least one year before selection according to Healthcare Common Procedure Coding System (HCPCS) codes (Appendix C). The generalizability of our findings are somewhat limited given the median age of HCC diagnosis in the U.S. is 63 years (12) and sampled cases were at least 66 years old. Yet, a comparable proportion of our case series (76.5%) and of all HCC cases diagnosed between 2000 and 2009 in California aged 64 years and younger (78.6%) resided in highly urbanized areas at diagnosis (202). Thus, our results potentially extend to this relatively younger population of cases when considering the entire state of California irrespective of rural/urban residence. The impact on rural results is less clear, although it is possible that younger cases in rural areas who were alive before the 1970s when organochlorines were still widely used were exposed. Furthermore, these individuals, although alive for a shorter period of time compared to older cases, might have been exposed during a critical period when pesticide exposure poses significant risks to biological development (231, 232). We excluded individuals enrolled in Part C due to their lack of Medicare claims. Between 2000 and 2009, approximately 31.0 to 37.9% of Medicare beneficiaries in California enrolled in Part C (208). Approximately 87.9% of cases excluded from our study due to HMOs resided in urban areas. When considering all California Medicare

beneficiaries enrolled in Part C sometime between 2000 and 2009, between 64.9 and 94.4% resided in urban areas (208). Assuming the association between urban residence and Part C enrollment is nondifferential between cases and controls, excluding Part C enrollees may have biased results towards the null.

The major limitation of our study was exposure misclassification. By carrying back the earliest available Medicare ZIP Code, we assumed individuals were residentially stable as early as their 30s and into midlife. This assumption may be conceivable given 89.6% of cases and 90.7% of controls had one to two ZIP Codes (using available Medicare ZIP Codes from 1991 until year before diagnosis/selection). Furthermore, only 3% of cases and 3% of controls moved from rural to urban areas or vice versa, potentially demonstrating that the exposure assigned to these individuals using their carried-back ZIP Code is representative of what they would have experienced anyway by virtue of typically residing in only urban, or only rural, areas during their lifetime. Medicare ZIP Codes represent the last available billing ZIP Code in that year, which may not reflect residence. ZIP Codes are coarse spatial resolution spatial variables, ranging from 0.005 to 2,028.7 mi^2 in California. ZIP Codes in rural areas, in which the population at risk for pesticide exposure reside, are typically larger than urban ZIP Codes. ZIP Codes are frequently modified and represent linear features created for the purposes of mail delivery (178). ZIP Code-level pesticide exposure metrics are ecologic, taking into account pesticide exposure occurring within an entire ZIP Code. As a result, exposure could have been over- or underestimated for any given study subject. A more meaningful GIS pesticide exposure metric would use finer spatial resolution data, such as geocoded residential locations, taking into account agricultural pesticide use occurring within a 500 m (radius) buffer (previously associated with impacting human health and used in epidemiologic studies) (24).

4.6 CONCLUSIONS

Among rural-dwelling California Medicare beneficiaries, previous ZIP Code-level organochlorine pesticide exposure was significantly associated with an increased risk of developing HCC after taking into account liver disease and diabetes. This is the first epidemiologic study using GIS to study pesticide exposure and HCC. Our study highlights another potential risk factor for HCC in the U.S. population that should be further examined. We used Medicare ZIP Codes to estimate pesticide exposure, which is a coarse spatial resolution variable subject to changes over time. Future research should explore the use of finer spatial resolution data, such as geocoded residences, in addition to collecting information regarding other routes of pesticide exposure to further elucidate the association between pesticides and HCC.

4.7 TABLES

Table 11. Included vs. Excluded Hepatocellular Carcinoma Cases From SEER-Medicare in California, 2000-2009

Characteristic	Included cases[a] (*n* = 3,034) *n* (%)	Excluded cases[a,b] (*n* = 7,374) *n* (%)	*P* Value[c]
Age in years: mean (SD)	75.1 (6.3)	66.1 (11.0)	<0.0001
Year			0.1754
2000	236 (7.8)	544 (7.4)	
2001	235 (7.8)	616 (8.4)	
2002	278 (9.2)	598 (8.1)	
2003	260 (8.6)	665 (9.0)	
2004	316 (10.4)	654 (8.9)	
2005	309 (10.2)	813 (11.0)	
2006	326 (10.7)	829 (11.2)	
2007	344 (11.3)	876 (11.9)	
2008	349 (11.5)	857 (11.6)	
2009	381 (12.6)	922 (12.5)	
Sex			<0.0001
Male	1,915 (63.1)	5,407 (73.3)	
Female	1,119 (36.9)	1,967 (26.7)	
Race			<0.0001
White	1,548 (51.0)	4,178 (56.7)	
Black	152 (5.0)	529 (7.2)	
Other	265 (8.7)	789 (10.7)	
Asian	793 (26.1)	1,185 (16.1)	
Hispanic	256 (8.4)	642 (8.7)	
Native American	20 (0.7)	51 (0.7)	
California residence			<0.0001
1-5 years	838 (27.6)	4,164 (56.5)	
6-10 years	1,077 (35.5)	1,611 (21.9)	
≥11 years	1,119 (36.9)	1,599 (21.7)	
State buy-in	1,586 (52.3)	2,149 (29.1)	<0.0001
Urban residence at diagnosis[d]	2,322 (76.5)	5,971 (81.0)	<0.0001

Abbreviations: SD, standard deviation.

[a] The source population of SEER-Medicare cases included all individuals diagnosed with hepatocellular carcinoma (ICD-O-3 C22.0 and 8170-8175) as a first cancer between 2000 and 2009, reported to a California registry, not of unknown race, and in the Medicare Enrollment DataBase (requirement to be included in SEER-Medicare data linkage).

[b] Cases were excluded from the study due to lack of diagnostic confirmation, being reporting on death certificate or autopsy only, <66 years old at diagnosis, not having ≥13 months of continuous Parts A and B, non-HMO enrollment before diagnosis, or no available California Medicare ZIP Codes by diagnosis.

[c] Two-sided *P* values from two-sample t-tests for continuous variables and from chi-square tests for categorical variables are presented.

[d] Urban residence was defined using 2003 Rural/Urban Continuum Codes (RUCC) corresponding to metropolitan vs. nonmetropolitan counties at diagnosis. Urban was defined as residence in counties with metropolitan areas of ≥1 million population (RUCC code 1).

Table 12. SEER-Medicare Hepatocellular Carcinoma Cases From Source Population, Included Cases, and From SEER (65 Years and Older) in California, 2000-2009

Characteristic	Source population of cases[a] (*n* = 10,408) *n* (%)	Included cases in study (*n* = 3,034) *n* (%)	SEER cases[b] (*n* = 17,291) *n* (%)	SEER cases ≥65 years old[c] (*n* = 7,185) *n* (%)
Age in years: mean (SD)	68.7 (10.7)	75.1 (6.3)	62.2 (14.6)	73.9 (6.0)
Year				
2000	780 (7.5)	236 (7.8)	1,252 (7.2)	585 (8.1)
2001	851 (8.2)	235 (7.8)	1,359 (7.9)	626 (8.7)
2002	876 (8.4)	278 (9.2)	1,396 (8.1)	655 (9.1)
2003	925 (8.9)	260 (8.6)	1,492 (8.6)	639 (8.9)
2004	970 (9.3)	316 (10.4)	1,657 (9.6)	692 (9.6)
2005	1,122 (10.8)	309 (10.2)	1,844 (10.7)	741 (10.3)
2006	1,155 (11.1)	326 (10.7)	1,934 (11.2)	756 (10.5)
2007	1,220 (11.7)	344 (11.3)	2,018 (11.7)	795 (11.1)
2008	1,206 (11.6)	349 (11.5)	2,035 (11.8)	812 (11.3)
2009	1,303 (12.5)	381 (12.6)	2,304 (13.3)	884 (12.3)
Sex				
Male	7,322 (70.4)	1,915 (63.1)	13,111 (75.8)	4,702 (65.4)
Female	3,086 (29.7)	1,119 (36.9)	4,180 (24.2)	2,483 (34.6)
Race[d]				
White	5,726 (55.0)	1,548 (51.0)	11,036 (63.8)	4,436 (61.7)
Black	681 (6.5)	152 (5.0)	1,373 (7.9)	412 (5.7)
Other	1,054 (10.1)	265 (8.7)	--	--
Asian	1,978 (19.0)	793 (26.1)	4,718 (27.3)	2,290 (31.9)
Hispanic	898 (8.6)	256 (8.4)	--	--
Native American	71 (0.7)	20 (0.7)	164 (0.9)	47 (0.7)
California residence[e]				
1-5 years	838 (27.6)	838 (27.6)	--	--
6-10 years	1,077 (35.5)	1,077 (35.5)	--	--
≥11 years	1,119 (36.9)	1,119 (36.9)	--	--
Urban residence at diagnosis	8,293 (79.7)	2,322 (76.5)	13,888 (80.3)	5,947 (82.8)

Abbreviations: SD, standard deviation.

[a] Includes all SEER-Medicare hepatocellular carcinoma cases included or excluded from the study.

[b] SEER cases were derived from SEER*Stat (202): ICD-O-3 C22.0 topography and 8170-8175 histology codes, sequence number 00 or 01, diagnosed between 2000 and 2009, reported to a California registry, diagnostically confirmed, not reported via autopsy or death certificate only, not of unknown race, and with no age restriction.

[c] SEER cases aged ≥65 years at diagnosis. Ages >85 years are not individually reported in SEER (*n* = 512 cases were categorized as 85+ years).

[d] SEER recoded race is comprised of four categories where Hispanics are not mutually exclusive of any of these groups. Other race is not a category.

[e] California residence was calculated using Medicare ZIP Codes, and was thus not available when using SEER.

Table 13. Population Characteristics of Hepatocellular Carcinoma Cases and Frequency-Matched Controls From SEER-Medicare in California, 2000-2009

Characteristic	Cases[a] (*n* = 3,034) *n* (%)	Controls[a] (*n* = 14,991) *n* (%)	*P* Value[b]
Age in years: mean (SD)	75.1 (6.3)	75.1 (6.3)	
Year			
2000	236 (7.8)	1,168 (7.8)	
2001	235 (7.8)	1,170 (7.8)	
2002	278 (9.2)	1,369 (9.1)	
2003	260 (8.6)	1,289 (8.6)	
2004	316 (10.4)	1,548 (10.3)	
2005	309 (10.2)	1,516 (10.1)	
2006	326 (10.7)	1,617 (10.8)	
2007	344 (11.3)	1,703 (11.4)	
2008	349 (11.5)	1,735 (11.6)	
2009	381 (12.6)	1,876 (12.5)	
Sex			
Male	1,915 (63.1)	9,469 (63.2)	
Female	1,119 (36.9)	5,522 (36.8)	
Race			
White	1,548 (51.0)	7,739 (51.6)	
Black	152 (5.0)	743 (5.0)	
Other	265 (8.7)	1,266 (8.5)	
Asian	793 (26.1)	3,924 (26.2)	
Hispanic	256 (8.4)	1,252 (8.4)	
Native American	20 (0.7)	67 (0.5)	
California residence			
1-5 years	838 (27.6)	4,145 (27.7)	
6-10 years	1,077 (35.5)	5,307 (35.4)	
≥11 years	1,119 (36.9)	5,539 (36.9)	
Duration of Medicare coverage (years)[c]			
1.1-4.3	747 (24.6)	3,761 (25.1)	
4.4-8.1	765 (25.2)	3,813 (25.4)	
8.2-11.5	775 (25.5)	3,524 (23.5)	
11.6-19	747 (24.6)	3,893 (26.0)	
Duration of Medicare coverage within 6 years of diagnosis/selection (years)[d]			
1.1-4.1	755 (24.9)	3,736 (24.9)	
4.2-6.1	2,279 (75.1)	11,255 (75.1)	
HCV	672 (22.2)	71 (0.5)	<0.0001
HBV	177 (5.8)	20 (0.1)	<0.0001
Unspecified hepatitis	379 (12.5)	75 (0.5)	<0.0001
Alcoholic liver disease	212 (7.0)	44 (0.3)	<0.0001
Non-specific cirrhosis	516 (17.0)	41 (0.3)	<0.0001
Liver disease[e]			<0.0001
None	1,861 (61.3)	14,783 (98.6)	
Hepatitis only	592 (19.5)	133 (0.9)	
Cirrhosis	581 (19.2)	75 (0.5)	
Diabetes	1,300 (42.9)	2,054 (13.7)	<0.0001
Obesity	125 (4.1)	365 (2.4)	<0.0001

Table 13 continued

Characteristic	Cases[a] (*n* = 3,034) *n* (%)	Controls[a] (*n* = 14,991) *n* (%)	*P* Value[b]
Rare genetic disorders	43 (1.4)	17 (0.1)	<0.0001
HIV[f]	<11	<11	0.0072
Smoking	262 (8.6)	795 (5.3)	<0.0001
State buy-in	1,586 (52.3)	6,070 (40.5)	<0.0001
ZIP Code percentage employed in agriculture: median (IQR)	0.2 (0.1, 0.7)	0.2 (0.1, 1.0)	0.0002
Urban residence at diagnosis/selection[g]	2,322 (76.5)	10,801 (72.1)	<0.0001

Abbreviations: HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; IQR, interquartile range; SD, standard deviation.

[a] Twenty-one controls subsequently became cases; 11,613 controls served once, 1,347 served twice, 195 served three times, 21 served four times, and 3 served five times.

[b] No statistical tests are presented for matching factors age, year, sex, race, and California residence. For all other variables, two-sided *P* values from univariable conditional logistic regression models (robust variance estimation) accounting for the matching factors are presented.

[c] Years of non-continuous enrollment in Parts A and B, non-HMO coverage. Coverage was categorized using quartiles among cases.

[d] Years of non-continuous enrollment in Parts A and B, non-HMO coverage within 6 years of diagnosis/selection. Coverage was categorized using the 25th percentile among cases (4.1 years).

[e] Liver disease was used in statistical modeling, representing none (no hepatitis, alcoholic liver disease, and non-specific cirrhosis), hepatitis only (hepatitis without alcoholic liver disease and cirrhosis), and cirrhosis (alcoholic liver disease or non-specific cirrhosis with or without hepatitis).

[f] In accordance with the SEER-Medicare data use agreement, cell sizes <11 are suppressed.

[g] Urban residence was defined using 2003 Rural/Urban Continuum Codes (RUCC) corresponding to metropolitan vs. nonmetropolitan counties at diagnosis. Urban was defined as residence in counties with metropolitan areas of ≥1 million population (RUCC code 1).

Table 14. Pesticide Exposure and Hepatocellular Carcinoma Using Pesticide Use Reports and SEER-Medicare in California, 2000-2009

	Pesticide exposure category (lb/ac)[a]	Cases (*n* = 3,034) *n* (%)	Controls (*n* = 14,991) *n* (%)	OR (95% CI)[b]	*P* Value[b]
ZIP Code pesticide exposure, annual pesticide application rate (lb/ac)					
All classes[c]					0.0578
Low exposure	≤0.001	1,054 (34.7)	4,949 (33.0)	--	
Moderate exposure	0.001-0.02	1,007 (33.2)	4,944 (33.0)	0.96 (0.87, 1.05)	
High exposure	≥0.02	973 (32.1)	5,098 (34.0)	0.89 (0.81, 0.98)	
Organophosphates					0.0366
Low exposure	≤0.0003	1,059 (34.9)	4,946 (33.0)	--	
Moderate exposure	0.0003-0.01	999 (32.9)	4,949 (33.0)	0.94 (0.86, 1.04)	
High exposure	≥0.01	976 (32.2)	5,096 (34.0)	0.89 (0.81, 0.98)	
Organochlorines					0.0499
Low exposure	≤0.00001	1,057 (34.8)	4,949 (33.0)	--	
Moderate exposure	0.00001-0.001	977 (32.2)	4,946 (33.0)	0.93 (0.84, 1.02)	
High exposure	≥0.001	1,000 (33.0)	5,096 (34.0)	0.92 (0.83, 1.01)	
Carbamates					0.1685
Low exposure	≤0.0001	1,039 (34.3)	4,945 (33.0)	--	
Moderate exposure	0.0001-0.004	1,018 (33.6)	4,950 (33.0)	0.98 (0.89, 1.08)	
High exposure	≥0.004	977 (32.2)	5,096 (34.0)	0.91 (0.83, 1.00)	
ZIP Code pesticide exposure, average annual applied pounds					
All classes[c]					0.0078
Low exposure	≤3.74	1,074 (35.4)	4,947 (33.0)	--	
Moderate exposure	3.74-158.48	1,008 (33.2)	4,948 (33.0)	0.94 (0.86, 1.03)	
High exposure	≥158.48	952 (31.4)	5,096 (34.0)	0.85 (0.78, 0.94)	
Organophosphates					0.0101
Low exposure	≤2.19	1,071 (35.3)	4,947 (33.0)	--	
Moderate exposure	2.19-87.87	1,025 (33.8)	4,946 (33.0)	0.96 (0.87, 1.05)	
High exposure	≥87.87	938 (30.9)	5,098 (34.0)	0.84 (0.77, 0.93)	
Organochlorines					0.0096
Low exposure	≤0.04	1,072 (35.3)	4,946 (33.0)	--	
Moderate exposure	0.04-13.27	1,005 (33.1)	4,949 (33.0)	0.94 (0.85, 1.03)	
High exposure	≥13.27	957 (31.5)	5,096 (34.0)	0.86 (0.78, 0.95)	
Carbamates					0.0789
Low exposure	≤0.46	1,049 (34.6)	4,947 (33.0)	--	
Moderate exposure	0.46-42.35	1,026 (33.8)	4,948 (33.0)	0.98 (0.89, 1.07)	
High exposure	≥42.35	959 (31.6)	5,096 (34.0)	0.88 (0.80, 0.97)	

Abbreviations: ac, acre; CI, confidence interval; lb, pound; OR, odds ratio.

