The role of socially patterned risk factors in childhood cancer incidence and survival

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Rebecca Diane Kehm

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Theresa L. Osypuk, Logan G. Spector

July 2017

Acknowledgements

I would like to acknowledge the support and contributions provided by each member of my dissertation committee. My advisor, Dr. Theresa Osypuk, provided me with invaluable mentorship and training throughout my four years in the PhD program. My co-advisor, Dr. Logan Spector, continuously offered his time and knowledge as I developed into a competent cancer epidemiologist. Dr. Jenny Poynter and Dr. David Vock both contributed expertise and guidance throughout the dissertation process, substantially strengthening my research.

I would also like to acknowledge collaboration with the National Center for Health Statistics, and Joyce Arbertha specifically, which made *manuscript 1* feasible. I would like to acknowledge collaboration with the Minnesota Department of Health, which made *manuscript 2* feasible. Specifically, I would like to acknowledge support from Margee Brown and Dr. Sally Bushhouse at the Minnesota Cancer Surveillance System, and Judy Palermo at the Minnesota Center for Health Statistics. Further, I would like to acknowledge collaboration with the National Cancer Institute, and Dr. Sean Altekruse specifically, which made *manuscript 3* feasible.

Finally, I would like to acknowledge the financial support I received from the National Institutes of Health Translational Pediatric Cancer Epidemiology Training Grant (T32CA099936) and from the J.B. Hawley Student Research Award.

Dedication

This dissertation is dedicated to my parents who taught me the importance and privilege of an education, and who provided me with unwavering support to pursue one; and to Nick, my partner in everything.

Abstract

A primary goal of childhood cancer research is to understand the role of non-genetic, socially patterned, risk factors in incidence and survival. Many unanswered questions persist in the literature that, if answered, may help address this aim. Such questions include: do pregnancy-related exposures contribute to the rise in childhood cancer incidence over time; is socioeconomic status (SES) associated with childhood cancer incidence; and what is the underlying role of social versus biological factors in explaining racial disparities in childhood cancer survival? This dissertation addressed each of these questions by leveraging population-based data from multiple existing data sources, and by employing advanced statistical methods, in three separate investigations.

In *manuscript 1*, we conducted a time series ecologic analysis at the county-level to test the hypothesis that the temporal rise in childhood cancer incidence is due to secular trends in established pregnancy-related risk factors including older maternal age, higher birthweight, and smaller family size. We linked population-based cancer registry data from the Surveillance, Epidemiology, and End Results (SEER) 9 database (1975-2013) to natality data from the National Center for Health Statistics (1970-2013). We compared the crude average annual percent change (AAPC) in incidence of combined (all diagnoses) and individual cancers among children, ages 0-4 years, to AAPCs adjusted for pregnancy-related and sociodemographic (race/ethnicity and poverty) factors. AAPCs were estimated from Poisson mixed models. In crude models, we found a statistically significant temporal rise in incidence of combined childhood cancers (AAPC, 0.71%; 95% CI, 0.55, 0.86), ALL (0.78%; 0.49, 1.07), AML (1.86%; 1.13, 2.59), CNS tumors

(1.31%; 0.94, 1.67), and hepatoblastoma (2.70%; 1.68, 3.72). Contrary to our hypothesis, AAPCs remained statistically significantly above 0% in models fully adjusted for county-level characteristics, though AAPCs were attenuated towards the null for AML (1.62%; 0.38, 2.87) and hepatoblastoma (2.36%; 0.71, 4.04). Therefore, we did not find conclusive evidence that secular trends in established pregnancy-related risk factors account for the temporal rise in cancer incidence rates among children, ages 0-4 years.

In *manuscript 2*, we tested whether SES, measured at multiple levels of exposure, is associated with incidence of childhood cancers after accounting for established demographic and pregnancy-related risk factors. We conducted a population-based case-cohort study using the Minnesota birth registry, 1989-2014, as the source cohort. Cases, ages 0-14 years, were identified in the Minnesota Cancer Surveillance System and linked to birth records through probabilistic record linkage. Controls were 4:1 frequency matched on birth year (2,947 cases, 11,907 controls). We measured individual-level SES using maternal education, and we measured neighborhood-level SES using a census tract composite index. Associations between SES and childhood cancer incidence were tested with logistic mixed models. In crude models, we found that higher maternal education was adversely associated with incidence of combined childhood cancers (OR, 1.08; 95% CI, 1.04, 1.13), ALL (OR, 1.10; 95% CI, 1.02, 1.19), CNS tumors (OR, 1.12; 95% CI, 1.04, 1.21), and neuroblastoma (OR, 1.15; 95% CI, 1.02, 1.30). These associations were attenuated towards the null, and no longer statistically significant, after adjusting for established demographic and pregnancy-related risk factors. Similar patterns were observed for neighborhood-level SES. A protective association, robust to covariate

control, was detected between higher maternal education and hepatoblastoma risk (adjusted OR, 0.70; 95% CI, 0.51, 0.98). Overall, results suggest that, unlike for some adult cancers that show strong socioeconomic gradients, associations between SES and many childhood cancers appear to be explained by established demographic and pregnancy-related risk factors (i.e. non-Hispanic white race/ethnicity, older maternal age, and higher birthweight).

In *manuscript 3*, we tested whether SES contributes to (i.e. mediates) black-white racial disparities in childhood cancer survival. We used population-based survival data from the SEER 18 database for black and white children, ages 0-19 years, diagnosed 2000-2011 (N=27,741). Race was recorded in SEER through medical record abstraction. We measured SES using a validated census tract composite index, and we tested treatment (first-course cancer-directed surgery and radiation) and distal stage at diagnosis as secondary potential mediators. We used the inverse odds weighting (IOW) method to test for mediation among combined and individual childhood cancers. Results showed that whites have a significant survival advantage over blacks for combined childhood cancers, leukemias, lymphomas, CNS tumors, neuroblastoma, and non-rhabdomyosarcoma soft tissue sarcomas (NRSTS). Significant black-white mortality hazard ratios ranged from 1.44 (95% CI, 1.11, 1.87) for NRSTS to 1.91 (95% CI, 1.52, 2.41) for astrocytomas. SES significantly mediated the race-survival association for combined childhood cancers and leukemias, accounting for 20% of the disparity for combined childhood cancers (indirect effect HR (iHR), 1.12; 95% CI, 1.08, 1.15), 44% for ALL (iHR, 1.18; 95% CI, 1.08, 1.29), and 28% for AML (iHR, 1.16; 95% CI, 1.04,

1.29). Treatment and stage significantly mediated race-survival associations for astrocytomas (27%; iHR, 1.19; 95% CI, 1.01, 1.40) and neuroblastoma (32%, iHR, 1.16; 95% CI, 1.04, 1.30). Overall, we found evidence that SES contributes to black-white racial disparities in childhood cancer survival, particularly for childhood leukemias. However, we could not rule out the possibility that non-social factors also contribute to survival differences by race, particularly for astrocytomas and neuroblastoma.