[a] Pesticide application rates or applied pounds were categorized using tertiles among controls.

[b] Odds ratios, 95% confidence intervals, and two-sided *P* values from univariable conditional logistic regression models (robust variance estimation) accounting for the matching factors are presented for cases vs. controls.

[c] Refers to all combined chemical classes: organophosphates, organochlorines, and carbamates.

Table 15. Factors Associated with Pesticide Exposure Among Controls Using SEER-Medicare in California, 2000-2009

Characteristic	Low exposure[a] (*n* = 4,949) ≤0.001 lb/ac *n* (%)	Moderate exposure[a] (*n* = 4,944) 0.001-0.02 lb/ac *n* (%)	High exposure[a] (*n* = 5,098) ≥0.02 lb/ac *n* (%)	*P* Value[b]
Age in years: mean (SD)	75.5 (6.3)	75.1 (6.3)	74.8 (6.2)	<0.0001
Year				0.2533
2000	401 (8.1)	372 (7.5)	395 (7.8)	
2001	381 (7.7)	384 (7.8)	405 (7.9)	
2002	446 (9.0)	478 (9.7)	445 (8.7)	
2003	434 (8.8)	413 (8.4)	442 (8.7)	
2004	482 (9.7)	533 (10.8)	533 (10.5)	
2005	480 (9.7)	506 (10.2)	530 (10.4)	
2006	520 (10.5)	511 (10.3)	586 (11.5)	
2007	581 (11.7)	584 (11.8)	538 (10.6)	
2008	611 (12.4)	559 (11.3)	565 (11.1)	
2009	613 (12.4)	604 (12.2)	659 (12.9)	
Sex				0.3161
Male	3,099 (62.6)	3,108 (62.9)	3,262 (64.0)	
Female	1,850 (37.4)	1,836 (37.1)	1,836 (36.0)	
Race				<0.0001
White	2,283 (46.1)	2,457 (49.7)	2,999 (58.8)	
Black	283 (5.7)	265 (5.4)	195 (3.8)	
Other	467 (9.4)	409 (8.3)	390 (7.7)	
Asian	1,630 (32.9)	1,433 (29.0)	861 (16.9)	
Hispanic	271 (5.5)	362 (7.3)	619 (12.1)	
Native American	15 (0.3)	18 (0.4)	34 (0.7)	
California residence				0.0065
1-5 years	1,406 (28.4)	1,376 (27.8)	1,363 (26.7)	
6-10 years	1,804 (36.5)	1,744 (35.3)	1,759 (34.5)	
≥11 years	1,739 (35.1)	1,824 (36.9)	1,976 (38.8)	
Duration of Medicare coverage within 6 years of diagnosis/selection (years)[c]				0.3116
1.1-4.1	1,203 (24.3)	1,267 (25.6)	1,266 (24.8)	
4.2-6.1	3,746 (75.7)	3,677 (74.4)	3,832 (75.2)	
HCV	28 (0.6)	31 (0.6)	12 (0.2)	0.0087
HBV[d]	13 (0.3)	<11	<11	0.0037
Unspecified hepatitis	30 (0.6)	26 (0.5)	19 (0.4)	0.2408
Alcoholic liver disease[d]	<11	<11	25 (0.5)	0.0051
Non-specific cirrhosis[d]	18 (0.4)	<11	16 (0.3)	0.0848
Liver disease				0.0006
None	4,871 (98.4)	4,873 (98.6)	5,039 (98.8)	
Hepatitis only	53 (1.1)	55 (1.1)	25 (0.5)	
Cirrhosis	25 (0.5)	16 (0.3)	34 (0.7)	
Diabetes	677 (13.7)	697 (14.1)	680 (13.3)	0.5415
Obesity	114 (2.3)	119 (2.4)	132 (2.6)	0.6417
Rare genetic disorders[d]	<11	<11	<11	0.8229
HIV[d,e]	<11	<11	<11	0.3643
Smoking	260 (5.3)	249 (5.0)	286 (5.6)	0.4315
State buy-in	2,257 (45.6)	2,028 (41.0)	1,785 (35.0)	<0.0001

Table 15 continued

Characteristic	Low exposure[a] (*n* = 4,949) ≤0.001 lb/ac *n* (%)	Moderate exposure[a] (*n* = 4,944) 0.001-0.02 lb/ac *n* (%)	High exposure[a] (*n* = 5,098) ≥0.02 lb/ac *n* (%)	*P* Value[b]
ZIP Code percentage employed in agriculture: median (IQR)	0.1 (0.1, 0.2)	0.2 (0.1, 0.4)	1.2 (0.3, 4.4)	<0.0001
Urban residence at diagnosis/selection[f]	4,365 (88.2)	4,105 (83.0)	2,331 (45.7)	<0.0001

Abbreviations: ac, acre; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; IQR, interquartile range; lb, pound; SD, standard deviation.

[a] Pesticide application rates from all chemical classes (organophosphates, organochlorines, and carbamates) were categorized using tertiles among controls.

[b] Two-sided *P* values from one-way analysis of variance for age, chi-square tests for categorical variables, and the Kruskal-Wallis test for occupation are presented.

[c] Years of non-continuous enrollment in Parts A and B, non-HMO coverage within 6 years of diagnosis/selection. Coverage was categorized using the 25th percentile among cases (4.1 years).

[d] In accordance with the SEER-Medicare data use agreement, cell sizes <11 are suppressed.

[e] Two-sided *P* value presented from Fisher's exact test.

[f] Urban residence was defined using 2003 Rural/Urban Continuum Codes (RUCC) corresponding to metropolitan vs. nonmetropolitan counties at diagnosis. Urban was defined as residence in counties with metropolitan areas of ≥1 million population (RUCC code 1).

Table 16. Random-Intercept Logistic Regression Examining Clustering Within ZIP Codes Using SEER-Medicare Cases and Controls in California, 2000-2009

	Random intercept: ZIP Code at diagnosis/selection Adjusted OR (95% CI)[a]	*P* Value[a]	Random intercept: ZIP Code occurring most frequently Adjusted OR (95% CI)[a]	*P* Value[a]
ICC	0.03		0.03	
ZIP Code pesticide exposure, annual pesticide application rate (lb/ac), all classes	1.00 (1.00, 1.00)	0.1710	1.00 (1.00, 1.00)	0.1220
Age in years	1.00 (0.99, 1.01)	0.8130	1.00 (0.99, 1.01)	0.7930
Year		>0.999		>0.999
2000	--		--	
2001	0.98 (0.81, 1.20)		0.99 (0.81, 1.20)	
2002	1.00 (0.82, 1.22)		1.00 (0.82, 1.22)	
2003	1.00 (0.82, 1.21)		1.00 (0.82, 1.21)	
2004	1.01 (0.84, 1.21)		1.01 (0.84, 1.21)	
2005	1.01 (0.83, 1.21)		1.01 (0.84, 1.21)	
2006	1.00 (0.83, 1.20)		1.00 (0.83, 1.21)	
2007	1.00 (0.83, 1.21)		1.00 (0.83, 1.20)	
2008	0.99 (0.82, 1.20)		0.99 (0.82, 1.19)	
2009	1.00 (0.84, 1.20)		1.00 (0.83, 1.20)	
Sex		0.8510		0.8300
Male	--		--	
Female	0.99 (0.91, 1.08)		0.99 (0.91, 1.08)	
Race		0.5861		0.5652
White	--		--	
Black	1.00 (0.82, 1.21)		1.00 (0.83, 1.21)	
Other	1.02 (0.88, 1.19)		1.02 (0.87, 1.19)	
Asian	0.95 (0.85, 1.06)		0.95 (0.85, 1.06)	
Hispanic	1.00 (0.86, 1.16)		1.01 (0.86, 1.17)	
Native American	1.49 (0.88, 2.55)		1.51 (0.89, 2.56)	
California residence		0.9718		0.9694
1-5 years	--		--	
6-10 years	1.00 (0.89, 1.12)		1.00 (0.89, 1.12)	
≥11 years	0.98 (0.85, 1.14)		0.99 (0.85, 1.14)	

Abbreviations: ac, acre; CI, confidence interval; ICC, intraclass correlation coefficient; lb, pound; OR, odds ratio.
[a] Adjusted odds ratios, 95% confidence intervals, and two-sided *P* values from random-intercept logistic regression models using robust variance estimation, defining the random intercept as the ZIP Code, and adjusting for all other variables are presented.

Table 17. Adjusted Odds Ratios for Hepatocellular Carcinoma: Pesticide Exposure in California Using SEER-Medicare, 2000-2009

Predictor	Adjusted OR[a]	95% CI[a]		*P* value[a]
ZIP Code pesticide exposure, annual pesticide application rate (lb/ac)				
All classes[b,c]	0.92	0.80	1.07	0.3009
Organophosphates	0.85	0.65	1.11	0.2322
Organochlorines	0.82	0.29	2.29	0.6981
Carbamates	0.83	0.56	1.22	0.3411
ZIP Code pesticide exposure, average annual applied pounds (per 10,000 lb)				
All classes[b]	0.98	0.96	1.01	0.1710
Organophosphates	0.98	0.94	1.02	0.3107
Organochlorines	0.95	0.80	1.12	0.5230
Carbamates	0.94	0.88	1.00	0.0663

Abbreviations: ac, acre; CI, confidence interval; lb, pound; OR, odds ratio.

[a] Odds ratios, 95% confidence intervals, and two-sided *P* values were estimated using multivariable conditional logistic regression (robust variance estimation) adjusting for liver disease, diabetes, rare genetic disorders, and state buy-in in addition to taking into account the matching factors.

[b] Refers to all combined chemical classes: organophosphates, organochlorines, and carbamates.

[c] The following are adjusted odds ratios for the other predictors in the multivariable conditional logistic regression model: hepatitis only vs. no liver disease (OR 35.99; 95% confidence interval 23.96, 54.05), cirrhosis vs. no liver disease (OR 62.43; 95% confidence interval 27.41, 142.23), diabetes (OR 4.44, 95% confidence interval 3.81, 5.16), rare genetic disorders (OR 6.17, 95% confidence 1.56, 24.45), state buy-in (OR 1.39, 95% confidence interval 1.21, 1.61).

Table 18. Rural Residents: Population Characteristics of Hepatocellular Carcinoma Cases and Frequency-Matched Controls From SEER-Medicare in California, 2000-2009

Characteristic	Cases[a] (*n* = 306) *n* (%)	Controls[a] (*n* = 1,758) *n* (%)	*P* Value[b]
Age in years: mean (SD)	74.9 (6.0)	74.6 (6.2)	
Year			
2000	24 (7.8)	155 (8.8)	
2001	27 (8.8)	142 (8.1)	
2002	32 (10.5)	157 (8.9)	
2003	29 (9.5)	160 (9.1)	
2004	31 (10.1)	178 (10.1)	
2005	40 (13.1)	190 (10.8)	
2006	28 (9.2)	196 (11.2)	
2007	25 (8.2)	169 (9.6)	
2008	34 (11.1)	189 (10.8)	
2009	36 (11.8)	222 (12.6)	
Sex			
Male	190 (62.1)	1,168 (66.4)	
Female	116 (37.9)	590 (33.6)	
Race[c]			
White	194 (63.4)	1109 (63.1)	
Black	<11	58 (3.3)	
Other	28 (9.2)	125 (7.1)	
Asian	23 (7.5)	151 (8.6)	
Hispanic	48 (15.7)	301 (17.1)	
Native American	<11	14 (0.8)	
California residence			
1-5 years	68 (22.2)	440 (25.0)	
6-10 years	105 (34.3)	589 (33.5)	
≥11 years	133 (43.5)	729 (41.5)	
Duration of Medicare coverage (years)[d]			
1.1-4.3	68 (22.2)	386 (22.0)	
4.4-8.1	63 (20.6)	414 (23.6)	
8.2-11.5	85 (27.8)	413 (23.5)	
11.6-19	90 (29.4)	545 (31.0)	
Duration of Medicare coverage within 6 years of diagnosis/selection (years)[e]			
1.1-4.1	66 (21.6)	408 (23.2)	
4.2-6.1	240 (78.4)	1,350 (76.8)	
Liver disease[f]	98 (32.0)	21 (1.2)	<0.0001
Diabetes	146 (47.7)	242 (13.8)	<0.0001
Obesity	21 (6.9)	57 (3.2)	0.0303
Rare genetic disorders[c]	<11	<11	<0.0001
HIV[c]	<11	<11	<0.0001
Smoking	29 (9.5)	98 (5.6)	0.0077
State buy-in	143 (46.7)	601 (34.2)	<0.0001
ZIP Code percentage employed in agriculture: median (IQR)	4.7 (2.6, 7.7)	5.0 (2.6, 7.5)	0.5097

Table 18 continued

Characteristic	Cases[a] (*n* = 306) *n* (%)	Controls[a] (*n* = 1,758) *n* (%)	*P* Value[b]
Urban residence at diagnosis/selection[g]	48 (15.7)	207 (11.8)	0.1629

Abbreviations: HCV, hepatitis C virus; HBV, hepatitis B virus; HIV, human immunodeficiency virus; IQR, interquartile range; SD, standard deviation.
[a] Four controls subsequently became cases; 1,583 controls served once, 157 served twice, and 18 served three times.
[b] No statistical tests are presented for matching factors age, year, sex, race, and California residence. For all other variables, two-sided *P* values from univariable conditional logistic regression models (robust variance estimation) accounting for the matching factors are presented.
[c] In accordance with the SEER-Medicare data use agreement, cell sizes <11 are suppressed.
[d] Years of non-continuous enrollment in Parts A and B, non-HMO coverage. Coverage was categorized using quartiles among cases from the full study sample.
[e] Years of non-continuous enrollment in Parts A and B, non-HMO coverage within 6 years of diagnosis/selection. Coverage was categorized using the 25th percentile among cases from the full study sample (4.1 years).
[f] Liver disease defined as yes/no hepatitis, alcoholic liver disease, or non-specific cirrhosis.
[g] Urban residence was defined using 2003 Rural/Urban Continuum Codes (RUCC) corresponding to metropolitan vs. nonmetropolitan counties at diagnosis. Urban was defined as residence in counties with metropolitan areas of ≥1 million population (RUCC code 1).

Table 19. Rural Residents: Pesticide Exposure and Hepatocellular Carcinoma Using Pesticide Use Reports and SEER-Medicare in California, 2000-2009

	Pesticide exposure category (lb/ac)[a]	Cases (*n* = 306) *n* (%)	Controls (*n* = 1,758) *n* (%)	OR (95% CI)[b]	*P* Value[b]
ZIP Code pesticide exposure, annual pesticide application rate (lb/ac):					
All classes[c]					0.7995
Low exposure	≤0.48	104 (34.0)	579 (32.9)	--	
Moderate exposure	0.48-0.92	94 (30.7)	582 (33.1)	0.92 (0.66, 1.29)	
High exposure	≥0.92	108 (35.3)	597 (34.0)	1.17 (0.83, 1.65)	
Organophosphates					0.0209
Low exposure	≤0.26	89 (29.1)	582 (33.1)	--	
Moderate exposure	0.26-0.50	110 (36.0)	574 (32.7)	1.37 (0.98, 1.92)	
High exposure	≥0.50	107 (35.0)	602 (34.2)	1.48 (1.05, 2.09)	
Organochlorines					0.0015
Low exposure	≤0.04	89 (29.1)	581 (33.1)	--	
Moderate exposure	0.04-0.10	98 (32.0)	580 (33.0)	1.44 (1.02, 2.04)	
High exposure	≥0.10	119 (38.9)	597 (34.0)	1.93 (1.35, 2.77)	
Carbamates					0.0160
Low exposure	≤0.13	81 (26.5)	579 (32.9)	--	
Moderate exposure	0.13-0.30	115 (37.6)	584 (33.2)	1.51 (1.06, 2.15)	
High exposure	≥0.30	110 (36.0)	595 (33.9)	1.44 (1.01, 2.06)	
ZIP Code pesticide exposure, average annual applied pounds					
All classes[c]					0.8192
Low exposure	≤22,301.80	103 (33.7)	576 (32.8)	--	
Moderate exposure	22,301.80-61,174.03	103 (33.7)	583 (33.2)	0.99 (0.72, 1.36)	
High exposure	≥61,174.03	100 (32.7)	599 (34.1)	1.08 (0.76, 1.54)	
Organophosphates					0.8147
Low exposure	≤12,402.68	102 (33.3)	581 (33.1)	--	
Moderate exposure	12,402.68-28,453.40	94 (30.7)	580 (33.0)	0.87 (0.62, 1.24)	
High exposure	≥28,453.40	110 (36.0)	597 (34.0)	1.23 (0.88, 1.72)	
Organochlorines					0.2031
Low exposure	≤1,795.52	103 (33.7)	581 (33.1)	--	
Moderate exposure	1,795.52-5,273.83	97 (31.7)	580 (33.0)	1.06 (0.76, 1.48)	
High exposure	≥5,273.83	106 (34.6)	597 (34.0)	1.38 (0.98, 1.94)	
Carbamates					0.5831
Low exposure	≤5,951.53	98 (32.0)	577 (32.8)	--	
Moderate exposure	5,951.53-18,425.59	108 (35.3)	581 (33.1)	1.14 (0.83, 1.56)	
High exposure	≥18,425.59	100 (32.7)	600 (34.1)	1.04 (0.73, 1.47)	

Abbreviations: ac, acre; lb, pound; OR, odds ratio.

[a] Pesticide application rates or applied pounds were categorized using tertiles among rural controls.

[b] Odds ratio, 95% confidence intervals, and two-sided *P* values from univariable conditional logistic regression models (robust variance estimation) accounting for the matching factors are presented for cases vs. controls.

[c] Refers to all combined chemical classes: organophosphates, organochlorines, and carbamates.

Table 20. Rural Residents: Factors Associated with Organochlorine Pesticide Exposure Among Controls Using SEER-Medicare in California, 2000-2009

Characteristic	Low exposure[a] (*n* = 581) ≤0.04 lb/ac *n* (%)	Moderate exposure[a] (*n* = 580) 0.04-0.10 lb/ac *n* (%)	High exposure[a] (*n* = 597) ≥0.10 lb/ac *n* (%)	*P* Value[b]
Age in years: mean (SD)	74.2 (6.1)	74.6 (6.1)	74.8 (6.2)	0.2070
Year				0.1194
2000	47 (8.1)	45 (7.8)	63 (10.6)	
2001	38 (6.5)	43 (7.4)	61 (10.2)	
2002	42 (7.2)	56 (9.7)	59 (9.9)	
2003	58 (10.0)	51 (8.8)	51 (8.5)	
2004	49 (8.4)	63 (10.9)	66 (11.1)	
2005	78 (13.4)	55 (9.5)	57 (9.6)	
2006	66 (11.4)	68 (11.7)	62 (10.4)	
2007	55 (9.5)	56 (9.7)	58 (9.7)	
2008	67 (11.5)	64 (11.0)	58 (9.7)	
2009	81 (13.9)	79 (13.6)	62 (10.4)	
Sex				0.2399
Male	396 (68.2)	391 (67.4)	381 (63.8)	
Female	185 (31.8)	189 (32.6)	216 (36.2)	
Race[c,d]				<0.0001
White	413 (71.1)	350 (60.3)	346 (58.0)	
Black	12 (2.1)	25 (4.3)	21 (3.5)	
Other	31 (5.3)	50 (8.6)	44 (7.4)	
Asian	38 (6.5)	56 (9.7)	57 (9.6)	
Hispanic	78 (13.4)	97 (16.7)	126 (21.1)	
Native American	<11	<11	<11	
California residence				0.1198
1-5 years	165 (28.4)	136 (23.5)	139 (23.3)	
6-10 years	174 (30.0)	202 (34.8)	213 (35.7)	
≥11 years	242 (41.7)	242 (41.7)	245 (41.0)	
Duration of Medicare coverage within 6 years of diagnosis/selection (years)[e]				0.5137
1.1-4.1	143 (24.6)	135 (23.3)	130 (21.8)	
4.2-6.1	438 (75.4)	445 (76.7)	467 (778.2)	
Liver disease[c]	<11	<11	<11	0.9981
Diabetes	57 (9.8)	87 (15.0)	98 (16.4)	0.0026
Obesity	18 (3.1)	18 (3.1)	21 (3.5)	0.8966
Rare genetic disorders[c,d]	<11	<11	<11	0.2187
HIV[c,d]	<11	<11	<11	0.6604
Smoking	24 (4.1)	39 (6.7)	35 (5.9)	0.1458
State buy-in	169 (29.1)	212 (36.6)	220 (36.9)	0.0066
ZIP Code percentage employed in agriculture: median (IQR)	3.7 (2.3, 5.8)	4.7 (2.7, 8.8)	5.4 (3.0, 9.6)	<0.0001
Urban residence at diagnosis/selection[f]	57 (9.8)	90 (15.5)	60 (10.1)	0.0029

Abbreviations: ac, acre; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; IQR, interquartile range; lb, pound; SD, standard deviation.

Table 20 continued

[a] Pesticide application rates from the organochlorine chemical class were categorized using tertiles among rural controls.
[b] Two-sided *P* values from one-way analysis of variance for age, chi-square tests for categorical variables, and the Kruskal-Wallis test for occupation are presented.
[c] In accordance with the SEER-Medicare data use agreement, cell sizes <11 are suppressed.
[d] Two-sided *P* value presented from Fisher's exact test.
[e] Years of non-continuous enrollment in Parts A and B, non-HMO coverage within 6 years of diagnosis/selection. Coverage was categorized using the 25th percentile among cases from the full study sample (4.1 years).
[f] Urban residence was defined using 2003 Rural/Urban Continuum Codes (RUCC) corresponding to metropolitan vs. nonmetropolitan counties at diagnosis. Urban was defined as residence in counties with metropolitan areas of ≥1 million population (RUCC code 1).

Table 21. Rural Residents: Adjusted Odds Ratios for Hepatocellular Carcinoma, Organochlorine Pesticide Exposure Using SEER-Medicare, 2000-2009

Predictor	Lag[a]	Adjusted OR[b]	95% CI[b]		*P* value[b]
ZIP Code organochlorine, annual pesticide application rate (lb/ac)					
	None[c]	3.32	0.89	12.39	0.0742
	10	2.91	0.98	8.63	0.0538
	15	2.74	1.03	7.30	0.0442
	20	2.43	1.04	5.70	0.0414
ZIP Code organochlorine rate (lb/ac): exposed if ≥50th pctl among rural controls					
≥0.06 lb/ac	None	1.52	1.02	2.28	0.0415
≥0.07 lb/ac	10	1.53	1.01	2.32	0.0459
≥0.08 lb/ac	15	1.85	1.22	2.81	0.0038
≥0.08 lb/ac	20	1.81	1.19	2.75	0.0058
ZIP Code organochlorine, average annual applied pounds (per 10,000 lb)					
	None	1.26	1.00	1.58	0.0501
	10	1.24	1.02	1.50	0.0314
	15	1.24	1.02	1.51	0.0284
	20	1.24	1.03	1.50	0.0242
ZIP Code organochlorine pounds: exposed if ≥75th pctl among rural controls					
≥8,292.13	None	1.61	1.02	2.54	0.0430
≥10,110.64	10	1.71	1.08	2.69	0.0219
≥9,897.86	15	1.54	0.98	2.43	0.0617
≥10,266.64	20	1.67	1.06	2.63	0.0274

Abbreviations: ac, acre; CI, confidence interval; lb, pound; OR, odds ratio; pctl, percentile.

[a] An exposure lag of 10, 15, or 20 years considered pesticide exposure occurring outside of that particular window before diagnosis/selection (e.g., 20-year lag considered exposure between 1974 and 1980 for case diagnosed in 2000).

[b] Odds ratios, 95% confidence intervals, and two-sided *P* values were estimated using multivariable conditional logistic regression (robust variance estimation) adjusting for liver disease (yes/no hepatitis, alcoholic liver disease, or non-specific cirrhosis) and diabetes, in addition to taking into account the matching factors.

[c] The following are adjusted odds ratios for the other predictors in the multivariable conditional logistic regression model: liver disease (OR 32.60; 95% confidence interval 6.42, 165.44), diabetes (OR 8.49; 95% confidence interval 4.93, 14.64).

Table 22. Rural Residents With No Known Risk Factors: Odds Ratios for Hepatocellular Carcinoma, Organochlorine Pesticide Exposure Using SEER-Medicare, 2000-2009

Predictor	Lag	OR[a]	95% CI[a]		*P* value[a]
ZIP Code organochlorine, annual pesticide application rate (lb/ac)					
	None	3.02	0.55	16.51	0.2015
	10	2.38	0.62	9.17	0.2067
	15	2.29	0.71	7.35	0.1650
	20	2.04	0.77	5.40	0.1490
ZIP Code organochlorine rate (lb/ac): exposed if ≥50th pctl among rural controls					
≥0.06 lb/ac	None	1.51	0.93	2.45	0.0970
≥0.07 lb/ac	10	1.35	0.84	2.18	0.2179
≥0.08 lb/ac	15	2.15	1.29	3.58	0.0032
≥0.08 lb/ac	20	1.69	1.02	2.82	0.0432
ZIP Code organochlorine, average annual applied pounds (per 10,000 lb)					
	None	1.15	0.90	1.47	0.2587
	10	1.12	0.91	1.37	0.2826
	15	1.12	0.91	1.37	0.2890
	20	1.14	0.93	1.39	0.2024
ZIP Code organochlorine pounds: exposed if ≥75th pctl among rural controls					
≥8,292.13	None	1.60	0.92	2.79	0.0994
≥10,110.64	10	1.94	1.12	3.34	0.0173
≥9,897.86	15	1.62	0.94	2.79	0.0830
≥10,266.64	20	1.63	0.95	2.79	0.0749

Abbreviations: ac, acre; CI, confidence interval; lb, pound; OR, odds ratio; pctl, percentile.
[a] Odds ratios, 95% confidence intervals, and two-sided *P* values were estimated using univariable conditional logistic regression models (robust variance estimation) taking into account the matching factors and only among those without HCC risk factors (no liver disease, diabetes, obesity, rare genetic disorders, smoking, and HIV) (*n* = 106 cases, *n* = 1,412 controls).

Table 23. Rural Residents Entitled to Medicare Due to Age: Adjusted Odds Ratios for Hepatocellular Carcinoma, Organochlorine Pesticide Exposure Using SEER-Medicare, 2000-2009

Predictor	Lag	Adjusted OR[a]	95% CI[a]		*P* value[a]
ZIP Code organochlorine, annual pesticide application rate (lb/ac)					
	None	4.20	1.11	15.88	0.0342
	10	3.39	1.14	10.03	0.0277
	15	3.10	1.18	8.16	0.0217
	20	2.70	1.16	6.28	0.0217
ZIP Code organochlorine rate (lb/ac): exposed if ≥50th pctl among rural controls					
≥0.06 lb/ac	None	1.86	1.22	2.85	0.0041
≥0.07 lb/ac	10	1.80	1.15	2.82	0.0106
≥0.08 lb/ac	15	2.03	1.31	3.17	0.0017
≥0.08 lb/ac	20	2.08	1.33	3.23	0.0012
ZIP Code organochlorine, average annual applied pounds (per 10,000 lb)					
	None	1.31	1.03	1.67	0.0298
	10	1.27	1.03	1.57	0.0268
	15	1.27	1.02	1.57	0.0295
	20	1.26	1.03	1.54	0.0277
ZIP Code organochlorine pounds: exposed if ≥75th pctl among rural controls					
≥8,292.13	None	1.81	1.12	2.93	0.0150
≥10,110.64	10	1.81	1.13	2.89	0.0134
≥9,897.86	15	1.68	1.05	2.70	0.0321
≥10,266.64	20	1.77	1.09	2.87	0.0203

Abbreviations: ac, acre; CI, confidence interval; lb, pound; OR, odds ratio; pctl, percentile.
[a] Odds ratios, 95% confidence intervals, and two-sided *P* values were estimated using multivariable conditional logistic regression (robust variance estimation) adjusting for liver disease (yes/no hepatitis, alcoholic liver disease, or non-specific cirrhosis) and diabetes, in addition to taking into account the matching factors and only among those originally entitled to Medicare due to attaining age 65 years (*n* = 258 cases, *n* = 1,543 controls).

4.8 FIGURES

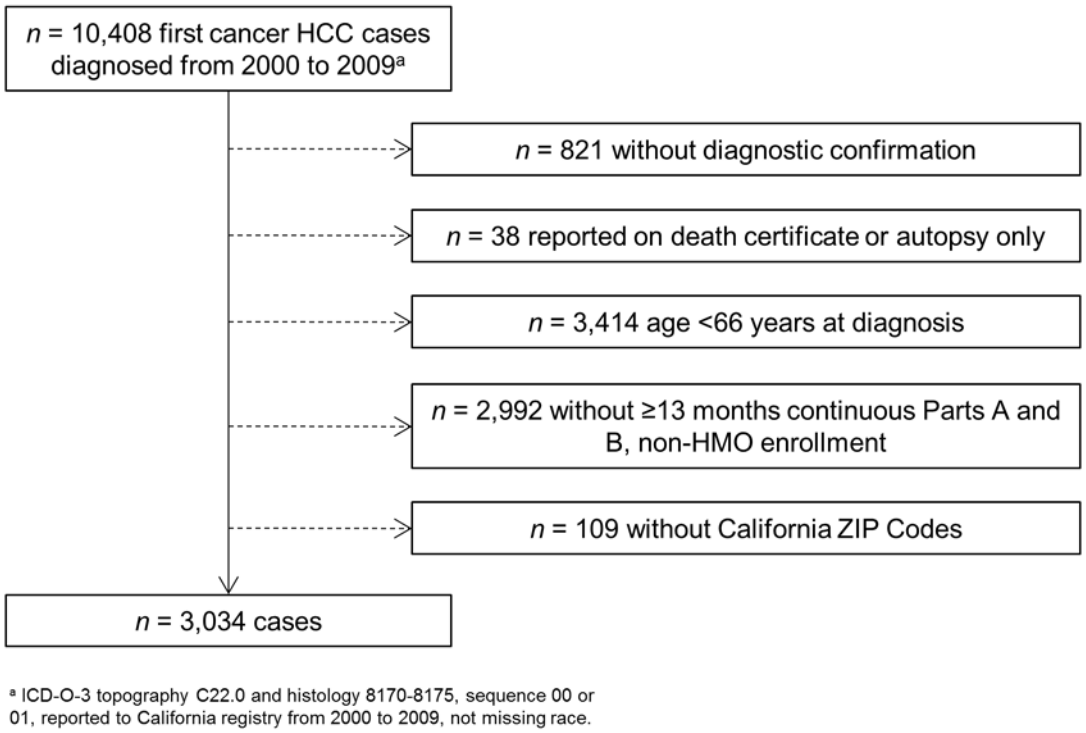


Figure 14. Eligibility Criteria Applied to First Cancer Hepatocellular Carcinoma Cases in California Diagnosed From 2000 to 2009 Using SEER-Medicare

5.0 GENERAL DISCUSSION

5.1 SUMMARY OF FINDINGS

Location is the fundamental focus of geographic information systems (GIS). Using a GIS, multiple data sources can be combined to both visualize and analyze spatial data, or data associated with locational information. Location can also play an important role in population health, where an individual's environment, including where an individual lives and works, has a direct effect on future health outcomes. Thus, location is extremely relevant to epidemiology, a branch of science seeking to describe and analyze determinants of human health. Specifically, GIS provides great utility in epidemiologic studies as a way to incorporate the location of both environmental exposures and humans to further our understanding of factors impacting human health. The three studies comprising this dissertation demonstrate the usage of GIS in studying human health via (1) presenting a new GIS method to estimate pesticide exposure for use in an epidemiologic study, (2) demonstrating the use of GIS to estimate pesticide exposure given aggregated spatial scale data, and (3) using GIS to link cancer and pesticide data to conduct an epidemiologic case-control study.

The first study presented a new GIS and remote sensing method to estimate individual-level agricultural pesticide exposure for use as an exposure metric in an epidemiologic study. This method, called the Landsat method, classifies Landsat satellite images into crop fields using

a maximum likelihood and per-field (segments) classification approach. These Landsat-classified crop fields are then matched to California Pesticide Use Report (PUR) data using the crop type and Public Land Survey System (PLSS) section of application. Pesticide exposure is subsequently estimated according to the pesticide-treated Landsat crop fields within a 500 m (radius) buffer around an individual's geocoded residence. Our study demonstrated that significantly more pesticide applications (60.3% vs. 57.4%; p=0.0002) were matched to Landsat crop fields compared to crop fields using the standard crop dataset called the land use survey (LUS). Furthermore, significantly more temporary crop pesticide applications were matched to Landsat crop fields compared to LUS crop fields (65.4% vs. 52.4%; p<0.0001). The Landsat method offers the opportunity to bridge the temporal gap between when pesticide exposure is to be estimated and the crop fields with which pesticide applications are matched, providing an alternative GIS pesticide exposure estimation method compared to the standard method that uses infrequently updated crop LUS's.