Taken together, findings from these three investigations suggest that while continued study of socioeconomic exposures may generate limited new insight into childhood cancer etiology, this line of research may be fruitful for understanding and, ultimately, addressing differences in childhood cancer outcomes.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Descriptive Epidemiology

Although rare, cancer is the leading cause of death by disease past infancy among children in the United States.[1] An estimated 10,270 children, under the age of 15 years, will be diagnosed with cancer in 2017 and an estimated 1,190 will die from the disease.[2] Childhood cancer incidence peaks in infancy at about 240 cases per million per year.[3,4] Incidence drops to a low of 128 cases per million per year between ages 5 and 9 years,[3,4] before rising to about 220 cases per million per year between ages 15 and 19 years.[3,4] Childhood cancer is not a single disease entity, but rather a spectrum of different malignancies that vary by histology, epidemiology, site of disease origin, and etiology.[5]

Classifications of Childhood Cancer

The types of cancers that develop in children often differ from those in adults. Unlike adult cancers, which are generally classified by the primary site of tumor origin, childhood cancers are more appropriately classified by morphology.[5] Leukemias are the most common type of cancer among children, accounting for approximately 30% of pediatric malignancies.[5] Leukemias arise in the hematopoietic system, most often involving malignant transformation of lymphoid progenitor cells or, to a lesser extent, myeloid progenitor cells.[5] Central nervous system (CNS) tumors, a heterogeneous group of neoplasms that originate in the brain, are the second most common type of cancer among children, accounting for approximately 17% of malignancies.[5] Astrocytomas, which arise from astrocytes, account for about half of all malignant CNS tumors.[5]

Lymphomas represent the third most common type of childhood cancer, accounting for approximately 15% of malignancies.[5] Lymphomas are rare among young children, with rising rates in adolescence.[5] Lymphomas arise from lymphocytes and are commonly classified into two etiologically distinct subgroups: Hodgkin lymphoma, which arises from B-lymphocytes with characteristic Reed-Sternberg cells, and non-Hodgkin lymphoma, which includes various histological subtypes.[5] Malignant bone tumors, such as osteosarcoma and Ewing's sarcoma, are also more common among older children and adolescents.[5] Conversely, embryonal tumors, which arise from aberrant development of stems cells, are more common among younger children.[5] Types of embryonal tumors include neuroblastoma (arising from neural crest cells), retinoblastoma (arising in the retina or, extremely rarely, in the pineal gland), Wilms' tumor (arising from primitive metanephric blastemal), and hepatoblastoma (arising from immature liver precursor cells).[5] Finally, soft tissue sarcomas (STS), which are primarily of mesenchymal cell origin, and germ cell tumors (GCT) are two heterogeneous groups of tumors that occur in children.[5] Other types of childhood cancers, such as carcinomas,[5] are rare and not often studied in isolation.

Etiology of Childhood Cancer

Due to the rarity of childhood cancers, progress in understanding the etiology of these diseases has been slow, though advances have been made over time.[3]

Demographic Factors – Incidence varies by demographic factors including age, sex, and race/ethnicity.[5] For most childhood cancers, incidence is slightly higher among males compared to females, though Wilms' tumor is a notable exception.[3,5] Incidence is

also higher among non-Hispanic white children for many, but not all, types of cancer.[5] For example, the incidence of acute lymphoblastic leukemia (ALL) is about 10% higher in Hispanic children compared to non-Hispanic white children.[3]

Genetic Factors – Efforts to sort out the role of genetic versus non-genetic risk factors in childhood cancer etiology are ongoing. Inherited syndromes caused by high-penetrance germline DNA mutations, chromosomal aneuploidy, or epigenetic disorders are estimated to account for 5 to 10% of childhood malignancies.[3] Examples of associated syndromes include trisomy 21 (Down syndrome) and leukemia;[6,7] Li-Fraumeni syndrome and STS, osteosarcoma, and CNS tumors;[8] and neurofibromatosis types 1 and 2 and CNS tumors.[8] The proportion of disease attributed to inherited syndromes may be much higher for rare cancers,[3] such as adrenocortical carcinomas.[9]

Findings from genome-wide association studies (GWAS) suggest that common genetic variants may explain a greater proportion of the population-attributable risk for childhood compared to adult cancers.[3] For example, GWAS conducted among racially and ethnically diverse subjects provide evidence of an association between the *ARID5B* rs10821936 polymorphism and increased risk of ALL.[3,10] Further, these studies demonstrate variability in allele frequency across racial and ethnic groups, with 33% presentation in Europeans, 27% in African Americans, and 47% in Hispanics.[3,10] Thus, this line of research may help explain noted racial and ethnic differences in incidence.

Infection – Exposure to infections may contribute to risk of some childhood cancers. There is evidence that Epstein-Barr virus increases risk of Burkitt's lymphoma and possibly some subtypes of Hodgkin lymphoma.[11] Further, studies of parental

occupational contact, population mixing, daycare attendance, and medical histories provide some indirect evidence of a possible infectious etiology of other childhood cancers, particularly ALL,[12,13] though no specific viruses have been identified.[3,12,13]

There are two main hypotheses for a possible association between delayed exposure to infection and ALL risk. Greaves proposed a two-hit hypothesis in which ALL arises as a consequence of at least two independent mutations.[14,15] The first leukemic translocation is initiated *in utero*.[14,15] If a child experiences limited exposure to infections in the first year of life, the risk of a subsequent mutation(s) increases due to improper immune system development, thus eliciting an abnormal proliferative response.[14,15] Greaves' hypothesis is supported by evidence from studies of the clonal relationship of concordant leukemia in identical, monozygotic, twins,[16] and from 'backtracking' studies that screened for fusion-gene sequences in DNA from archived Guthrie cards.[17,18] Alternatively, Kinlen proposed that clusters of childhood ALL arise from population mixing, which occurs when a relatively unexposed, immunologically naïve, population is introduced to a new virus, which then leads to lymphoproliferative stress.[19,20] Kinlen's hypothesis is supported by a recent meta-analysis that estimated an increased risk of ALL in rural settings of population mixing.[21] Other hypotheses include Smith's hygiene hypothesis, which theorizes that improved hygiene decreases a mother's chances of having protective antibodies against an unknown leukemic agent,[22] and Schmiegelow's adrenal hypothesis, which theorizes that early exposure to infection reduces ALL risk through a direct anti-leukemic cortisol effect and modification of the T-helper-1/T-helper-2 balance.[23]

Pregnancy-Related Factors – Given the early onset of many childhood cancers, pregnancy-related factors are thought to contribute to etiology. Prior studies have linked maternal age, birthweight, and birth order with many, if not most, childhood cancers.[3] A large pooled analysis of population-based record-linkage studies, conducted by Johnson *et al*. (2009), found significant positive linear trends between increasing maternal age and risk of combined childhood cancers (all diagnoses), leukemias, lymphomas, CNS tumors, neuroblastoma, Wilms' tumor, bone tumors, and STS.[24] Estimated associations ranged from 6 to 15% increased risk per 5-year increase in maternal age.[24] The underlying cause of the association between advanced maternal age and childhood cancer risk is not currently known. Proposed mechanisms include differential expression of cell cycle control genes, DNA damage to response and repair pathways, age-related decreases in oocyte gene expression, and increased *de novo* epimutations in oocyte genes.[24,25] Other potential explanations include assisted reproductive technology use[26] and age-related changes in hormone levels during pregnancy.[24]

Risk of ALL, CNS tumors, neuroblastoma, and Wilms' tumor are shown to rise as a linear function of birthweight, though to varying magnitudes.[3] Acute myeloid leukemia (AML) risk is elevated with low and high birthweight.[3] Hepatoblastoma risk is inversely associated with birthweight, with strikingly elevated risk among the smallest infants.[3,27] Recent studies using alternative measures of birth size, such as size for gestational age and percent of optimal birthweight, demonstrate similar associations.[3] As with maternal age, the reasons for birthweight associations are not known, but may include prenatal growth hormone exposure, an increased number of cells at risk of carcinogenic

transformation, or underlying genetic factors associated with birthweight.[3] For example, the role of the insulin-like growth factor system in childhood cancer risk is a promising area of research.[28] The strong inverse association between birthweight and hepatoblastoma risk may be due to medical exposures in the neonatal intensive care unit, such as irradiation and oxidative stress, combined with the immature defense mechanisms of premature infants, though no definitive exposures have been identified.[27]