The second study explored three different methods of estimating ZIP Code-level pesticide exposure. The basis for each pesticide exposure estimation method was PLSS section pesticide application rates. Gold standard pesticide application rates were calculated according to pesticide-treated PLSS sections intersecting 500 m buffers around centroids of populated census block-habitable class polygons. ZIP Code rates were calculated using area weighting (PLSS section rates weighted by proportion of ZIP Code's area comprised by section), population weighting (section rates weighted by proportion of ZIP Code's population comprised by section), and road weighting (section rates weighted by proportion of ZIP Code's roads comprised by section). Area weighting achieved moderate agreement with the gold standard overall (weighted kappa 0.55; 95% confidence interval [CI] 0.52, 0.58) and substantial agreement in rural areas

(weighted kappa 0.63; 95% CI 0.57, 0.69). The results of this second study directly informed the methodological approach to estimate pesticide exposure in the third study, where the Surveillance, Epidemiology, and End Results (SEER)-Medicare database provided ZIP Codes for each study subject as a spatial variable capable of being linked to PUR data in a GIS.

The third study is the primary dissertation project - an epidemiologic case-control study examining the association between agricultural pesticide exposure and hepatocellular carcinoma (HCC). Cases and controls were sampled from the SEER-Medicare database, representing SEER cancer cases linked with Medicare enrollment data along with non-cancer Medicare beneficiaries in SEER catchment areas. Comorbidities as potential confounders were extracted from Medicare claims. Pesticide exposure was calculated using Medicare ZIP Codes, where area-weighted ZIP Code rates calculated using PLSS section pesticide application rates were summed and divided by the number of years of available Medicare data. Among rural California Medicare beneficiaries, previous annual ZIP Code exposure to ≥0.06 lb/ac of applied organochlorine pesticides was associated with a 52% increase in HCC risk compared to exposure <0.06 lb/ac after adjusting for liver disease and diabetes (adjusted odds ratio [OR] 1.52; 95% CI 1.02, 2.28; p=0.0415). HCC risk increased after accounting for a 20-year exposure lag (adjusted OR 1.81; 95% CI 1.19, 2.75; p=0.0058). Our study represents the first epidemiologic study examining pesticide exposure and HCC using GIS as well as the first epidemiologic study conducted among the U.S. population not relying on self-report to estimate pesticide exposure.

5.2 PUBLIC HEALTH SIGNIFICANCE

All three dissertation studies demonstrate significance in the field of public health as representing research that seeks to promote population health and disease prevention. Public health is an all-encompassing field, comprised of such disciplines as epidemiology and biostatistics, and is focused on measures to improve health via implementing educational programs, informing policymaking, and conducting research on disease detection and prevention. This dissertation contributes to the mission of public health by providing a new methodological technique by which to estimate pesticide exposure, associated with adverse human health outcomes such as cancers, that can be used in an epidemiologic study investigating pesticides and a disease outcome. Most importantly, this dissertation adds to a body of epidemiologic literature linking particular pesticides with HCC. In the context of understanding the role of an environmental exposure in human health, the capacity for the field of public health to impact and promote population health partly hinges on the research conducted to elucidate the impact of such exposures on diseases. If research can provide a substantive link between a purported exposure and disease, then efforts to mitigate exposure can be subsequently explored. And very importantly, the field of public health as a whole can move forward as we have formed a more comprehensive understanding of the factors that contribute to developing a particular disease.

The first study presented a new exposure method to capture an environmental determinant of health - pesticides. Accurate estimation of environmental exposures that could potentially impact human health allows us to form a better understanding of their exact role in human health outcomes. This specifically contributes to public health from the vantage point of providing another methodological tool to accurately estimate this environmental exposure that

incorporates multiple relevant data sources and is capable of considering multiple points in time. Thus, this study serves to push the bounds of using GIS in epidemiologic studies, adding to a research toolbox of environmental exposure assessment methods that can be used by public health investigators seeking to estimate this exposure.

The second study is relevant to public health from a research standpoint via demonstrating how data available at different scales can be combined for use in an epidemiologic study. An common issue encountered when conducting research using spatial data manifests itself as a discrepancy between the scale at which a phenomenon of interest operates, also called the operational scale, and the scale at which data is available, also called the analysis scale. For example, for purposes of patient confidentiality, a variable of interest may be available at some aggregated scale, such as ZIP Codes, while the scale at which the variable operates is much finer (e.g., residential level). The second study presented GIS methods to reconcile pesticide exposure estimation between the two spatial scales of PLSS sections and ZIP Codes. GIS overlay operations allowed for California PUR data, reported at the 1 mi^2 PLSS section level, to be aggregated, or scaled up, to the ZIP Code level, which are typically much larger in size compared to PLSS sections. This study addressed an important practical issue regarding the SEER-Medicare study (third study), where PURs reported according to PLSS sections had to be reconciled with SEER-Medicare-provided ZIP Codes for study subjects. Our approach of applying various spatial weighting methods to scale up spatial data and comparisons to a finer scale gold standard can be applied to study other environmental exposures relevant to public health research. Specifically, exploring different spatial weighting methods relevant to addressing human activity (i.e., road and population weighting) and a weighting method capable of being applied using any GIS (i.e., area weighting) can be applied towards investigating other

environmental exposures aggregated at other spatial scales (e.g., census tracts) in public health research.

The third and primary study addresses a disease significantly increasing in incidence in the U.S. population and associated with low survival - hepatocellular carcinoma. From a public health perspective, a disease associated with vast consequences in terms of morbidity, mortality, and quality of life demands research efforts to elucidate its risk factors. Furthermore, up to 50% of all HCC cases diagnosed in the U.S. population occur among those with no known risk factors. Given the epidemiologic evidence regarding pesticide exposure as a risk factor for HCC in the U.S. is inconsistent, the opportunity to utilize SEER-Medicare, a population-based data source providing access to comorbidities as potential confounders and high quality cancer outcomes data, served as an advantageous public health research endeavor addressing this often fatal disease. The results of this study demonstrated that previous exposure to organochlorines, a pesticide chemical class comprised of many chemicals, e.g., dichlorodiphenyltrichloroethane (DDT), that have since been banned in the U.S. due to environmental persistence and adverse health effects, was associated with an increased risk of developing HCC between 2000 and 2009 among rural Californians. Therefore, this research provides potential evidence of organochlorine exposure affecting HCC risk that warrants further investigation, especially as past agricultural pesticide application practices may continue to impact human health outcomes in U.S. populations outside of California. This research provides an epidemiologic foundation upon which to build and improve upon our understanding of agricultural pesticide exposure and HCC.

5.3 STRENGTHS AND LIMITATIONS

The three studies of this dissertation offer many methodological strengths that both improve on previous studies and provide tools and evidence with which future research can be pursued. The strengths of the first study include presenting a novel GIS method to estimate pesticide exposure that is superior to the standard method in using temporally current remote sensing data sources. This work is built on a small body of previous research that pioneered the use of remote sensing in pesticide exposure estimation through using modern and improved image classification methods (maximum likelihood and per-field via segmentation) as compared to previously utilized methods (minimum distance classification). Ultimately, we sought to create a new and improved GIS-based pesticide exposure estimation method that would address a current limitation of the standard GIS pesticide exposure method in California - land use changes. The standard method relies on a crop field dataset (LUS's) that can be up to 10 years removed from the year in which pesticide exposure is to be estimated. During this temporal gap, substantial land use changes can occur such as urbanization, crop field changes, etc. The fundamental issue underlying this limitation is that California pesticide data, available from 1974 until present day, is matched to crop fields from a LUS in a particular PLSS section that may no longer be present or have since changed. Therefore, the primary strength of this research is that it allows an investigator to estimate agricultural pesticide exposure using a GIS for any given time period without having to use outdated crop fields. Rather, Landsat remotely sensed satellite images can be classified into crop fields using images from the year in which pesticide exposure is to be estimated, which attempts to close the temporal gap between pesticide application data and crop field data.

Limitations of the first study include a hard classification method where only one crop type is assigned to each segment. This is problematic as some farmers practice intra-annual crop rotations, where crops can be rotated within a given year, while our method was better suited to capture annual crop rotations. However, our results demonstrated that even when we were unable to achieve tier 1 matches between pesticide data and Landsat crop fields, we were able to achieve tier 2 matches and classify truly agricultural crop fields into some type of agricultural land use. In other words, even if a direct pesticide application-to-crop field match could not be achieved (tier 1), tier 2 matches were achieved, which distributed pesticides to all other agricultural land uses as classified by Landsat images in a PLSS section (preferred over distributing to the entire section [tier 3 match]). Another limitation was the sole reliance on Normalized Difference Vegetation Index (NDVI) values to classify images into crop fields. Although we used a time series of NDVI images to help classify crops, crop misclassification was still manifest in the accuracy assessment, particularly among native vegetation, field, and idle crops. Crop misclassification has a direct impact on pesticide exposure estimation in terms of the aforementioned tiers 1 to 3 matches. For example, if tier 1 matches are not achieved when they should be since there truly were particular crop types in a PLSS section in 1985, pesticide exposure estimated using 500 m buffers around residential locations nearby to these crop fields will be over- or underestimated.

Strengths of the second study include being the first study, to the best of our knowledge, to explore the accuracy of ecologic metrics estimating pesticide exposure. Particularly when examining spatial data that is inherently associated with some geographic location on Earth, issues will emerge regarding spatial scale discrepancies related to data availability and useful/relevant data. To address working with datasets of different spatial scales, the second

paper offers a framework that can be used to explore other environmental exposures available at various spatial scales: incorporate ancillary data to spatially weight one scale of data up to the other (i.e., spatial aggregation if source unit of data in its current form is smaller in size relative to the target unit at which data are available) and calculate statistics (e.g., weighted kappa) to compare to a finer scale (relative to the ecologic metric) gold standard. The greatest strength of this research was internal, where the results of this project directly informed the primary SEER-Medicare study. In order to perform the data linkage that would allow for exploring the relationship between agricultural pesticide exposure and HCC using California PURs and SEER-Medicare, respectively, a method in good agreement with a gold standard was desired to scale up PURs reported at the PLSS section level to the Medicare ZIP Code level. Thus, based on the results of the second study, area weighting was used to perform this data linkage.

A limitation of the second study is exposure misclassification associated with using ecologic metrics. When the spatial scale of the data is being altered, whether via scaling up (as used in this study) or scaling down (e.g., areal interpolation to disaggregate data), the variable of interest will inevitably be altered. For example, in the case of the second study, area-weighted ZIP Code pesticide exposure tended to underestimate gold standard pesticide exposure. In applying area weighting, pounds of applied pesticides are essentially being distributed across an entire ZIP Code, which in rural areas, can span over hundreds of square miles. Therefore, the pesticide exposure assigned at the ZIP Code level may over- or underestimate an individual's true pesticide exposure experience. However, as pesticide exposure was estimated using the same methods for both cases and controls in the third study, any bias would be nondifferential, potentially attenuating the true association between pesticides and HCC.

Strengths of the third study include its novelty in using GIS to study pesticides and HCC. GIS and the spatial sciences are a burgeoning scientific discipline with many tools and concepts that are relevant and useful in the context of conducting epidemiologic studies. In the case of this third study, GIS allowed for a novel data linkage between two large population-based data sources in the U.S. - California PURs and SEER-Medicare - using PLSS sections (geographic level of reporting of the PUR database) and Medicare ZIP Codes for study subjects. Therefore, GIS allowed the third study to explore whether or not agricultural pesticide exposure is associated with HCC in California. Another methodological strength was not relying on self-reported pesticide exposure and occupation. All four previous U.S. studies examining this topic have relied on self-reported occupation reported on death certificates, medical records, and from the individual or next-of-kin. Self-reported measures are subject to recall bias, where exposure may be over- or underestimated. The third study was able to use the PUR database, the world's most comprehensive pesticide reporting system, to identify specific chemicals applied on crops reported by California farmers and commercial pest control operators. Recall bias is minimized by using standardized pesticide reports submitted to the California government with information on specific chemicals dating back to 1974 to reconstruct historical exposure. One of the interesting results from the third study was the demonstration of a potential exposure lag, which is relevant to studying chronic diseases such as cancers. It is hypothesized that exposure lags of 20 years or more may exist between initial exposure to various HCC risk factors and clinical diagnosis of HCC. Therefore, demonstration of statistically significant results when examining exposure lags of 10, 15, and 20 years both bolstered scientific evidence regarding the clinical progression of this disease and demonstrated the great utility of using the historical PUR dataset to reconstruct past pesticide exposure in order to examine a potential exposure lag.

A limitation of the third study was the use of Medicare claims to extract potential confounders. A motivating factor for using the SEER-Medicare dataset to conduct this study was access to information regarding chronic hepatitis C virus (HCV) and hepatitis B virus (HBV), which are established risk factors for HCC. It was hypothesized that negative confounding potentially existed, where failure to adjust for viral hepatitis would underestimate the true effect of pesticide exposure on HCC. This is due to how HCV and HBV are hypothesized to be largely urban phenomena by virtue of their predominant routes of transmission via intravenous drug use and illicit sexual behavior, and agricultural pesticide use is a rural phenomenon. However, an inherit limitation of using administrative data for comorbidities is the lack of sensitivity for particular health conditions and also how our study was limited to examining claims data beginning at Medicare entitlement typically at the age of 65 years. Therefore, usage of Medicare claims may be associated with residual confounding as all individuals with viral hepatitis were not captured in the SEER-Medicare study population.

The primary limitation of the third study was exposure misclassification due to using ZIP Codes to estimate pesticide exposure. ZIP Codes are often large features created for the purposes of mail delivery. Pesticide exposure estimated in urban areas was potentially in better agreement with what that particular urban individual experienced (urban ZIP Codes are typically small in size). However, rural ZIP Codes can be quite large in size (hundreds of square miles) and rural areas were of interest to the study as the independent variable of interest, pesticide exposure, predominantly occurs in these areas. Ultimately, given the SEER-Medicare data source, we cannot pinpoint where each study subject resided within a ZIP Code. Furthermore, usage of ZIP Codes likely under- or overestimated pesticide exposure compared to what would be estimated had we had access to finer scale geographic data that is used in practice to capture agricultural

pesticide exposure occurring via residential proximity to pesticide applications (geocoded residential locations). However, as previously mentioned in the context of the second study, as ZIP Codes were used in the same way to estimate pesticide exposure for each case and control, any bias would be nondifferential and attenuate study results. Furthermore, observing that the majority of cases and controls had one to two Medicare ZIP Codes ever and minimal rural-to-urban or urban-to-rural ZIP Code changes over a person's Medicare-captured lifetime lends confidence to the methods of the third study capturing, as best as possible, the pesticide exposure experience of these study subjects given their Medicare ZIP Code data.

5.4 FUTURE RESEARCH AND CONCLUSIONS

In order to expand the bounds of using GIS in epidemiology, spatial tools and concepts can be directly applied in epidemiologic studies to not only improve the methods used to estimate an environmental exposure of interest, for example, but to help link multiple data sources in order to address timely and salient public health issues. The studies comprising this dissertation demonstrate how incorporation of geospatial techniques can improve how we measure environmental determinants of health by addressing some concepts fundamental to epidemiology and studying chronic diseases - latency periods, specificity of measured chemicals, and ambient exposure.

The centerpiece of this dissertation is the SEER-Medicare epidemiologic case-control study examining the association between agricultural pesticides and HCC. This study now represents the 18th epidemiologic study conducted researching this topic and provided evidence of a statistically significant positive association between HCC and organochlorine pesticides, a

chemical class comprising pesticides previously associated with HCC in other literature, including DDT. This study is associated with many methodological advantages compared to the four previously conducted U.S.-based case-control studies (only one of which reported a statistically significant positive association), including information on potential confounders and non-reliance on self-report for pesticide exposure. Therefore, overall, this study makes a significant contribution to a modestly sized body of epidemiologic literature investigating pesticides and HCC and provides evidence that warrants future research to further clarify their association. Identifying and understanding risk factors for HCC, particularly in the U.S., is a significant public health issue as this disease is predicted to continue increasing in incidence and remains associated with a low five-year relative survival rate.

The results of this dissertation provide many future research opportunities. Regarding the first study, additional ancillary information, e.g., elevation, along with an improved classification approach, e.g., Classification and Regression Tree (CART), would improve the Landsat image classification method. This would allow for an enhanced ability to better distinguish between crop types within PLSS sections and thus match to more PUR pesticide applications to improve pesticide exposure estimation for use in epidemiologic studies. Regarding the second study, a validity study could be conducted to determine the extent to which measures of association from an epidemiologic study are impacted when using a ZIP Code pesticide exposure metric vs. a more large scale metric (e.g., geocoded residences). Regarding the third study, a future study investigating the link between pesticides and HCC in the U.S. population where pesticide exposure is estimated using biological samples (e.g., serum), considered the gold standard method of addressing human pesticide exposure from all routes, and non-administrative data are

used to collect comorbidity information would be valuable in helping to elucidate their potential association.