A large pooled analysis of population-based record-linkage studies, conducted by Von Behren *et al*. (2011), reported an inverse association between birth order and risk of combined childhood cancers, ALL, CNS tumors, neuroblastoma, bilateral retinoblastoma, Wilms' tumor, and rhabdomyosarcoma.[29] A positive association between increasing birth order and AML risk was also observed.[29] Birth order may be a marker of other potential risk factors, such as infection.[29] Later born children may be exposed to infections at earlier ages than first born children through contact with older siblings, though this may not hold if the birth interval is large, or if infections are acquired from other sources, such as daycare attendance.[29] Birth order may also be a marker of differential hormone exposures *in utero*,[29] which may influence cancer risk.[30,31] Estrogen levels in maternal and umbilical cord blood samples are shown to be somewhat higher in the first pregnancy compared with second and third pregnancies.[32-34]

Other pregnancy-related exposures have been studied in the childhood cancer literature including maternal substance use,[35-37] vitamin use,[38] and exposure to residential and occupational pesticides.[39-41] However, no definitive associations have emerged among these types of exposures.[3]

Socioeconomic Status – Socioeconomic status (SES) is a multidimensional construct that encompasses both economic resources and social standing.[42,43] SES is consistently linked to a range of health outcomes in children and adults including many adulthood cancers.[44-46] However, evidence of an association between SES and childhood cancer incidence remains limited and inconsistent. Prior studies of SES have been largely confined to childhood leukemias, and ALL in particular. This is likely because of the relatively high incidence of leukemias in children, as well as the hypothesized infectious etiology of ALL, which may be socially patterned.[13,47,48] Due to the lack of individual-level socioeconomic data in cancer registries and medical records,[49] studies of SES and childhood cancer risk have been mostly ecological in design, especially in earlier years. Ecologic studies have used a variety of measures to operationalize SES, including area-level income, education, occupation, and composite indices.[50-54] Findings from these studies generally suggest an adverse association between higher SES and childhood cancer risk.[50-54] Though few studies have tested associations of SES and childhood cancer risk at the individual-level, among those that have, null associations between SES and childhood cancer risk have mostly been reported.[43,55-59] A recent study by Heck *et al.* (2016) did find lower risk of several cancers among Hispanic children of foreign born mothers compared to Hispanic children of US born mothers.[60] Although nativity is not a direct measure of SES, this study provides some individual-level evidence that environmental, social, and cultural factors may contribute to childhood cancer risk.[60] Given the potential biases that can arise in ecologic studies,[61] as well as the limited

evidence of an association at the individual-level, further research is needed to clarify associations of SES and childhood cancer risk.

Despite the lack of conclusive empirical evidence, there is theoretical grounding for a potential association between SES and childhood cancer risk. As previously stated, SES may drive patterns of exposure to infections, which in turn may influence risk of some childhood cancers.[11,12] The differential distribution of ALL and Hodgkin lymphoma in the pediatric population of high-income versus low-income countries provides some indirect support for this hypothesis.[52] SES may also influence risk through the social patterning of established pregnancy-related exposures, such as maternal age, birthweight, and birth order.[3] Women of higher SES are more likely to delay childbearing and have fewer children than women of lower SES.[62] Further, high SES women are at lower risk of adverse birth outcomes, including low birthweight and small for gestational age, than low SES women.[63] Finally, SES may influence risk through environmental exposures, such as air pollutants, pesticides, and parental substance use.[64-66] However, besides high-dose ionizing radiation and prior chemotherapy, no environmental exposures have emerged as definitive causes of childhood cancer.[3] A better understanding of the relationship between SES and childhood cancer risk may provide etiologic insight into the role of such exogenous exposures.

Temporal Trends in Incidence

The incidence of combined childhood cancers has gradually increased in the United States by approximately 0.6% per year since 1975.[2] During this period, incidence has also increased for several individual childhood cancers including leukemias, CNS

tumors, hepatoblastoma, STS, and GCT.[67] The underlying cause of increasing incidence rates is not currently known. Some researchers speculate that rising rates may be an artifact of changes over time in diagnostic technology, disease classification, and registry completeness.[68-70] For example, the sharp increase in CNS tumor diagnoses in the 1980s is thought to be attributed to the introduction of magnetic resonance imaging and stereotactic biopsy.[68] Diagnostic changes are less likely to explain the observed rise in other types of childhood cancer, such as childhood leukemias.[68]

Others hypothesize that rising rates reflect a true increase in incidence attributed to changes over time in exogenous risk factors, such as environmental exposures (e.g. low-level radiation).[71-73] However, as previously noted, there is little empirical evidence linking such exposures to childhood cancer risk.[3] Further, some environmental pollutants have decreased in the United States since the introduction of the Clean Air Act in 1970.[74] A more plausible hypothesis is that the rise in childhood cancer incidence is due to an increase over time in established pregnancy-related exposures, such as older maternal age, heavier birthweight, and decreased parity.[3] A population-level shift in recent decades towards delayed childbearing and smaller family size is well documented in the United States,[62] and rising overweight and obesity rates are speculated to have influenced birthweight trends.[75,76] To date, limited work has been done to empirically test this hypothesis. An age-period-cohort study in Piedmont, Italy, conducted by Maule *et al*. (2007), found a positive association between increasing maternal age and ALL incidence over time.[77] However, this study did not account for potential confounders, such as birthweight and parity, nor did it consider other childhood cancers, besides ALL, that also

demonstrate associations with maternal age.[24] Additional time series research is needed to better understand the role of pregnancy-related exposures, as well as other potential risk factors, in explaining temporal trends in childhood cancer incidence.

Childhood Cancer Survival

Childhood cancer survival has improved markedly over the past 30 years due to new and improved treatments and standardization in care of the pediatric population.[67] The 5-year survival rate for combined childhood cancers increased from 58% in the mid-1970s to 82% between 2003 and 2009.[78] However, survival rates vary considerably by patient and tumor characteristics, such as age at diagnosis and cancer type.[5] Survival also varies by race and ethnicity, with the largest disparity documented between non-Hispanic black and non-Hispanic white children.[67] Based on 2005-2011 estimates, the 5-year survival rate for combined childhood cancers was only 78% among non-Hispanic blacks, ages 0 to 19 years, compared to 85% among non-Hispanic whites.[4] Black-white racial disparities in survival are also documented for several individual childhood cancers including leukemias,[79-81] lymphomas,[82-84] CNS tumors,[85] and some extracranial solid tumors.[86-88] The reasons for differences in survival by race are not well understood, with both biological and socioeconomic mechanisms proposed in the literature.[89]

There is some evidence suggesting that underlying genetic variations associated with ancestry may lead to differences in tumor biology for some childhood cancers.[90,91] For example, several studies demonstrate that high white blood cell count and older age at presentation contribute to the overrepresentation of black children in the ALL high-risk group.[92-94] A study by Metzger et al. (2008) reported a greater proportion of high-risk

features among black compared to white Hodgkin lymphoma pediatric patients including advanced stage at diagnosis, low hemoglobin levels, and high erythrocyte sedimentation.[95] A study by Baker et al. (2002) reported a greater proportion of high-risk features among non-white compared to white rhabdomyosarcoma pediatric patients including presentation with invasive T2 tumors, positive regional lymph nodes, larger tumor size, and advanced stage at diagnosis.[96] Because these studies largely relied on comparisons of clinical characteristics, a more sophisticated molecular understanding of differences in disease biology across racial and ethnic groups is needed.[90]

There is also some evidence that variability in pharmacogenomics may contribute to survival differences by race.[90] For example, the antimetabolite 6-mercaptopurine, which is used to treat ALL, is found to be more active in patients with genetic deficiency of thiopurine methyl transferase, an enzyme involved in detoxification.[97] Variation in the frequency and distribution of mutant alleles by ancestry has been shown.[98-102] Variability in enzymes, such as CYP3A4, responsible for the metabolism of drugs, such as alkylating agents and vinca alkaloids, may also contribute to observed survival differences.[93]

However, race is a socially constructed taxonomy that is not synonymous with ancestry.[91] Race is highly correlated with SES,[103,104] especially in the United States where embedded institutionalized racism continues to place racial and ethnic minorities at high risk of low SES.[105] Empirical evidence of an association between SES and childhood cancer survival is emerging. A 2014 systematic review by Gupta *et al*. positively linked higher SES to improved survival from combined childhood cancers,[106] a 2014 meta-analysis by Petridou *et al.* positively linked SES to survival from childhood leukemias,[107]

and evidence is emerging positively linking SES to survival from types of childhood solid tumors.[85,87,108] Therefore, it is plausible that SES also contributes to documented survival differences by race. Underlying mechanisms linking SES to childhood cancer survival may include early diagnosis and entry into care, ready access to quality health care including clinical trial enrollment, and sufficient time and energy to maintain adherence to therapy.[90] More work is needed to disentangle the relative role of social versus biological mechanisms in explaining racial and ethnic disparities in childhood cancer survival.

Dissertation Objective

A primary goal of childhood cancer research is to understand the role of non-genetic, socially patterned, risk factors in incidence and survival. Many unanswered questions persist in the literature that, if answered, may help address this aim. Such questions include: do pregnancy-related exposures contribute to the rise in childhood cancer incidence over time; is there an association between SES and childhood cancer incidence; and what is the underlying role of social versus biological factors in explaining racial survival disparities? Efforts to address these questions have been hampered by the dearth of robust, population-based, data available to rigorously test childhood cancer associations. Progress has also been slowed by the underutilization of advanced statistical methods, such as multilevel modeling, in the childhood cancer literature. Innovative approaches to data collection and analysis are needed to move the field forward in understanding socially patterned risk factors of childhood cancer incidence and survival. This dissertation contributes to the childhood cancer literature by rigorously testing each

of the previously defined research questions in three separate investigations using population-based registry data from multiple existing data sources and advanced statistical methods.

Manuscript 1 – The aim of this study was to investigate whether temporal trends in pregnancy-related risk factors contribute to the rise in childhood cancer incidence over time. To achieve this aim, we conducted a time series ecologic analysis at the county-level. We used population-based birth and cancer registry data linked from the National Center for Health Statistics and the Surveillance, Epidemiology, and End Results (SEER) program. We tested temporal associations using Poisson mixed models. We assessed multiple pregnancy-related risk factors including maternal age, birthweight, and birth order; we adjusted for temporal trends in county-level sociodemographic factors including race, ethnicity, and poverty; and we tested associations for combined and individual childhood cancers including types of leukemias and solid tumors. We hypothesized that population-level shifts towards older maternal age, higher birthweight, and smaller family size contribute to the rise in childhood cancer incidence over time.

Manuscript 2 – The aim of this study was to determine if SES, measured at multiple levels of exposure, is associated with childhood cancer incidence. We conducted a population-based case-cohort study through linkage of data from Minnesota birth and cancer registries. We tested measures of SES at both the individual-level and neighborhood-level to potentially reconcile contradictory findings from prior ecologic and individual-level studies. We tested associations using logistic mixed models, and we accounted for established demographic and pregnancy-related risk factors that may

confound or mediate associations of SES and childhood cancer incidence. We hypothesized that higher SES is adversely associated with childhood cancer incidence, which is due, at least in part, to the social patterning of established demographic and pregnancy-related risk factors.

Manuscript 3 – The aim of this study was to determine if SES mediates the association between race (black versus white) and childhood cancer survival and, if so, to what degree. We also tested prognostic factors, including tumor stage and treatment, as other potential mediators of the race-survival association. We employed the semiparametric inverse odds weighting method to formally test for mediation. We used population-based SEER data representative of the US pediatric cancer population, and we assessed survival from combined and individual childhood cancers to determine if mediation differs across tumor groups. We hypothesized that SES contributes to (i.e. mediates) the association between race and survival, and thus, after accounting for the mediating pathway of SES, black-white racial disparities in childhood cancer survival will be attenuated.

Chapter 2

Manuscript 1: Do pregnancy-related risk factors contribute to rising childhood cancer incidence rates in the United States?

Introduction

Since 1975, the overall childhood cancer incidence rate has been gradually increasing in the United States at an annual rate of approximately 0.6% per year.[2] During this period, incidence rates have also increased for several individual types of childhood cancer including leukemias, central nervous system (CNS) tumors, and hepatoblastoma.[67] The underlying cause of rising rates is not known, and few studies have conducted time series analyses to understand trends. Some researchers speculate that rising rates may reflect changes over time in diagnostic technology, disease classification, and registry completeness.[68-70] For example, it is thought that the sharp increase in CNS tumor diagnoses in the 1980s was due to the introduction of magnetic resonance imaging and stereotactic biopsy.[68] However, there are also plausible explanations for why the documented rise may reflect a true increase in disease. We hypothesize that rising incidence rates are attributed to increases over time in established pregnancy-related risk factors including older maternal age, higher birthweight, and reduced parity.[3] A population-level shift in recent decades towards delayed childbearing and smaller family size is well documented in the United States.[62] Further, rising overweight and obesity rates are speculated to have influenced birthweight trends.[109,110] Mounting evidence links these pregnancy-related exposures to many, if not most, childhood cancers.[3,24,29,111]

Therefore, it is reasonable to suppose that some proportion of the rise in childhood cancer incidence is due to secular trends in these established risk factors.

To our knowledge, only one prior study has empirically tested temporal associations between pregnancy-related risk factors and cancer incidence in the childhood cancer literature. Maule *et al.* (2007) conducted an individual-level age-period-cohort study in Piedmont, Italy to test temporal associations between maternal age and acute lymphoblastic leukemia (ALL) incidence.[77] Study results suggest that a trend towards older maternal age does explain, at least in part, the increase in ALL incidence over time.[77] However, this study was confined to only 252 cases of ALL diagnosed between 1980 and 1997. Other types of childhood cancer, which also demonstrate individual-level associations with maternal age,[24] were not considered. Further, this study did not account for potential confounding by other risk factors, such as birthweight and birth order. Additional research is needed to more thoroughly describe temporal associations between established pregnancy-related risk factors and childhood cancer incidence. We conducted a time series ecologic analysis using population-based registry data in the United States to test associations between county-level trends in pregnancy-related risk factors and childhood cancer incidence over a 39-year period. We assessed multiple risk factors including maternal age, birthweight, and birth order; we adjusted for temporal trends in county-level sociodemographic characteristics including race, ethnicity, and poverty; and we tested associations for combined and individual childhood cancers including types of leukemias and solid tumors. We hypothesized that adjustment for county-level temporal

trends in pregnancy-related risk factors would attenuate the estimated average annual percent change (AAPC) in childhood cancer incidence.