In conclusion, these three studies demonstrate the integration of GIS in public health research, specifically in epidemiology and environmental exposure assessment. GIS, which allows for the application of methods to study spatial phenomena, allowed for the successful execution of a case-control study examining the role of (geographically reported) pesticides as a risk factor for HCC. Most notably, the third and primary study of this dissertation demonstrated the use of GIS to perform a novel data linkage between two population-based data sources, California PURs and SEER-Medicare data, to conduct an epidemiologic study using Medicare ZIP Codes. As the first epidemiologic study conducted in the U.S. population demonstrating a statistically significant positive association between pesticide exposure (not based on self-report) and HCC after adjusting for potential confounders, future research should explore using an improved pesticide exposure metric and a more sensitive and specific comorbidity data source to further characterize the role of pesticides as a risk factor for HCC - a disease that continues to impact numerous individuals in the U.S. and around the world.

APPENDIX A: LANDSAT IMAGES USED FOR NDVI SIGNATURES AND CLASSIFICATION

Table 24. Landsat Images Classified for Inclusion in the 1990 NDVI Signatures Extent

	Path 41[a]		Path 42[a]	
	Row 35	Row 36	Row 35	Row 36
Month	Acquisition date (% cloud cover)	Acquisition date (% cloud cover)	Acquisition date (% cloud cover)	Acquisition date (% cloud cover)
January	1/22/1990 (10)	1/22/1990 (0)	1/29/1990 (0)	1/29/1990 (0)
February	2/15/1990 (0)	2/15/1990 (0)	2/14/1990 (0)	2/14/1990 (0)
March	3/27/1990 (10)	Excluded[b]	3/18/1990 (0)	3/18/1990 (10)
April	4/28/1990 (40)	4/28/1990 (20)	4/3/1990 (0)	4/3/1990 (40)
May	5/30/1990 (10)	5/30/1990 (10)	5/5/1990 (0)	5/5/1990 (10)
June	Excluded[b]	Excluded[b]	6/6/1990 (0)	6/6/1990 (10)
July	None available	None available	7/8/1990 (0)	7/8/1990 (50)
August	8/18/1990 (10)	8/18/1990 (10)	8/25/1990 (0)	8/25/1990 (0)
September	9/3/1990 (0)	9/3/1990 (0)	9/10/1990 (0)	9/10/1990 (20)
October	10/5/1990 (0)	10/5/1990 (0)	10/12/1990 (0)	10/28/1990 (0)

[a] These Landsat images were processed for inclusion in the 1990 NDVI signatures extent used as training and classification data for the accuracy assessment and as training data for the classification of 1985 Landsat images. All images were captured using the Landsat 4 or 5 Thematic Mapper sensor.
[b] Images were excluded due to excessive cloud cover.

Table 25. Landsat Images Classified for Inclusion in the 1985 Imagery Extent

	Path 41[a]		Path 42[a]	
	Row 35	Row 36	Row 35	Row 36
Month	Acquisition date (% cloud cover)	Acquisition date (% cloud cover)	Acquisition date (% cloud cover)	Acquisition date (% cloud cover)
January	Excluded[b]	Excluded[b]	1/31/1985 (20)	1/31/1985 (0)
February	2/25/1985 (10)	2/25/1985 (10)	None available	None available
March	Excluded[b]	Excluded[b]	3/20/1985 (0)	3/20/1985 (0)
April	4/14/1985 (10)	4/14/1985 (10)	4/5/1985 (0)	4/5/1985 (50)
May	5/16/1985 (10)	5/16/1985 (10)	5/23/1985 (0)	5/23/1985 (20)
June	6/1/1985 (10)	6/17/1985 (1)	6/8/1985 (0)	6/8/1985 (18)
July	7/3/1985 (0)	7/3/1985 (0)	7/26/1985 (1)	None available
August	None available	8/20/1985 (10)	8/11/1985 (0)	8/11/1985 (50)
September	9/21/1985 (0)	9/21/1985 (0)	9/12/1985 (1)	9/12/1985 (0)
October	Excluded[b]	Excluded[b]	10/14/1985 (0)	10/14/1985 (0)

[a] These Landsat images were processed for inclusion in the 1985 imagery extent to estimate agricultural pesticide exposure In 1985. All images were captured using the Landsat 5 Thematic Mapper sensor.
[b] Images were excluded due to excessive cloud cover.

APPENDIX B: LANDSAT AND LAND USE SURVEY (LUS) THREE-TIER PESTICIDE APPLICATION MATCHING METHODOLOGY

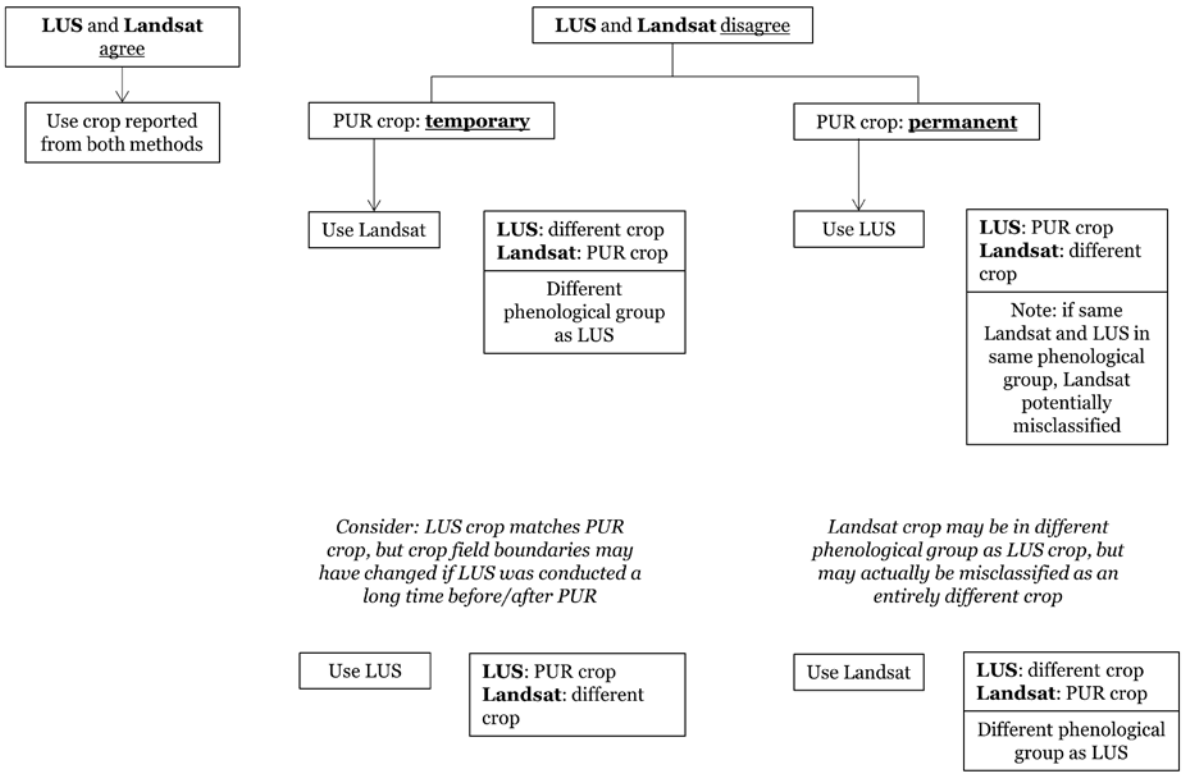


Figure 15. Combined Landsat and LUS Pesticide-to-Crop Matching Three-Tier Method Considering Temporary and Permanent Crops

APPENDIX C: SUPPLEMENTAL ANALYSIS: HEPATITIS TESTING IN THE SEER-MEDICARE POPULATION

C.1 INTRODUCTION

It is expected that a high proportion of hepatocellular carcinoma (HCC) cases in our study tested for hepatitis. However, it was important to explore the proportion of controls that tested for hepatitis. Both chronic hepatitis B virus (HBV) and hepatitis C virus (HCV) infection were assessed as confounders in our statistical analyses. Estimating the proportion of controls testing for these health conditions would give us an idea of potential undiagnosed hepatitis among controls. Furthermore, some controls might have been cases had they been tested for hepatitis and underwent further clinical inspection. We also wanted to assess the extent to which there might be a self-selection bias in our study, where selection into our study could depend on pesticide exposure. Individuals living near pesticide-exposed areas might be concerned about their health and ask their doctor to be tested for hepatitis, which would prompt further investigation and uncover HCC.

C.2 METHODS

We examined Healthcare Common Procedure Coding System (HCPCS) codes, also known as Current Procedural Terminology (CPT)-4 codes, using Medicare claims of all cases and controls. HCPCS codes are a proprietary standardized numeric coding system maintained by the American Medical Association (AMA) used by healthcare professionals to bill for their services. We extracted the HCPCS codes in Table 26 from Medicare outpatient and carrier claims. The inpatient file only included major surgical procedure codes and was thus excluded from this analysis. HCPCS codes ending in 'F' are Category II codes, or used for performance measurement such as patient history. None of these Category II codes were found in the Medicare claims included in our analyses. Pesticide exposure is defined as ZIP Code pesticide application rates from all pesticide chemical classes included in the study (organophosphates, organochlorines, and carbamates) using available Medicare ZIP Codes from 1974 up until the year before selection for controls. The earliest available ZIP Code was carried back to 1974. Pesticide application rates were categorized using tertiles among controls.

C.3 RESULTS

As expected, a higher proportion of cases compared to controls tested for HBV and/or HCV before diagnosis. This is observed when examining the full study sample (Table 27) and the rural study sample (Table 28). High pesticide exposure was associated with not testing for hepatitis among all controls, while among rural controls, pesticide exposure was slightly higher among

those testing for hepatitis (Table 29). Table 30 shows the distribution of age, year of selection, sex, race, and California residence with respect to hepatitis testing among all controls and rural controls. Among all controls, whites are less likely to test for hepatitis, while Asians are more likely to test. These relationships are slightly diminished when examining rural controls. Among rural residents who did not test for hepatitis at least one year before diagnosis/selection, a higher proportion of cases were exposed to the high pesticide tertile compared to controls (Table 31).

C.4 DISCUSSION

As 1% of the elderly U.S. population is estimated to have chronic HCV infection (233) and <2% have chronic HBV infection (234), a relatively high proportion of controls were tested for hepatitis. This lends confidence to Medicare claims being able to capture the controls truly having hepatitis, and also to the unlikelihood of cases potentially misclassified as controls. However, it cannot be ruled out that while a high proportion of individuals were tested for hepatitis, those who truly have hepatitis were not well captured among this tested population. The high proportion of controls testing for hepatitis may reflect recommended HCV testing among high risk populations (e.g., intravenous drug users). The latest a case/control was born in our study was 1943, so there were no “Baby Boomers” included.

The relationship between hepatitis testing and pesticide exposure might also explain the null effects in the primary analysis. Cases were more frequently tested for hepatitis compared to controls. Not testing for hepatitis was associated with higher pesticide exposure. This may reflect how those not testing for hepatitis reside in rural areas with less access to healthcare compared to

their urban counterparts. Their rural residence explains their higher pesticide exposure. As not testing for hepatitis was more common among controls, this selection bias may have manifested itself as an artificially inflated pesticide exposure among controls, biasing results towards the null hypothesis.

Due to increased awareness and enhanced medical surveillance, preferential disease detection in pesticide-exposed individuals might create a non-causal, artificial association between pesticide exposure and HCC. Table 29 shows a two-fold greater prevalence of hepatitis testing (≥1 year before selection) among rural controls in high vs. low pesticide tertiles (32/565 vs. 15/564). Therefore, among rural residents, a medical detection bias might have explained some or all of the association observed between pesticides and HCC in our study. However, after exploring the relationship between pesticides and HCC among rural residents who did not test for hepatitis, cases are still more likely than controls to have been exposed to the high vs. low pesticide tertile (91/565 vs. 88/564) (Table 31).

C.5 TABLES

Table 26. Healthcare Common Procedure Coding System (HCPCS) Codes for Hepatitis C and B Virus Testing

HCPCS code	Description
80074	Acute hepatitis panel This panel must include the following: Hepatitis A antibody (HAAb), IgM antibody (86709) Hepatitis B core antibody (HBcAb), IgM antibody (86705) Hepatitis B surface antigen (HBsAg) (87340) Hepatitis C antibody (86803)
86704	Hepatitis B core antibody (HBcAb); total
86705	Hepatitis B core antibody (HBcAb); IgM antibody
86706	Hepatitis B surface antibody (HBsAb)
86707	Hepatitis Be antibody (HBeAb)
86803	Hepatitis C antibody
86804	Hepatitis C antibody confirmatory test (e.g., immunoblot or RIBA)
87340	Infectious agent antigen detection by enzyme immunoassay technique, qualitative or semiquantitative, multiple step method; hepatitis B surface antigen (HBsAg)
87341	Infectious agent antigen detection by enzyme immunoassay technique, qualitative or semiquantitative, multiple step method; hepatitis B surface antigen (HBsAg) neutralization
87350	Infectious agent antigen detection by enzyme immunoassay technique, qualitative or semiquantitative, multiple step method; hepatitis Be antigen (HBeAg)
87515	HepB antigen
87516	HepB antigen
87517	HepB antigen
87520	Hepatitis C antigen, direct probe technique
87521	Hepatitis C antigen, amplified probe technique
87522	Hepatitis C antigen, quantification
87902	Infectious agent genotype analysis by nucleic acid (DNA or RNA); Hepatitis C virus
87912	Infectious agent genotype analysis by nucleic acid (DNA or RNA); Hepatitis B virus
3217F	RNA testing for Hepatitis C viremia ordered at initial evaluationor previously performed (HEP-C), code deleted 1/1/2009
3218F	RNA testing for Hepatitis C documented as performed within 6 months prior to initiation of antiviral treatment for Hepatitis C (HEP-C)
3219F	Hepatitis C genotype testing documented as performed prior toinitiation of antiviral treatment for Hepatitis C (HEP-C), deleted on 1/1/2009
3220F	Hepatitis C quantitative RNA testing documented as performed at 12 weeks from initiation of antiviral treatment (HEP-C)
3265F	Ribonucleic acid (RNA) testing for Hepatitis C viremia ordered or results documented (HEP C)
3266F	Hepatitis C genotype testing documented as performed prior to initiation of antiviral treatment for Hepatitis C (HEP C)
3513F	Hepatitis B screening documented as performed (HIV)
3514F	Hepatitis C screening documented as performed (HIV)
4150F	Patient receiving antiviral treatment for Hepatitis C (HEP-C)
4151F	Patient not receiving antiviral treatment for Hepatitis C (HEP-C)

Table 27. Hepatitis B and C Testing Among Cases and Controls (Full Study Sample)

	Cases (n=3,034) n (%)	Controls (n=14,991) n (%)
Hepatitis testing any time before diagnosis/selection		
Yes	1,275 (42.0)	1,947 (13.0)
No	1,759 (58.0)	13,044 (87.0)
Hepatitis testing ≥1 year before diagnosis/selection		
Yes	863 (28.4)	1,586 (10.6)
No	2,171 (71.6)	13,405 (89.4)

Table 28. Hepatitis B and C Testing Among Cases and Controls (Rural Study Sample)

	Cases (n=306) n (%)	Controls (n=1,758) n (%)
Hepatitis testing any time before diagnosis/selection		
Yes	99 (32.4)	93 (5.3)
No	207 (67.6)	1,665 (94.7)
Hepatitis testing ≥1 year before diagnosis/selection		
Yes	53 (17.3)	71 (4.0)
No	253 (82.7)	1,687 (96.0)

Table 29. Pesticide Exposure and Hepatitis Testing: Full and Rural Controls

	All controls: pesticide exposure, n (%)[a]			Rural controls: pesticide exposure, n (%)[a]		
	Low exposure	Moderate exposure	High exposure	Low exposure	Moderate exposure	High exposure
	≤0.001 lb/ac	0.001-0.02 lb/ac	≥0.02 lb/ac	≤0.48 lb/ac	0.48-0.92 lb/ac	≥0.92 lb/ac
Hepatitis testing any time before selection						
Yes	779 (40.0)	731 (37.5)	437 (22.4)	18 (19.4)	33 (35.5)	42 (45.2)
No	4,170 (32.0)	4,213 (32.3)	4,661 (35.7)	561 (33.7)	549 (33.0)	555 (33.3)
Hepatitis testing ≥1 year before selection						
Yes	648 (40.9)	586 (37.0)	352 (22.2)	15 (21.1)	24 (33.8)	32 (45.1)
No	4,301 (32.1)	4,358 (32.5)	4,746 (35.4)	564 (33.4)	558 (33.1)	565 (33.5)

Abbreviations: ac, acre; lb, pound.

[a] Pesticide exposure, calculated as ZIP Code pesticide application rates using organophosphates, organochlorines, and carbamates from 1974 until the year before selection, were categorized according to tertiles among controls.