Methods

Study Population

Our study sample consisted of 194 counties from eight cancer registries included in the Surveillance, Epidemiology, and End Results (SEER) 9 database (1975-2013): Atlanta, Connecticut, Detroit, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah.[78] We linked county-level aggregated cancer data, based on addresses at time of diagnosis, to county-level aggregated birth data with the assumption that county of residence was stable from birth to diagnosis. To minimize bias from this assumption, we restricted county incidence rates to younger children, ages 0 to 4 years, given that older children are more residentially mobile.[112] Between 1975 and 2013, average county population size of children, ages 0 to 4 years, in our sample was 8,769 (SD = 20,139), ranging from 22 children in Harding County, NM (2002, 2004) to 201,142 children in Wayne County, MI (1975).

Data Linkage

Our analysis was made feasible through linkage of data from three sources: (1) SEER, (2) the National Center for Health Statistics (NCHS), and (3) the US Census Bureau. Data were merged at the county-level using Federal Information Processing Standard (FIPS) identifiers. We obtained annual county-level incidence rates, 1975 to 2013, for children, aged 0 to 4 years, from the SEER database. Denominators were based on US Census Bureau annual population estimates. Incidence rates were estimated for

combined childhood cancers (all diagnoses) and for individual cancers occurring among younger children including ALL, acute myeloid leukemia (AML), combined CNS tumors, neuroblastoma, retinoblastoma, Wilms' tumor, and hepatoblastoma. Cancers were classified using the International Classification of Childhood Cancer, third edition.[113] We were granted access to annual birth files with county identifiers, 1970 to 2013, from the NCHS. We aggregated pregnancy-related factors from birth records including maternal age (years), birthweight (grams), and live birth order (0-8+) and estimated county-level means, averaged over the 6-year birth window corresponding to year of diagnosis. For example, 1975 cancer incidence rates for children, ages 0 to 4 years, include cases born between 1970 and 1975; therefore 1975 incidence rates were linked to birth data averaged 1970-1975. We linked county-level demographic and socioeconomic measures from the US Census including proportion of residents, ages 0-4 years, classified as white race (%white) and Hispanic ethnicity (%Hispanic); and proportion of residents, all ages, below the poverty line (%poverty). Data on race were available from annual population estimates, while data on ethnicity (1980-2010) and poverty (1970-2010) were only available from decennial censuses. We linearly interpolated intercensal years so that, for example, a county with 10% Hispanic children in 1990 and 20% in 2000 would be assigned a value of 11% Hispanic in 1991. Because ethnicity data were not available from the 1970 census, we used 1980 values to impute %Hispanic for years 1975-1979.

Statistical Analysis

Among our sample of 194 counties, we descriptively assessed overall trends in cancer incidence and county-level characteristics by graphing annual rates and means between 1975 and 2013. We estimated annual trends in county-level characteristics from linear mixed models with a random intercept for repeated county-level measures over time; a quadratic term for time was included in the model predicting birthweight. We tested associations of county-level characteristics predicting childhood cancer incidence in Poisson mixed models with a random intercept for repeated county-level measures over time. The model implicitly controls for heterogeneity in population size by specifying incidence rate denominators (county population of children aged 0-4) in the offset term. We tested associations between each county-level characteristic and childhood cancer incidence, first in models adjusted only for year of diagnosis, and then in models further adjusted for all other county-level characteristics. Rate ratios (RR) and 95% confidence intervals (CI) were estimated for a 5-year change in county-level average maternal age, a 500-gram change in county-level average birthweight, a 1-unit change in county-level average birth order, and a 10% change in county-level %white, %Hispanic, and %poverty. We tested for interactions between county-level characteristics and year of diagnosis, but found none were significant in adjusted models.

We estimated the crude AAPC from a Poisson mixed model with only year of diagnosis predicting cancer incidence [AAPC = $(e^{\beta yeardx} - 1)*100)$]. For cancers in which incidence significantly changed over time in crude models, we estimated the percent change between the crude AAPC and AAPCs adjusted for county-level characteristics

[%change = ((βyeardx$_{adjusted}$–βyeardx$_{crude}$)/ βyeardx$_{crude}$)]. We calculated AAPCs from models adjusted for each county-level characteristic one at a time (e.g. a model specifying year of diagnosis and maternal age predicting incidence), and from models fully adjusted for all county-level characteristics simultaneously. As a secondary analysis, we tested models specifying the baseline value and the change from baseline for each county-level characteristic, but found no substantial differences from crude estimates (results not presented). We tested the null hypothesis (AAPC = 0%) using a two-sided test; statistical significance was determined as $p < .05$. Analyses were performed using Stata 14.2 (College Station, TX).[114]

Results

In Table 2-1, we present county-level characteristics averaged over the first (1975-1980) and last (2008-2013) six years of the study period. Over the study period, average maternal age increased from 24.8 years (standard deviation (SD), 0.7) in 1975-1980 to 27.0 years (SD, 1.4) in 2008-2013, while average birthweight slightly decreased from 3,361 grams (SD, 102) to 3,300 grams (SD, 89). Average %white decreased from 95.0% (SD, 10.5) in 1975-1980 to 89.4% (SD, 13.3) in 2008-2013, while average %Hispanic increased from 9.8% (SD, 19.0) to 19.2% (SD, 21.2). Average live birth order and %poverty remained stable over the study period.

Annual trends in county-level characteristics are depicted in Figure 2-1. Temporal trends further demonstrate that, over the study period, average county-level maternal age (β, 0.068; 95% CI, 0.067, 0.070) and %Hispanic (β, 0.30; 95% CI, 0.29, 0.30) increased, average %white (β, -0.17; 95% CI, -0.17, -0.16) decreased, and average birth order (β, -

0.0007; 95% CI, -0.0009, -0.0005) and %poverty (β, -0.022; 95% CI, -0.026, -0.017) remained relatively stable. Though there was an overall decrease in average birthweight between 1975 and 2013, a steady increase in average birthweight occurred in earlier years before reversing direction in the late 1980s (linear β, 4.26; 95% CI, 4.00, 4.52; quadratic β, -0.17; 95% CI, -0.18, -0.17). Several county-level characteristics were highly correlated (Table 2-2), though standard errors remained stable in fully adjusted models.

We present associations between county-level characteristics and combined childhood cancer incidence in Table 2-3. After adjusting for year of diagnosis, county-level average maternal age (rate ratio (RR), 1.22; 95% CI, 1.11, 1.34), average birthweight (RR, 1.57; 95% CI, 1.37, 1.80), and %white (RR, 1.03; 95% CI, 1.01, 1.04) were positively associated with county-level incidence of combined childhood cancers over the 39-year study period. County-level average birth order (RR, 0.86; 95% CI, 0.76, 0.97), %Hispanic (RR, 0.97; 95% CI, 0.95, 0.98), and %poverty (RR, 0.87; 95% CI, 0.83, 0.90) were inversely associated with county-level combined childhood cancer incidence. However, associations between county-level characteristics and combined childhood cancer incidence were attenuated towards the null, and no longer statistically significant, in fully adjusted models. Similar patterns of association were observed among individual childhood cancers, though to varying magnitudes and with some differences in directionality (Table 2-4). For example, %Hispanic was positively associated with incidence of ALL (fully adjusted RR, 1.05; 95% CI, 1.02, 1.08).

As illustrated in Figure 2-2, the incidence of combined childhood cancers significantly increased between 1975 and 2013 at an AAPC of 0.71% per year (95% CI,

0.55, 0.86). Between 1975 and 2013, a statistically significant increase in incidence was also observed for ALL (AAPC, 0.78%; 95% CI, 0.49, 1.07), AML (1.86%; 1.13, 2.59), CNS tumors (1.31%; 0.94, 1.67), and hepatoblastoma (2.70%; 1.68, 3.72). There was no statistically significant change over time in incidence of neuroblastoma (AAPC, 0.28%; 95% CI, -0.12, 0.68), retinoblastoma (0.25%; 95% CI, -0.38, 0.88), or Wilms' tumor (-0.15%; 95% CI, -0.65, 0.36).

In Table 2-5, we compare crude and adjusted AAPCs in incidence of childhood cancers with significant temporal trends (combined, ALL, AML, CNS tumors, and hepatoblastoma). Across all the county-level characteristics and all the cancers tested, the most notable reduction in AAPC in cancer incidence occurred after adjustment for county trends in maternal age. Adjustment for county-level average maternal age reduced the AAPC in incidence of combined childhood cancers to 0.32% per year (95% CI, 0.08, 0.56), a 55% reduction from the crude AAPC, and reduced AAPCs in incidence of individual cancers by between 8% for hepatoblastoma to 40% for ALL. However, even after adjustment for maternal age, AAPCs in county-level cancer incidence remained significant from zero. Adjustment for other county-level characteristics either had no effect on county-level cancer incidence rates over time (e.g. birth order), or increased the AAPC, suggesting a masking effect. For example, adjustment for county-level average birthweight increased the AAPC in combined childhood cancer incidence by 21% from the crude (adjusted AAPC, 0.86; 95% CI, 0.69, 1.02).

AAPCs in incidence of combined and individual childhood cancers remained statistically significantly above 0% in models fully adjusted for all county-level

characteristics, with only an attenuation towards the null observed for AML (adjusted AAPC, 1.62%; 95% CI, 0.38, 2.87; 13% reduction) and hepatoblastoma (2.36%; 95% CI, 0.71, 4.04; 12%). Fully adjusted AAPCs increased from the crude for combined childhood cancers (0.75%; 95% CI, 0.48, 1.03; 6% increase), ALL (0.81%; 95% CI, 0.31, 1.31; 4%), and CNS tumors (1.69%; 95% CI, 1.03, 2.35; 29%).

Discussion

This is the first study to test temporal associations between pregnancy-related risk factors and cancer incidence rates in the US pediatric population. Through the use of linked population-based registry data over almost four decades, we confirmed that the combined cancer incidence rate among children, 0 to 4 years of age, has been gradually increasing in the United States by about 0.7% per year since 1975.[2] We also confirmed increasing trends in incidence rates of ALL, AML, CNS tumors, and hepatoblastoma.[67] For these cancers, we tested whether temporal trends in select county-level characteristics were associated with rising incidence rates.

We found preliminary evidence of a temporal association between county-level average maternal age and childhood cancer incidence rates. Descriptive analysis revealed that, of all the county-level characteristics assessed, the temporal trend in county-level average maternal age most closely aligns with trends in childhood cancer incidence rates. Specifically, we showed that county-level average maternal age has been steadily increasing in the United States at an annual rate of about 0.07 years since 1975 (Figure 2-1). Further, we found that estimated AAPCs in incidence of combined and individual childhood cancers were substantially attenuated towards the null in models adjusted only

for maternal age, with a reduction from the crude of between 8% for hepatoblastoma to 55% for combined childhood cancers (Table 2-5). However, we note that AAPCs adjusted only for maternal age remained significantly above the null value of 0% change. Further, AAPCs estimated from fully adjusted models indicated a much weaker temporal association between maternal age and childhood cancer incidence.

Besides maternal age, we found no evidence that other county-level characteristics including birthweight, birth order, and sociodemographic factors, are associated with rising childhood cancer incidence rates over time. Adjustment for these factors resulted in an estimated AAPC in childhood cancer incidence that was either higher than or no different than the crude. This suggests that, if anything, trends in these county-level characteristics counteract rising childhood cancer incidence rates over time in the United States.

Contrary to our study hypothesis, estimated AAPCs in incidence rates of combined and individual childhood cancers remained significantly above 0% after comprehensive adjustment for county-level sociodemographic and pregnancy-related characteristics. In fact, only the fully adjusted AAPCs for AML and hepatoblastoma were attenuated towards the null; all other fully adjusted AAPCs were higher than the crude (combined, ALL, CNS tumors). While individual-level associations between maternal age and AML have been previously documented in the literature, there is currently little evidence of an association between maternal age and hepatoblastoma risk.[24] Therefore, the observed reduction in the hepatoblastoma incidence trend attributed to maternal age should be interpreted with caution, especially given small case counts.

Limitations

There are several limitations to our analysis. First, as with any ecologic study, group-level risk factors may not be associated with incidence at the individual-level.[61] This is especially true for larger counties in which greater heterogeneity in pregnancy-related and sociodemographic characteristics is expected. To assess this limitation, we conducted a sensitivity analysis in which we estimated crude and adjusted AAPCs in incidence of combined childhood cancers within subsets of our sample restricted to (1) counties of <10,000 children (averaged over 1975-2013), (2) counties of <5,000 children, and (3) counties of <1,000 children. This revealed that maternal age accounted for a much greater portion of the annual trend in combined childhood cancer incidence for smaller counties than what we observed in our analysis of all counties with available data (Table 2-6). This is consistent with results we would expect from measurement error of area-level maternal age, which could be higher in larger counties (given larger variation), that serves to minimize the percent reduction after adjustment. However, we note that estimates lost precision upon restriction to smaller counties.

Second, because county of birth is not available in SEER, we made the assumption that county of residence was stable from birth to diagnosis. To minimize potential bias due to this assumption, we restricted our sample to younger children, ages 0 to 4 years, given that residential mobility increases with age.[112] We acknowledge that a California-based study reported that 38.5% of leukemia cases, ages 0 to 4 years, had moved away from county of birth by time of diagnosis, indicating that residential mobility is relatively common among younger children, at least in California.[112]

Nevertheless, this study found no significant differences in sociodemographic characteristics of counties at birth and diagnosis among residentially mobile cases.[112] Further, a recent case-cohort study in Minnesota reported high correlation between the socioeconomic status of residential census tracts at birth and diagnosis among cancer cases, ages 0 to 14 years.[115] Therefore, even if our assumption of stable residency is incorrect, it may be appropriate to assume stable county-level characteristics throughout early childhood.

Third, there were limitations to our measures of county-level characteristics. For example, our measure of county-level poverty was not specific to the early childhood population (0 to 4 years). We also did not consider all potential pregnancy-related risk factors, such as paternal age, due to issues of missing data and potential collinearity.[24] Further, we note potential concerns of a few strong correlations among county-level characteristics (Table 2-2), which may have hindered our ability to fully disentangle associations, especially in ecologic data.[61] To further tease apart county-level characteristics, we conducted a sensitivity analysis in which we tested temporal associations between pregnancy-related risk factors and cancer incidence among only white children within counties, which produced similar overall patterns (Table 2-7).

Finally, this study was limited by sample size, both in terms of observational units (N=194 counties) and small case counts within counties. The rarity of childhood cancers required us to combine all diagnoses among ages 0 and 4 years into a single rate, which resulted in less precise exposure measures aggregated over the six-year birth window.