Table 30. Characteristics of Controls Testing for Hepatitis

	All controls				Rural controls			
	Hepatitis testing any time before selection		Hepatitis testing ≥1 year before selection		Hepatitis testing any time before selection		Hepatitis testing ≥1 year before selection	
	Yes n (%)	No n (%)	Yes n (%)	No n (%)	Yes n (%)	No n (%)	Yes n (%)	No n (%)
Age at selection, mean (SD)	75.7 (5.8)	75.0 (6.3)	75.9 (5.7)	75.0 (6.3)	75.2 (6.0)	74.5 (6.2)	74.8 (6.0)	74.5 (6.2)
Year of selection[a]								
2000-2002	263 (13.5)	3,444 (26.4)	172 (10.8)	3,535 (26.4)	<11	446 (26.8)	<11	448 (26.6)
2003-2005	485 (24.9)	3,868 (29.7)	374 (23.6)	3,979 (29.7)	19 (20.4)	509 (30.6)	13 (18.3)	515 (30.5)
2006-2009	1,199 (61.6)	5,732 (43.9)	1,040 (65.6)	5,891 (44.0)	66 (71.0)	710 (42.6)	52 (73.2)	724 (42.9)
Sex								
Male	1,219 (62.6)	8,250 (63.3)	982 (61.9)	8,487 (63.3)	53 (57.0)	1,115 (67.0)	43 (60.6)	1,125 (66.7)
Female	728 (37.4)	4,794 (36.8)	604 (38.1)	4,918 (36.7)	40 (43.0)	550 (33.0)	28 (39.4)	562 (33.3)
Race[a]								
White	491 (25.2)	7248 (55.6)	401 (25.3)	7,338 (54.7)	52 (55.9)	1,057 (63.5)	40 (56.3)	1,069 (63.4)
Black	86 (4.4)	657 (5.0)	64 (4.0)	679 (5.1)	<11	54 (3.2)	<11	55 (3.3)
Other	168 (8.6)	1,098 (8.4)	141 (8.9)	1,125 (8.4)	<11	115 (6.9)	<11	116 (6.9)
Asian	1,027 (52.8)	2,897 (22.2)	835 (52.7)	3,089 (23.0)	<11	146 (8.8)	<11	148 (8.8)
Hispanic	174 (8.9)	1,078 (8.3)	144 (9.1)	1,108 (8.3)	22 (23.7)	279 (16.8)	16 (22.5)	285 (16.9)
Native American	<11	66 (0.5)	<11)	66 (0.5)	<11	14 (0.8)	<11	14 (0.8)
California residence[a]								
1-5 years	416 (21.4)	3,729 (28.6)	288 (18.2)	3,857 (28.8)	<11	433 (26.0)	<11	435 (25.8)
6-10 years	739 (38.0)	4,568 (35.0)	609 (38.4)	4,698 (35.1)	34 (36.6)	555 (33.3)	29 (40.9)	560 (33.2)
≥11 years	792 (40.7)	4,747 (36.4)	689 (43.4)	4,850 (36.2)	52 (55.9)	677 (40.7)	37 (52.1)	692 (41.0)

[a] In accordance with the SEER-Medicare data use agreement, cell sizes <11 have been suppressed.

Table 31. Pesticide Exposure and Hepatocellular Carcinoma Among Rural Residents Not Testing for Hepatitis

	Cases (n=253)[a] n (%)	Controls (n=1,687)[a] n (%)
Pesticide exposure[b]		
Low: ≤0.48 lb/ac	88 (34.8)	564 (33.4)
Moderate: 0.48-0.92 lb/ac	74 (29.3)	558 (33.1)
High: ≥0.92 lb/ac	91 (36.0)	565 (33.5)

[a] Only considered rural residents who did not test for hepatitis at least one year before diagnosis/selection.
[b] Pesticide exposure, calculated as ZIP Code pesticide application rates using organophosphates, organochlorines, and carbamates from 1974 until the year before selection, were categorized according to tertiles among rural controls.

APPENDIX D: GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION (GWLR): ORGANOCHLORINE PESTICIDE EXPOSURE AND HEPATOCELLULAR CARCINOMA RISK AMONG RURAL CALIFORNIA RESIDENTS

D.1 INTRODUCTION

Geographically weighted regression (GWR) is an exploratory spatial method that allows some or all of the estimates from the independent variables in a regression model to vary across space (235). We specifically used geographically weighted logistic regression (GWLR) for a binary outcome to visualize the risk of hepatocellular carcinoma (HCC) conferred by organochlorine pesticide exposure across rural California.

D.2 METHODS

Rural residents from the SEER-Medicare case-control study in Section 4 were analyzed (n=306 cases and n=1,768 controls). Organochlorine pesticide exposure was defined as an annual ZIP Code pesticide application rate (lb/ac) calculated using Medicare ZIP Codes. Pesticide Use Report (PUR) applied pounds were matched to PLSS sections and divided by section acreage to calculate rates. PLSS section rates were aggregated up to the ZIP Code level using area

weighting (Section 4). For each study subject, using available California ZIP Codes from 1991 up until the year before diagnosis/selection and carrying back the earliest available ZIP Code to 1974, ZIP Code rates were summed and divided by the number of years of California residence. Each case and control was assigned to the centroid of the ZIP Code that they resided in the majority of the time. Note that multiple cases and controls could have been assigned to the same centroid. These centroids associated with cases and controls are referred to as data or sample points, which were used to estimate each regression model.

Regression points (n=1,020) were created within the 220 rural ZIP Codes in which the cases and controls resided. At least five points ≥1,000 m apart within each rural ZIP Code were randomly generated. At each regression point, a logistic regression model was fit using the data points within a kernel centered on that regression point. The kernel is adaptive, varying in size so that the number of data points within each kernel (i.e., bandwidth) used to estimate each logistic regression model remained constant. The adaptive kernel is recommended for data points not distributed evenly across space. The GWR4 program selected the optimal bandwidth of 776. In other words, 776 cases and controls (some associated with the same ZIP Code centroid) falling within a kernel centered on each regression point were used to estimate a logistic regression model at each regression point. A Gaussian weighting scheme was selected, where data points farther in distance from a given regression point were weighted less when estimating that particular logistic regression model. The Gaussian weight decreases continuously and gradually as one moves away from the regression point. Variables in each logistic regression model examining the dependent variable of hepatocellular carcinoma (yes/no) included: categorical annual ZIP Code organochlorine pesticide exposure (exposed if ≥0.06 lb/ac [median among rural controls]), liver disease, diabetes, and the five matching factors (year, age, sex, race, and years of

California residence). All predictors were allowed to vary geographically. Adjusted odds ratio (ORs) for organochlorine pesticide exposure estimated at each regression point were subject to inverse distance-weighted (IDW) interpolation to create a smooth, continuous raster surface of adjusted ORs (12 nearest neighbors used to calculate OR at each unmeasured location; masked using rural ZIP Codes). All spatial analyses were conducted using ArcGIS and GWR4 (212, 236).

D.3 RESULTS

Figure 16 shows interpolated adjusted odds ratios for hepatocellular carcinoma conferred by previous annual ZIP Code organochlorine pesticide exposure to ≥0.06 lb/ac. This maps provides visual evidence of HCC risk potentially varying across rural ZIP Codes in California. Some areas within Contra Costa, Imperial, Riverside, Sacramento, San Joaquin, Solano, and Yolo Counties were associated with null to borderline null effects (adjusted ORs close to 1). The highest adjusted ORs were observed in the agriculturally intensive Central Valley region, including Fresno and Kern Counties. Geographic variation in risk may be attributed to varying pesticide application practices, differences in predominant crop types, agricultural occupation, and/or sample size.

D.4 DISCUSSION

GWLR is a useful exploratory spatial method allowing for the visualization of spatial data to describe spatial patterns of disease. We observed some visual evidence of varying HCC risk conferred by organochlorine pesticide exposure. However, an improved approach would be to use point data, for example, geocoded residential locations, which would provide a more accurate depiction of potentially geographically varying individual-level HCC risk. Usage of ZIP Code centroids is problematic as cases and controls may not reside at these particular locations, let alone be exposed according to the levels measured at the ZIP code level. Furthermore, GWLR allows for a preliminary visualization of the data. Sophisticated spatial methods, such as spatial generalized linear mixed modeling, should be used to explore this topic further.

D.5 FIGURES

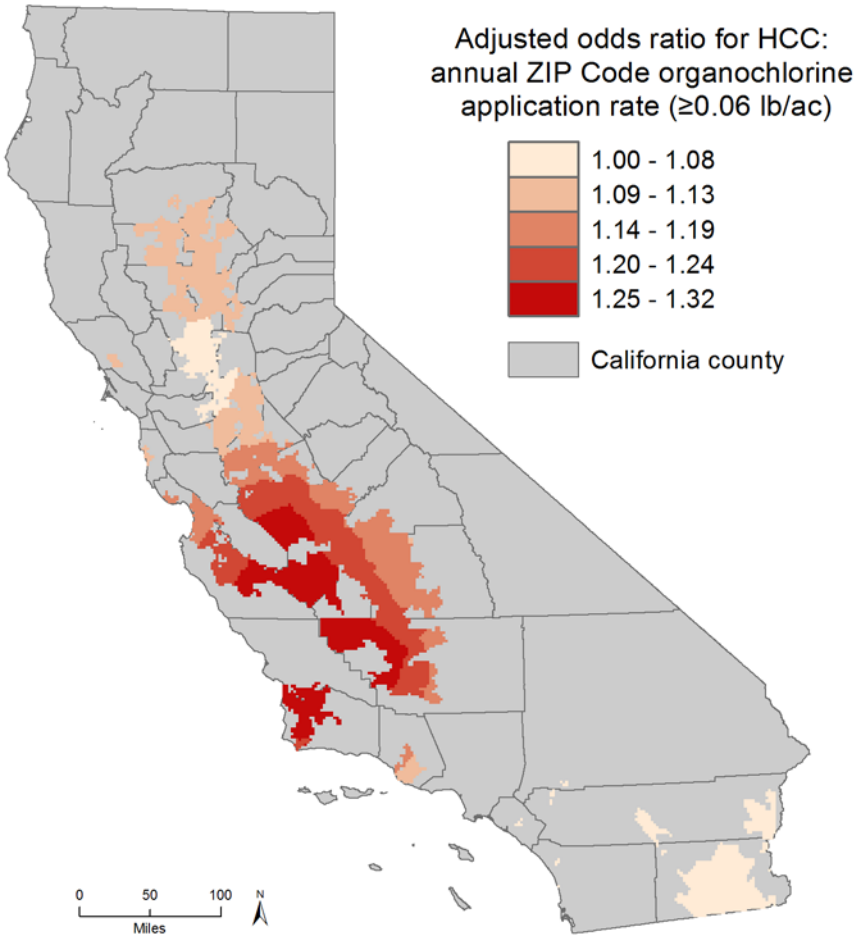


Figure 16. Geographically Weighted Logistic Regression (GWLR): Adjusted Odds Ratios for Hepatocellular Carcinoma Conferred by Organochlorine Pesticide Exposure Among Rural California Residents Using SEER-Medicare, 2000-2009

BIBLIOGRAPHY

1. McGlynn KA, London WT. The global epidemiology of hepatocellular carcinoma: present and future. *Clinics in liver disease* 2011;15(2):223-43, vii-x.
2. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA: a cancer journal for clinicians* 2011;61(2):69-90.
3. Carr BI. *Hepatocellular Carcinoma: Diagnosis and Treatment*. 2nd ed. Philadelphia, PA: Humana Press; 2010.
4. Ferlay J SH, Bray F, Forman D, Mathers C and Parkin DM. GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 Lyon, France: International Agency for Research on Cancer; 2010. Accessed March 14 2013.
5. Chen JG, Egner PA, Ng D, et al. Reduced aflatoxin exposure presages decline in liver cancer mortality in an endemic region of China. *Cancer prevention research* 2013;6(10):1038-45.
6. El-Serag HB. Epidemiology of hepatocellular carcinoma in USA. *Hepatology research : the official journal of the Japan Society of Hepatology* 2007;37 Suppl 2:S88-94.
7. National Cancer Institute. Cancer Trends Progress Report: 2011/2012 Update. 2012. (http://progressreport.cancer.gov/doc_detail.asp?pid=1&did=2009&chid=93&coid=920&mid#changing). Accessed January 1, 2013.
8. Altekruse SF, McGlynn KA, Reichman ME. Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009;27(9):1485-91.
9. Bolstad P. *GIS Fundamentals: A First Text on Geographic Information Systems*. 3rd ed. White Bear Lake, MN: Eider Press; 2008.
10. Tong S. Hepatitis B virus, a sex hormone-responsive virus. *Gastroenterology* 2012;142(4):696-9.
11. Centers for Disease Control and Prevention. Hepatocellular Carcinoma --- United States, 2001--2006. 2010. (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5917a3.htm). Accessed January 1, 2013.
12. National Cancer Institute (NCI). SEER Stat Fact Sheets: Liver and Intrahepatic Bile Duct Cancer. (http://seer.cancer.gov/statfacts/html/livibd.html). Accessed October 1, 2014.
13. Alter MJ. Prevention of spread of hepatitis C. *Hepatology* 2002;36(5 Suppl 1):S93-8.
14. Alter MJ, Kruszon-Moran D, Nainan OV, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *The New England journal of medicine* 1999;341(8):556-62.

15. Welzel TM, Graubard BI, Quraishi S, et al. Population-Attributable Fractions of Risk Factors for Hepatocellular Carcinoma in the United States. *The American journal of gastroenterology* 2013.
16. Ng J, Wu J. Hepatitis B- and hepatitis C-related hepatocellular carcinomas in the United States: similarities and differences. *Hepatitis monthly* 2012;12(10 HCC):e7635.
17. McGlynn KA, Quraishi, S., Welzel, T.A., Davila, J.A., El-Serag, H.B., Graubard, B.I. Attributable Risks for Hepatocellular Carcinoma in the United States. *American Association for Cancer Research*. Philadelphia, PA, 2010.
18. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *International journal of cancer Journal international du cancer* 2006;118(12):3030-44.
19. Yu MC, Yuan JM. Environmental factors and risk for hepatocellular carcinoma. *Gastroenterology* 2004;127(5 Suppl 1):S72-8.
20. El-Serag HB. Hepatocellular carcinoma. *The New England journal of medicine* 2011;365(12):1118-27.
21. Gomaa AI, Khan SA, Toledano MB, et al. Hepatocellular carcinoma: epidemiology, risk factors and pathogenesis. *World journal of gastroenterology : WJG* 2008;14(27):4300-8.
22. Environmental Protection Agency. Pesticides. 2012. (http://www.epa.gov/pesticides/). Accessed January 1, 2013.
23. Alavanja MC, Hoppin JA, Kamel F. Health effects of chronic pesticide exposure: cancer and neurotoxicity. *Annual review of public health* 2004;25:155-97.
24. Rull RP, Ritz B. Historical pesticide exposure in California using pesticide use reports and land-use surveys: an assessment of misclassification error and bias. *Environmental health perspectives* 2003;111(13):1582-9.
25. Franklin C, Worgan, John *Occupational and Residential Exposure Assessment for Pesticides*. Hoboken, NJ: Wiley; 2005.
26. Centers for Disease Control and Prevention. Fourth National Report on Human Exposure to Environmental Chemicals. 2009. (http://www.cdc.gov/exposurereport/pdf/FourthReport.pdf). Accessed January 1, 2013.
27. Dich J, Zahm SH, Hanberg A, et al. Pesticides and cancer. *Cancer causes & control : CCC* 1997;8(3):420-43.
28. Rossi L, Barbieri O, Sanguineti M, et al. Carcinogenicity study with technical-grade dichlorodiphenyltrichloroethane and 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene in hamsters. *Cancer research* 1983;43(2):776-81.
29. Turusov VS, Day NE, Tomatis L, et al. Tumors in CF-1 mice exposed for six consecutive generations to DDT. *Journal of the National Cancer Institute* 1973;51(3):983-97.
30. Figa-Talamanca I, Mearelli I, Valente P. Mortality in a cohort of pesticide applicators in an urban setting. *International journal of epidemiology* 1993;22(4):674-6.
31. Giordano F, Dell'Orco V, Giannandrea F, et al. Mortality in a cohort of pesticide applicators in an urban setting: sixty years of follow-up. *International journal of immunopathology and pharmacology* 2006;19(4 Suppl):61-5.
32. Cocco P, Kazerouni N, Zahm SH. Cancer mortality and environmental exposure to DDE in the United States. *Environmental health perspectives* 2000;108(1):1-4.
33. Brown DP. Mortality of workers employed at organochlorine pesticide manufacturing plants--an update. *Scandinavian journal of work, environment & health* 1992;18(3):155-61.