Given the noted limitations of our study, further time series research is needed in larger analytic samples, and using more robust data, to confirm temporal associations between pregnancy-related risk factors and childhood cancer incidence rates. For example, future studies should test temporal associations in data aggregated over smaller observational units, such as at the census tract-level. Studies should also employ birth and cancer registry data linked at the individual-level to address such questions as whether cancer incidence rates have increased over time when restricted to children born to older mothers (e.g. > 35 years). We emphasize the need for thorough covariate control in future time series research. This is underscored by the fact that we only identified significant temporal associations between maternal age and childhood cancer incidence in minimally adjusted models, and that the one prior study to report a significant temporal association between maternal age and ALL incidence did not adjust for potential confounders.[77]

Conclusion

Overall, we did not find conclusive evidence to support our hypothesis that the rise in pregnancy-related risk factors accounts for the documented rise in childhood cancer incidence over time. While preliminary findings suggested an association between rising maternal age and childhood cancer incidence over time, comprehensive adjustment for all county-level characteristics had little impact on estimated incidence trends in combined and individual childhood cancers. Therefore, we cannot rule out alternative explanations for increasing trends, including the possibility that rising incidence rates are an artifact of changes over time in diagnostic technology, disease classification, and registry completeness. We also cannot rule out the possibility that rising childhood cancer

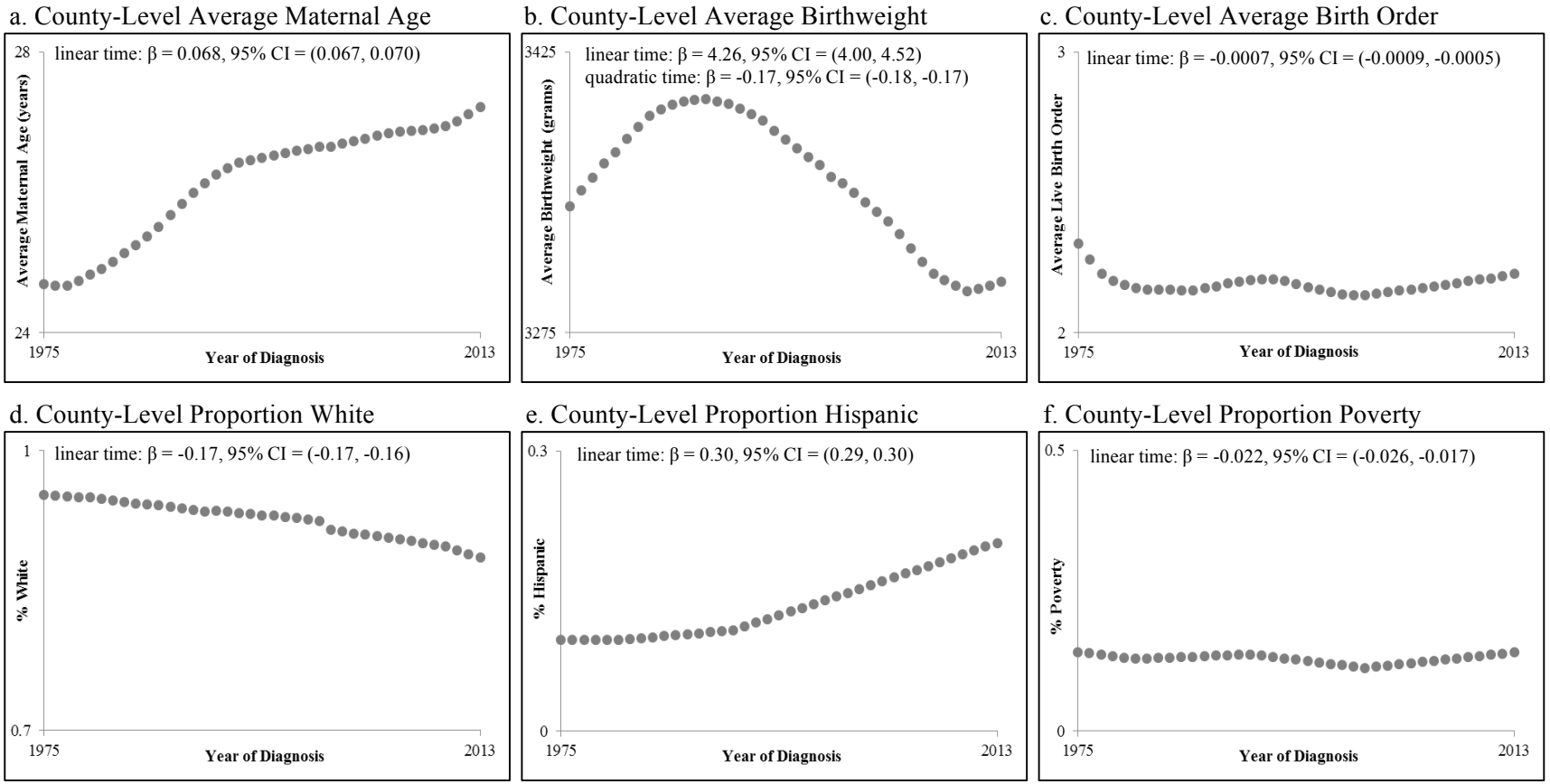
incidence rates may be attributed to changes over time in other risk factors. For example, it has been hypothesized that environmental exposures, such as low-level radiation, may contribute to the rise in childhood cancer incidence over time.[71-73] However, there is currently little evidence linking such exogenous exposures to childhood cancer risk,[3] and levels of some environmental pollutants may have actually decreased in the United States during our study period.[74] Moreover, although exposure to environmental hazards is generally greater among African American children,[64] their rate of ALL and many solid tumors is substantially lower than among white children.[67] Therefore, the rise in childhood cancer incidence rates remains a topic of debate in the literature, and thus further time series research is needed that can build upon our findings and ultimately pinpoint the underlying cause of temporal trends.

Table 2-1 County-level pregnancy-related and sociodemographic characteristics of SEER 9 counties (N=194)

	1975-1980					2008-2013				
County-Level Characteristic	Mean	Std	25th Pct	Median	75th Pct	Mean	Std	25th Pct	Median	75th Pct
Average Maternal Age (years)	24.8	0.7	24.3	24.7	25.1	27.0	1.4	26.2	26.9	27.6
Average Birthweight (grams)	3,361	102	3,312	3,388	3,434	3,300	89	3,238	3,320	3,366
Average Live Birth Order	2.2	0.3	2.0	2.1	2.4	2.2	0.2	2.1	2.1	2.3
% White	95.0	10.5	95.6	98.8	99.5	89.4	13.3	87.0	94.5	97.1
% Hispanic	9.8	19.0	0.4	1.5	5.1	19.2	21.2	4.5	10.8	24.0
% Poverty	13.5	6.4	9.4	12.1	15.3	13.5	5.1	9.9	12.5	16.2

Notes: Std = standard deviation; Pct =percentile.

Figure 2-1 Temporal trends in county-level characteristics, SEER 9 counties (N=194), 1975-2013



Notes: Annual county-level estimates are averaged across the sample of 194 counties (i.e. points represent the annual average of county-level averages). Time beta coefficients (β) and 95% confidence intervals (CI) estimated from linear mixed models specifying year of diagnosis predicting specified county-level characteristic. County-level proportion white and proportion Hispanic refer to the childhood population, ages 0-4 years, within counties. County-level proportion poverty refers to the entire population, all ages, within counties below the poverty line.