34. el-Zayadi AR, Badran HM, Barakat EM, et al. Hepatocellular carcinoma in Egypt: a single center study over a decade. *World journal of gastroenterology : WJG* 2005;11(33):5193-8.
35. Shibata A, Fukuda K, Toshima H, et al. The role of cigarette smoking and drinking in the development of liver cancer: 28 years of observations on male cohort members in a farming and fishing area. *Cancer detection and prevention* 1990;14(6):617-23.
36. Wang Y, Lewis-Michl EL, Hwang SA, et al. Cancer incidence among a cohort of female farm residents in New York State. *Archives of environmental health* 2002;57(6):561-7.
37. Cocco P, Blair A, Congia P, et al. Long-term health effects of the occupational exposure to DDT. A preliminary report. *Annals of the New York Academy of Sciences* 1997;837:246-56.
38. Cocco P, Fadda D, Billai B, et al. Cancer mortality among men occupationally exposed to dichlorodiphenyltrichloroethane. *Cancer research* 2005;65(20):9588-94.
39. Ferrand JF, Cenee S, Laurent-Puig P, et al. Hepatocellular carcinoma and occupation in men: a case-control study. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 2008;50(2):212-20.
40. Reif J, Pearce N, Fraser J. Cancer risks in New Zealand farmers. *International journal of epidemiology* 1989;18(4):768-74.
41. Kato I, Tominaga S, Ikari A. An epidemiological study on occupation and cancer risk. *Japanese journal of clinical oncology* 1990;20(2):121-7.
42. Keller JE, Howe HL. Case-control studies of cancer in Illinois farmers using data from the Illinois State Cancer Registry and the U.S. Census of Agriculture. *Eur J Cancer* 1994;30A(4):469-73.
43. Andersen A, Barlow L, Engeland A, et al. Work-related cancer in the Nordic countries. *Scandinavian journal of work, environment & health* 1999;25 Suppl 2:1-116.
44. Dossing M, Petersen KT, Vyberg M, et al. Liver cancer among employees in Denmark. *American journal of industrial medicine* 1997;32(3):248-54.
45. Kristensen P, Andersen A, Irgens LM, et al. Incidence and risk factors of cancer among men and women in Norwegian agriculture. *Scandinavian journal of work, environment & health* 1996;22(1):14-26.
46. Laakkonen A, Pukkala E. Cancer incidence among Finnish farmers, 1995-2005. *Scandinavian journal of work, environment & health* 2008;34(1):73-9.
47. Olsen JH, Jensen OM. Occupation and risk of cancer in Denmark. An analysis of 93,810 cancer cases, 1970-1979. *Scandinavian journal of work, environment & health* 1987;13 Suppl 1:1-91.
48. Pukkala E, Martinsen JI, Lynge E, et al. Occupation and cancer - follow-up of 15 million people in five Nordic countries. *Acta Oncol* 2009;48(5):646-790.
49. Ronco G, Costa G, Lynge E. Cancer risk among Danish and Italian farmers. *British journal of industrial medicine* 1992;49(4):220-5.
50. Wiklund K. Swedish agricultural workers. A group with a decreased risk of cancer. *Cancer* 1983;51(3):566-8.
51. Wiklund K, Dich J. Cancer risks among female farmers in Sweden. *Cancer causes & control : CCC* 1994;5(5):449-57.
52. Wiklund K, Dich J. Cancer risks among male farmers in Sweden. *Eur J Cancer Prev* 1995;4(1):81-90.

53. Wiklund K, Dich J, Holm LE, et al. Risk of cancer in pesticide applicators in Swedish agriculture. *British journal of industrial medicine* 1989;46(11):809-14.
54. Ezzat S, Abdel-Hamid M, Eissa SA, et al. Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt. *International journal of hygiene and environmental health* 2005;208(5):329-39.
55. McGlynn KA, Abnet CC, Zhang M, et al. Serum concentrations of 1,1,1-trichloro-2,2-bis(p-chlorophenyl)ethane (DDT) and 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (DDE) and risk of primary liver cancer. *Journal of the National Cancer Institute* 2006;98(14):1005-10.
56. Persson EC, Graubard BI, Evans AA, et al. Dichlorodiphenyltrichloroethane and risk of hepatocellular carcinoma. *International journal of cancer Journal international du cancer* 2012;131(9):2078-84.
57. Soliman AS, Hung CW, Tsodikov A, et al. Epidemiologic risk factors of hepatocellular carcinoma in a rural region of Egypt. *Hepatology international* 2010;4(4):681-90.
58. Zhao B, Shen H, Liu F, et al. Exposure to organochlorine pesticides is an independent risk factor of hepatocellular carcinoma: A case-control study. *Journal of exposure science & environmental epidemiology* 2011;21(6):601-8.
59. Austin H, Delzell E, Grufferman S, et al. Case-control study of hepatocellular carcinoma, occupation, and chemical exposures. *Journal of occupational medicine : official publication of the Industrial Medical Association* 1987;29(8):665-9.
60. Badawi AF, Michael MS. Risk factors for hepatocellular carcinoma in Egypt: the role of hepatitis-B viral infection and schistosomiasis. *Anticancer research* 1999;19(5C):4565-9.
61. Brownson RC, Reif JS, Chang JC, et al. Cancer risks among Missouri farmers. *Cancer* 1989;64(11):2381-6.
62. Cordier S, Le TB, Verger P, et al. Viral infections and chemical exposures as risk factors for hepatocellular carcinoma in Vietnam. *International journal of cancer Journal international du cancer* 1993;55(2):196-201.
63. Hardell L, Bengtsson NO, Jonsson U, et al. Aetiological aspects on primary liver cancer with special regard to alcohol, organic solvents and acute intermittent porphyria--an epidemiological investigation. *British journal of cancer* 1984;50(3):389-97.
64. Heinemann K, Willich SN, Heinemann LA, et al. Occupational exposure and liver cancer in women: results of the Multicentre International Liver Tumour Study (MILTS). *Occup Med (Lond)* 2000;50(6):422-9.
65. Kauppinen T, Riala R, Seitsamo J, et al. Primary liver cancer and occupational exposure. *Scandinavian journal of work, environment & health* 1992;18(1):18-25.
66. London WT, Evans AA, McGlynn K, et al. Viral, host and environmental risk factors for hepatocellular carcinoma: a prospective study in Haimen City, China. *Intervirology* 1995;38(3-4):155-61.
67. Porru S, Placidi D, Carta A, et al. Primary liver cancer and occupation in men: a case-control study in a high-incidence area in Northern Italy. *International journal of cancer Journal international du cancer* 2001;94(6):878-83.
68. Stemhagen A, Slade J, Altman R, et al. Occupational risk factors and liver cancer. A retrospective case-control study of primary liver cancer in New Jersey. *American journal of epidemiology* 1983;117(4):443-54.
69. Suarez L, Weiss NS, Martin J. Primary liver cancer death and occupation in Texas. *American journal of industrial medicine* 1989;15(2):167-75.

70. Evans AA, Chen G, Ross EA, et al. Eight-year follow-up of the 90,000-person Haimen City cohort: I. Hepatocellular carcinoma mortality, risk factors, and gender differences. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2002;11(4):369-76.
71. Blair A, Zahm SH. Agricultural exposures and cancer. *Environmental health perspectives* 1995;103 Suppl 8:205-8.
72. Austin H, Delzell E, Grufferman S, et al. A case-control study of hepatocellular carcinoma and the hepatitis B virus, cigarette smoking, and alcohol consumption. *Cancer research* 1986;46(2):962-6.
73. Cockburn M, Mills P, Zhang X, et al. Prostate cancer and ambient pesticide exposure in agriculturally intensive areas in California. *American journal of epidemiology* 2011;173(11):1280-8.
74. Environmental Protection Agency. Types of Pesticides. 2012. (http://www.epa.gov/opp00001/about/types.htm#chemical). Accessed January 1, 2013.
75. Environmental Protection Agency. Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates. 2011.
76. Nasca PC, Pastides, H. *Fundamentals of Cancer Epidemiology*. Second ed. Sudbury, MA: Jones and Bartlett Publishers; 2008.
77. Percy C, Stanek E, 3rd, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American journal of public health* 1981;71(3):242-50.
78. Flanders WD. Inaccuracies of death certificate information. *Epidemiology* 1992;3(1):3-5.
79. Szklo M, and F.J. Nieto. *Epidemiology Beyond the Basics*. Second ed.: Jones and Bartlett Publishers, LLC; 2007.
80. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins,, 2008:x, 758 p. ill., map 27 cm.
81. Maxwell SK, Meliker JR, Goovaerts P. Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research. *Journal of exposure science & environmental epidemiology* 2010;20(2):176-85.
82. Ritz B, Rull RP. Assessment of environmental exposures from agricultural pesticides in childhood leukaemia studies: challenges and opportunities. *Radiation protection dosimetry* 2008;132(2):148-55.
83. Bolstad P. *GIS Fundamentals: A First Text on Geographic Information Systems*. Third ed. White Bear Lake, Minnesota: Eider Press; 2008.
84. Auchincloss AH, Gebreab SY, Mair C, et al. A review of spatial methods in epidemiology, 2000-2010. *Annual review of public health* 2012;33:107-22.
85. Blair A, Zahm, S.H., Cantor, K.P., Stewart, P.A. Estimating Exposure to Pesticides in Epidemiological Studies of Cancer. *Biological Monitoring for Pesticide Exposure: Measurement, Estimation, and Risk Reduction*: American Chemical Society, 1988:38-46.
86. Ward MH, Nuckols JR, Weigel SJ, et al. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. *Environmental health perspectives* 2000;108(1):5-12.
87. EPA (Environmental Protection Agency). Pesticide issues in the works: pesticide volatilization. 2009. (http://www.epa.gov/pesticides/about/intheworks/volatilization.htm). Accessed 1 Aug 2013.

88. Gunier RB, Ward MH, Airola M, et al. Determinants of agricultural pesticide concentrations in carpet dust. *Environmental health perspectives* 2011;119(7):970-6.
89. Gunier RB, Harnly ME, Reynolds P, et al. Agricultural pesticide use in California: pesticide prioritization, use densities, and population distributions for a childhood cancer study. *Environmental health perspectives* 2001;109(10):1071-8.
90. Franklin C, Worgan J. *Occupational and Residential Exposure Assessment for Pesticides*. Hoboken, NJ: Wiley; 2005.
91. CDPR (California Department of Pesticide Regulation). Pesticide Use Reporting (PUR). 2014. (http://www.cdpr.ca.gov/docs/pur/purmain.htm). Accessed 1 Aug 2013.
92. Rull RP, Ritz B, Shaw GM. Validation of self-reported proximity to agricultural crops in a case-control study of neural tube defects. *Journal of exposure science & environmental epidemiology* 2006;16(2):147-55.
93. Goldberg DW, Zhang X, Marusek JC, et al. Development of an automated pesticide exposure analyst for California's central valley. Presented at Proc Urban Regional Info Syst Assoc GIS Public Health Conf, New Orleans2007.
94. Nuckols JR, Gunier RB, Riggs P, et al. Linkage of the California Pesticide Use Reporting Database with spatial land use data for exposure assessment. *Environmental health perspectives* 2007;115(5):684-9.
95. USGS (U.S. Geological Survey). Landsat: A Global Land-Imaging Mission. 2014. (http://landsat.usgs.gov/). Accessed 1 Feb 2014.
96. Maxwell SK, Airola M, Nuckols JR. Using Landsat satellite data to support pesticide exposure assessment in California. *International journal of health geographics* 2010;9:46.
97. Maxwell SK. Downscaling Pesticide Use Data to the Crop Field Level in California Using Landsat Satellite Imagery: Paraquat Case Study. *Remote Sensing* 2011;3(9):1805-16.
98. Campbell JB, Wynne RH. *Introduction to Remote Sensing*. Fifth ed. New York, NY: The Guilford Press; 2011.
99. Guerschman J, Paruelo J, Bella CD, et al. Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data. *International Journal of Remote Sensing* 2003;24(17):3381-402.
100. De Wit A, Clevers J. Efficiency and accuracy of per-field classification for operational crop mapping. *International journal of remote sensing* 2004;25(20):4091-112.
101. Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing* 2007;28(5):823-70.
102. Turker M, Ozdarici A. Field-based crop classification using SPOT4, SPOT5, IKONOS and QuickBird imagery for agricultural areas: a comparison study. *International Journal of Remote Sensing* 2011;32(24):9735-68.
103. USDA (U.S. Department of Agriculture). The Central Valley Region. 2003. (www.fs.fed.us/.../psw_gtr187_05.pdf). Accessed 1 January 2014.
104. USDA (U.S. Department of Agriculture). State Fact Sheets: California. 2014. (http://www.ers.usda.gov/data-products/state-fact-sheets/state-data.aspx?StateFIPS=06&StateName=California). Accessed 1 Jan 2014.
105. U.S. Census Bureau. Census 2010 Summary File 1; using American FactFinder. 2010. (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml). Accessed 1 Feb 2014.

106. U.S. Census Bureau. State & County QuickFacts: Kern County, California 2014. (http://quickfacts.census.gov/qfd/states/06/06029.html). Accessed 1 Jan 2014.
107. CDPR (California Department of Pesticide Regulation). Pesticide Use Reporting. 2011. (http://www.cdpr.ca.gov/docs/pur/pur11rep/lbsby_co_11.pdf). Accessed 1 Aug 2013.
108. USDA (U.S. Department of Agriculture). Census of Agriculture. 2014. (http://www.agcensus.usda.gov/index.php). Accessed 1 Jan 2014.
109. County of Kern. Kern County Crop Reports. 2014. (http://www.kernag.com/caap/crop-reports/crop-reports.asp). Accessed 1 Jan 2014.
110. FAO (Food and Agriculture Organization of the United Nations). Glossary. 2013. (http://faostat.fao.org/site/375/default.aspx). Accessed 2014.
111. County of Kern. Kern County, California Planting and Harvesting Dates. 2014. (http://www.kernag.com/dept/stats/crop-stats.asp). Accessed 1 Jan 2014.
112. National Atlas. The Public Land Survey System (PLSS). 2014. (http://www.nationalatlas.gov/articles/boundaries/a_plss.html). Accessed 1 Apr 2013.
113. CDWR (California Department of Water Resources). Land Use Survey Overview. 2014. (http://www.water.ca.gov/landwateruse/lusrvymain.cfm). Accessed 1 Aug 2013.
114. Kern County Assessor. Use Codes. 2012. (http://recorder.co.kern.ca.us/assessor/FYI.php?page=usecodes). Accessed 1 Apr 2013.
115. U.S. Census Bureau. TIGER Products. 2014. (http://www.census.gov/geo/maps-data/data/tiger.html). Accessed 1 Jan 2014.
116. CDPR (California Department of Pesticide Regulation). Appendix C: California's Pesticide Use Report An Assessment of Spatial Data Quality. 2000. (http://www.cdpr.ca.gov/docs/pur/appendix_c_dataq_ldr.pdf). Accessed 1 Aug 2013.
117. VoPham T. *Integrating Landsat and California Pesticide Exposure Estimation at Aggregated Analysis Scales: Accuracy Assessment of Rurality*. Los Angeles, CA: USC (University of Southern California); 2014.
118. CDPR (California Department of Pesticide Regulation). DPR Pesticide Use Reporting: An Overview of California's Unique Full Reporting System. 2000. (http://www.cdpr.ca.gov/docs/pur/purovrvw/ovr52000.pdf). Accessed 1 Aug 2013.
119. AgroPages. Crop Protection Database. 2014. (http://www.agropages.com/AgroData/). Accessed 1 Aug 2013.
120. Greene SA, Pohanish RP. *Sittig's handbook of pesticides and agricultural chemicals*. Norwich, N.Y.: William Andrew Pub.; 2005.
121. Rull RP, Gunier R, Von Behren J, et al. Residential proximity to agricultural pesticide applications and childhood acute lymphoblastic leukemia. *Environmental research* 2009;109(7):891-9.
122. Rull RP, Ritz B, Shaw GM. Neural tube defects and maternal residential proximity to agricultural pesticide applications. *American journal of epidemiology* 2006;163(8):743-53.
123. Wood A. Compendium of Pesticide Common Names. 2010. (http://www.alanwood.net/pesticides/class_pesticides.html). Accessed 1 Aug 2013.
124. Wells J. Pesticide Use Trends in California Agriculture. 2011. (http://www.foodandfarming.info/docs.asp?id=76). Accessed 1 Feb 2014.
125. Clark Labs. IDRISI Selva. Worcester, Massachusetts, United States, 2014.

126. Chander G, Markham BL, Helder DL. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote sensing of environment* 2009;113(5):893-903.
127. Chavez PS. Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing* 1996;62(9):1025-35.
128. Song C, Woodcock CE, Seto KC, et al. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote sensing of environment* 2001;75(2):230-44.
129. YCEO (Yale Center for Earth Observation). Yale Guide to Landsat 8 Image Processing. 2013. (http://www.yale.edu/ceo/Documentation/Landsat%208%20image%20processing.pdf). Accessed 2013.
130. Mather PM, Koch M. Chater 7: Filtering Techniques. *Computer Processing of Remotely-Sensed Images: An Introduction, 4th Edition*. Oxford, UK: Wiley Blackwell, 2011.
131. Vassiliou A, Boulianne M, Blais J. On the application of averaging median filters in remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on* 1988;26(6):832-8.
132. USGS (U.S. Geological Survey). NDVI, the Foundation for Remote Sensing Phenology 2011. (http://phenology.cr.usgs.gov/ndvi_foundation.php). Accessed 2013.
133. CDWR (California Department of Water Resources). Metadata for the 2009 Eastern Fresno County Land Use Survey Data. 2009. (http://www.water.ca.gov/landwateruse/docs/landusedata/shapes/09eFR.zip). Accessed 1 Aug 2013.
134. Richards JA. *Remote Sensing Digital Image Analysis: An Introduction*. 5th ed. New York: Springer; 2013.
135. Esri. ArcGIS 10.2 Help. 2014. (http://resources.arcgis.com/en/help/). Accessed 1 Aug 2013.
136. Van Niel TG, McVicar TR. Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. *Computers and Electronics in Agriculture* 2004;45(1):91-108.
137. Simoniello T, Carone M, Lanfredi M, et al. Landscape-scale characterization of vegetation phenology using AVHRR-NDVI and Landsat-TM data. Presented at Remote Sensing2004.
138. Suzuki R, Nomaki T, Yasunari T. Spatial distribution and its seasonality of satellite-derived vegetation index (NDVI) and climate in Siberia. *International Journal of Climatology* 2001;21(11):1321-35.
139. U.S. Census Bureau. 2000 Census Urban and Rural Classification and Urban Area Criteria. 2000. (http://www.census.gov/geo/www/ua/ua_2k.html). Accessed 2013.
140. Martinuzzi S, Gould WA, Ramos González OM. Creating Cloud-Free Landsat ETM+ Data Sets in Tropical Landscapes: Cloud and Cloud-Shadow Removal. 2007. (http://www.fs.fed.us/global/iitf/pubs/iitf-gtr32.pdf). Accessed 1 Aug 2013.
141. Esri. ArcGIS 10.1. Redlands, California, United States, 2014.
142. SAS. SAS. Cary, North Carolina, United States, 2014.
143. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *biometrics* 1977:159-74.

144. Szklo M, Nieto FJ. *Epidemiology Beyond the Basics*. Second ed.: Jones and Bartlett Publishers, LLC; 2007.
145. Chen X, Li BL, Allen MF. Characterizing urbanization, and agricultural and conservation land-use change in Riverside County, California, USA. *Annals of the New York Academy of Sciences* 2010;1195(s1):E164-E76.
146. Maxwell S, Meliker J, Goovaerts P. Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research. *Journal of exposure science & environmental epidemiology* 2010;20(2):176-85.
147. Godfrey L, Rosenheim J, Goodell P. Cotton aphid emerges as major pest in SJV cotton. *California Agriculture* 2000;54(6):26-9.
148. PAN (Pesticide Action Network). PAN Pesticides Database - California Pesticide Use. 2013. (http://www.pesticideinfo.org/Search_Use.jsp). Accessed 1 Aug 2013.
149. Gliessman SR. Multiple cropping systems: A basis for developing an alternative agriculture. In: Assessment OoT, ed. *Innovative Technologies for Lesser Developed Countries*. Washington D.C.: U.S. Congress, 1985:69-86.
150. Sullivan P. Intercropping Principles and Production Practices. Appropriate Technology Transfer for Rural Areas (ATTRA); 2003. (https://attra.ncat.org/publication.html). Accessed.
151. Liebman M, Dyck E. Crop rotation and intercropping strategies for weed management. *Ecological Applications* 1993:92-122.
152. USDA (U.S. Department of Agriculture). Soil Tillage and Crop Rotation. 2013. (http://www.ers.usda.gov/topics/farm-practices-management/crop-livestock-practices/soil-tillage-and-crop-rotation.aspx). Accessed 1 Jan 2014.
153. UC IPM (University of California Statewide Integrated Pest Management Program). Alfalfa: Crop Rotation. 2006. (http://www.ipm.ucdavis.edu/PMG/r1900811.html). Accessed 1 Jan 2013.
154. UC IPM (University of California Statewide Integrated Pest Management Program). Potato: Crop Rotation. 2008. (http://www.ipm.ucdavis.edu/PMG/r607900511.html). Accessed 1 Jan 2013.
155. Yu Q, Gong P, Clinton N, et al. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing* 2006;72(7):799.
156. Hill R. Image segmentation for humid tropical forest classification in Landsat TM data. *International Journal of remote sensing* 1999;20(5):1039-44.
157. AMIRA. The almond tree and its fruit. 2007. (http://www.amira.ca/en/info_almond.htm#almondReferences). Accessed 1 Feb 2014.
158. Conrad C, Fritsch S, Zeidler J, et al. Per-field irrigated crop classification in arid Central Asia using SPOT and ASTER data. *Remote Sensing* 2010;2(4):1035-56.
159. NIH (National Institutes of Health). Sample Surveys. Office of Behavioral & Social Sciences Research; 2014. (http://www.esourceresearch.org/tabid/374/default.aspx). Accessed 1 Feb 2014.
160. Forkel M, Carvalhais N, Verbesselt J, et al. Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology. *Remote Sensing* 2013;5(5):2113-44.

161. Maxwell SK, Sylvester KM. Identification of “ever-cropped” land (1984-2010) using Landsat annual maximum NDVI image composites: Southwestern Kansas case study. *Remote sensing of environment* 2012;121:186-95.
162. Bin W, Muramatsu K, Fujiwara N. Automated detection and removal of clouds and their shadows from Landsat TM images. *IEICE Transactions on information and systems* 1999;82(2):453-60.
163. USDA FSA Aerial Photography Field Office (U.S. Department of Agriculture Farm Service Agency Aerial Photography Field Office). Kern County compressed county mosaic (CCM). Salt Lake City, Utah: National Agriculture Imagery Program (NAIP) via Cal-Atlas Geospatial Clearinghouse, 2012.
164. Waller LA, Gotway CA. *Applied spatial statistics for public health data*. John Wiley & Sons; 2004.
165. Montello DR. Scale in Geography. In: Smelser NJ, Baltes, P.B., ed. *International Encyclopedia of the Social & Behavioral Sciences*: Pergamon Press, 2001:13501-4.
166. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environmental health perspectives* 2004;112(9):1007-15.
167. Ritz B, Costello S. Geographic model and biomarker-derived measures of pesticide exposure and Parkinson's disease. *Ann N Y Acad Sci* 2006;1076:378-87.
168. Boscoe FP, Ward MH, Reynolds P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *International journal of health geographics* 2004;3(1):28.
169. Gotway CA, Young LJ. Combining incompatible spatial data. *Journal of the American Statistical Association* 2002;97(458):632-48.
170. Bian L, Butler R. Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric Engineering and Remote Sensing* 1999;65:73-84.
171. Clary T, Ritz B. Pancreatic cancer mortality and organochlorine pesticide exposure in California, 1989-1996. *American journal of industrial medicine* 2003;43(3):306-13.
172. Economic Research Service (ERS). Agricultural Productivity in the U.S.: Overview. 2014. (http://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us.aspx). Accessed 1 May 2014.
173. State of California. Facts. 2012. (http://www.ca.gov/About/Facts.html). Accessed 30 May 2014.
174. U.S. Census Bureau. State & County QuickFacts: California. 2014. (http://quickfacts.census.gov/qfd/states/06000.html). Accessed 1 July 2014.
175. U.S. Census Bureau. History: 2000 Census of Population and Housing (Volume 1). Washington D.C.: U.S. Government Printing Office; 2009. (http://www.census.gov/history/pdf/Census2000v1.pdf). Accessed 1 May 2014.
176. U.S. Census Bureau. Census 2000 Summary File 1 Technical Documentation. 2001. (http://www.census.gov/prod/cen2000/doc/sf1.pdf). Accessed 1 January 2014.
177. U.S. Census Bureau. What are census blocks? ; 2011. (http://blogs.census.gov/2011/07/20/what-are-census-blocks/). Accessed 2014.
178. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International journal of health geographics* 2006;5:58.

179. U.S. Census Bureau. ZIP Code™ Tabulation Areas (ZCTAs™). 2014. (https://www.census.gov/geo/reference/zctas.html). Accessed 2014.
180. U.S. Census Bureau. Census 2000 TIGER/Line Files Technical Documentation. Washington D.C.; 2000. (http://www.census.gov/geo/maps-data/data/pdfs/tiger/tiger2k/tiger2k.pdf). Accessed 1 January 2014.
181. Vogelmann JE, Howard SM, Yang L, et al. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing* 2001;67(6).
182. MRLC (Multi-Resolution Land Characteristics Consortium). National Land Cover Dataset 1992 (NLCD1992). 2014. (http://www.mrlc.gov/nlcd1992.php). Accessed 1 May 2014.
183. GreenInfo Network. California Protected Areas Data Portal. 2014. (http://www.calands.org/). Accessed 1 May 2014.
184. Lim E, Pon A, Djoumbou Y, et al. T3DB: a comprehensively annotated database of common toxins and their targets. *Nucleic acids research* 2010;38(suppl 1):D781-D6.
185. Bell EM, Hertz-Picciotto I, Beaumont JJ. A case-control study of pesticides and fetal death due to congenital anomalies. *Epidemiology* 2001;12(2):148-56.
186. USGS (U.S. Geological Survey). Dasymetric Mapping: An Alternative Approach To Visually And Statistically Enhancing Population Density. 2014. (http://geography.wr.usgs.gov/science/dasymetric/index.htm). Accessed 1 May 2014.
187. Eicher CL, Brewer CA. Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science* 2001;28(2):125-38.
188. Kang C, Qaqish B, Monaco J, et al. Kappa statistic for clustered dichotomous responses from physicians and patients. *Statistics in medicine* 2013;32(21):3700-19.
189. Rural Health Research Center. RUCA Data. 2006. (http://depts.washington.edu/uwruca/ruca-download.php). Accessed 1 May 2014.
190. NPR (National Public Radio). California's Central Valley. 2002. (http://www.npr.org/programs/atc/features/2002/nov/central_valley/). Accessed 1 January 2014.
191. McGee S. Simplifying likelihood ratios. *Journal of general internal medicine* 2002;17(8):647-50.
192. Sleeter R. Dasymetric mapping techniques for the San Francisco Bay region, California. Presented at Urban and regional information systems association annual conference proceedings2004.
193. Matisziw TC, Grubesic TH, Wei H. Downscaling spatial structure for the analysis of epidemiological data. *Computers, Environment and Urban Systems* 2008;32(1):81-93.
194. Jacquez GM. Current practices in the spatial analysis of cancer: flies in the ointment. *International journal of health geographics* 2004;3(1):22.
195. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of epidemiology and community health* 2001;55(2):111-22.
196. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *Bmj* 2004;329(7458):168-9.
197. O'Sullivan DO, Unwin DJ. *Geographic Information Analysis*. 2nd ed.: John Wiley & Sons; 2010.

198. County of Santa Clara. Guideline for Policies Governing the Exchange of an Existing Williamson Act Contract for an Open Space Easement. 2011. (http://www.sccgov.org/sites/planning/PlansPrograms/Williamson/Documents/Guideline-for-Exchange-for-OSE-Approved-10-18-11.pdf). Accessed 1 May 2014.
199. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. (http://globocan.iarc.fr). Accessed October 1, 2014.
200. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA: a cancer journal for clinicians* 2011;61(2):69-90.
201. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2013 Sub (2000-2011) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969-2012 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2014 (updated 5/7/2014), based on the November 2013 submission.
202. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2013 Sub (1973-2011 varying) - Linked To County Attributes - Total U.S., 1969-2012 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2014 (updated 5/7/2014), based on the November 2013 submission.
203. National Cancer Institute (NCI). PDQ® Adult Primary Liver Cancer Treatment. Bethesda, MD: National Cancer Institute. (http://www.cancer.gov/cancertopics/pdq/treatment/adult-primary-liver/HealthProfessional). Accessed October 1 2014.
204. Environmental Protection Agency (EPA). Pesticides. (http://www.epa.gov/pesticides/). Accessed October 28, 2014.
205. U.S. Department of Agriculture. Agricultural Productivity in the U.S. (http://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us.aspx). Accessed October 1, 2014.
206. Engels EA, Pfeiffer RM, Ricker W, et al. Use of surveillance, epidemiology, and end results-medicare data to conduct case-control studies of cancer among the US elderly. *American journal of epidemiology* 2011;174(7):860-70.
207. National Cancer Institute (NCI). SEER-Medicare Linked Database. (http://appliedresearch.cancer.gov/seermedicare/). Accessed October 28, 2014.
208. The Kaiser Family Foundation's State Health Facts. Data Source: Medicare Enrollment Reports MSCPF. Medicare Prescription Drug Plans: Stand Alone PDP Enrollment. 2007. Accessed October 28, 2014.
209. National Cancer Institute (NCI). Surveillance, Epidemiology, and End Results Program. (http://seer.cancer.gov/). Accessed October 1, 2014.
210. Fritz A, Percy C, Shanmugaratnam K, et al. *International Classification of Diseases for Oncology: ICD-O*. 3rd ed. Geneva, Switzerland: World Health Organization; 2000.
211. Davila JA, Morgan RO, Shaib Y, et al. Diabetes increases the risk of hepatocellular carcinoma in the United States: a population based case control study. *Gut* 2005;54(4):533-9.

212. Esri. ArcGIS for Desktop: Release 10.1. Redlands, CA: Esri, 2012.
213. California Department of Pesticide Regulation (CDPR). Pesticide Use Reporting (PUR). (http://www.cdpr.ca.gov/docs/pur/purmain.htm). Accessed October 1, 2014.
214. Reynolds P, Hurley SE, Goldberg DE, et al. Residential proximity to agricultural pesticide use and incidence of breast cancer in the California Teachers Study cohort. *Environmental research* 2004;96(2):206-18.
215. Census 2000 TIGER/Line® Files [machine-readable data files]/prepared by the U.S. Census Bureau-Washington DC. 2000.
216. Welzel TM, Graubard BI, Zeuzem S, et al. Metabolic syndrome increases the risk of primary liver cancer in the United States: a study in the SEER-Medicare database. *Hepatology* 2011;54(2):463-71.
217. Centers for Medicare & Medicaid Services (CMS). Medicare Savings Programs. (http://www.medicare.gov/your-medicare-costs/help-paying-costs/medicare-savings-program/medicare-savings-programs.html). Accessed October 1, 2014.
218. Census 2000 Summary File 3—[California]/prepared by U.S. Census Bureau. 2002.
219. University of Washington Rural Health Research Center. RUCA ZIP Version 2.0 Codes. 2006. http://depts.washington.edu/uwruca/index.php. Accessed October 1, 2014.
220. Warren J, Smalley KB. *Rural Public Health: Best Practices and Preventive Models*. Springer Publishing Company; 2014.
221. Environmental Protection Agency (EPA). Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates. 2011.
222. Wells J. Pesticide Use Trends in California Agriculture. Environmental Solutions Group, 2011.
223. Environmental Protection Agency (EPA). Recognition and Management of Pesticide Poisonings. 2013.
224. Environmental Protection Agency (EPA). Pesticides: Reregistration, Agreement to Terminate All Uses of Aldicarb. (http://www.epa.gov/oppsrrd1/REDs/factsheets/aldicarb_fs.html). Accessed October 1, 2014.
225. Centers for Disease Control and Prevention (CDC). Fourth National Report on Human Exposure to Environmental Chemicals. 2009.
226. Longnecker MP, Rogan WJ, Lucier G. The human health effects of DDT (dichlorodiphenyltrichloroethane) and PCBS (polychlorinated biphenyls) and an overview of organochlorines in public health. *Annual review of public health* 1997;18:211-44.
227. Franklin Claire, Worgan John. *Occupational and Residential Exposure Assessment for Pesticides*. Hoboken, NJ: Wiley; 2005.
228. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
229. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Medical care* 2002;40(8):IV-3-IV-18.
230. Social Security Administration. Disability Evaluation Under Social Security. 2014. (http://www.ssa.gov/disability/professionals/bluebook/). Accessed November 30, 2014.
231. Gilden RC, Huffling K, Sattler B. Pesticides and health risks. *Journal of obstetric, gynecologic, and neonatal nursing : JOGNN / NAACOG* 2010;39(1):103-10.

232. Ma X, Buffler PA, Gunier RB, et al. Critical windows of exposure to household pesticides and risk of childhood leukemia. *Environmental health perspectives* 2002;110(9):955-60.
233. Marcus EL, Tur-Kaspa R. Chronic hepatitis C virus infection in older adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2005;41(11):1606-12.
234. Ott JJ, Stevens GA, Groeger J, et al. Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 2012;30(12):2212-9.
235. Fotheringham AS, Brunsdon C, Charlton M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, England: John Wiley & Sons, LTD; 2002.
236. Nakaya T, Charlton M, Lewis P, et al. GWR4. Kyoto, Japan, 2014.