Table 2-2 Correlation matrix of county-level characteristics, SEER 9 counties (N=194), 1975-2013

	Pearson Correlations							
	1	2	3	4	5	6	7	**VIF**
1. Year of Diagnosis	1							1.84
2. Maternal Age	0.57	1						2.25
3. Birthweight	-0.27	0.13	1					2.66
4. Birth Order	-0.03	-0.12	-0.16	1				1.37
5. % White	-0.15	-0.17	0.34	0.07	1			1.52
6. % Hispanic	0.16	-0.21	-0.69	-0.08	-0.14	1		3.07
7. % Poverty	-0.04	-0.43	-0.49	0.24	-0.28	0.61	1	2.56

Notes: VIF = variance inflation factor.

Table 2-3 Associations between county-level characteristics and combined childhood cancer incidence, ages 0 to 4 years, SEER 9 counties (N=194), 1975-2013

	Adjusted Only for Year of Diagnosis			Fully Adjusted for All Characteristics		
County-Level Characteristic	RR	95% CI	P	RR	95% CI	P
Maternal Age (per 5-year increase)	1.22	(1.11, 1.34)	<.001	1.08	(0.97, 1.20)	.14
Birthweight (per 500-gram increase)	1.57	(1.37, 1.80)	<.001	1.17	(0.97, 1.41)	.10
Birth Order (per 1-unit increase)	0.86	(0.76, 0.97)	.015	0.93	(0.83, 1.05)	.26
% White (per 10% increase)	1.03	(1.01, 1.04)	.002	1.02	(1.00, 1.03)	.079
% Hispanic (per 10% increase)	0.97	(0.95, 0.98)	<.001	0.98	(0.97, 1.00)	.081
% Poverty (per 10% increase)	0.87	(0.83, 0.90)	<.001	0.95	(0.89, 1.02)	.16

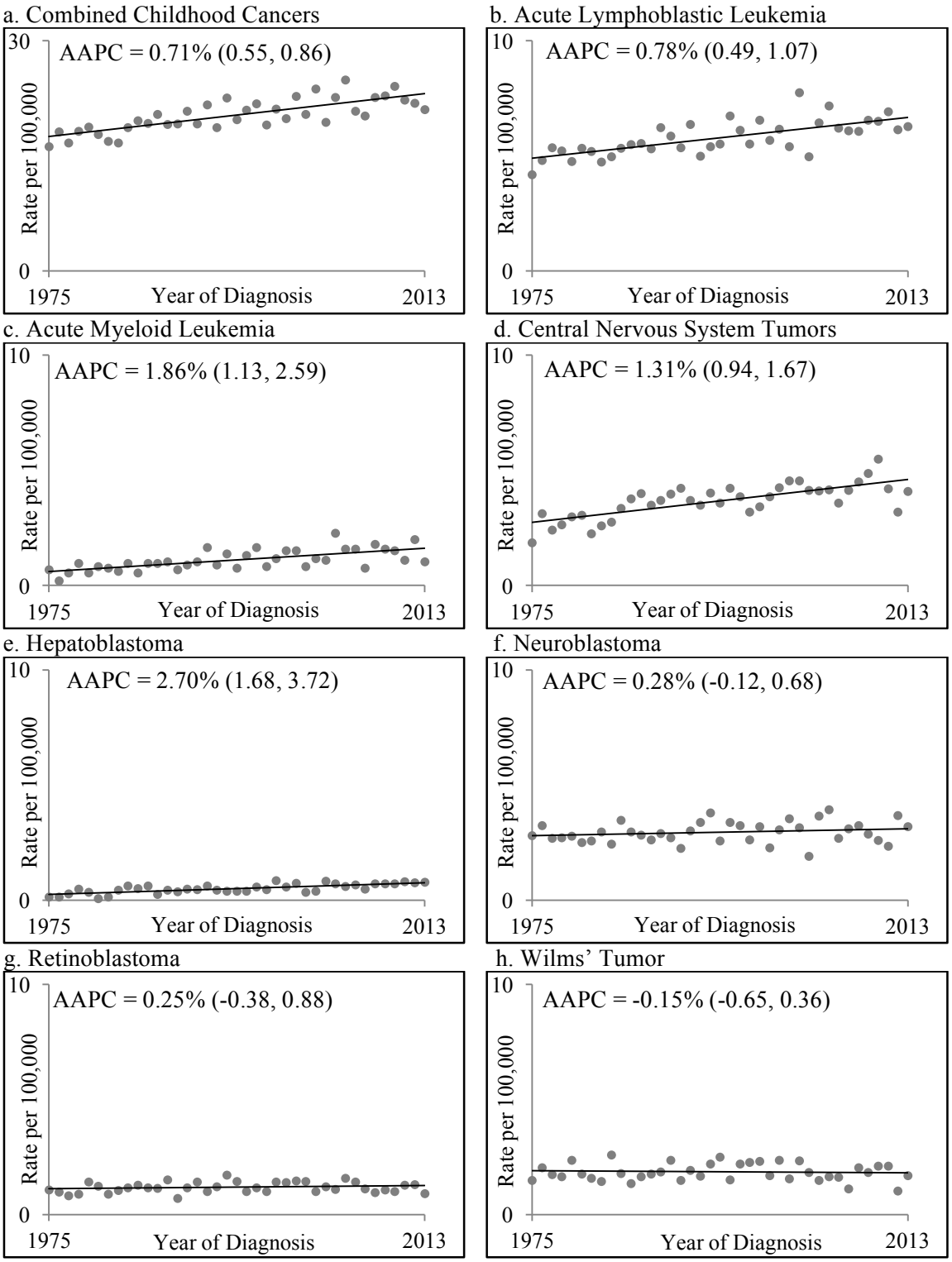
Notes: Rate ratios (RR) and 95% confidence intervals (CI) estimated from Poisson mixed models with a random intercept for repeated county-level measures over time. Fully adjusted models control for year of diagnosis and all county-level characteristics simultaneously.

Table 2-4 Associations between county-level characteristics and childhood cancer incidence, ages 0 to 4 years, by cancer type, SEER 9 counties (N=194), 1975-2013

	Combined Cancers				Acute Lymphoblastic Leukemia			
	Adjusted for Year of Diagnosis		Fully Adjusted for All Characteristics		Adjusted for Year of Diagnosis		Fully Adjusted for All Characteristics	
County-Level Characteristic	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Maternal Age	1.22	(1.11, 1.34)	1.08	(0.97, 1.20)	1.17	(0.99, 1.39)	1.03	(0.85, 1.25)
Birthweight	1.57	(1.37, 1.80)	1.17	(0.97, 1.41)	1.53	(1.15, 2.02)	1.37	(0.97, 1.91)
Birth Order	0.86	(0.76, 0.97)	0.93	(0.83, 1.05)	0.82	(0.67, 1.02)	0.90	(0.73, 1.11)
% White	1.03	(1.01, 1.04)	1.02	(1.00, 1.03)	1.06	(1.03, 1.09)	1.04	(1.01, 1.07)
% Hispanic	0.97	(0.95, 0.98)	0.98	(0.97, 1.00)	1.01	(0.99, 1.04)	1.05	(1.02, 1.08)
% Poverty	0.87	(0.83, 0.90)	0.95	(0.89, 1.02)	0.87	(0.80, 0.95)	0.92	(0.82, 1.03)
County-Level	Acute Myeloid Leukemia				CNS Tumors			
Characteristic	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Maternal Age	1.22	(0.85, 1.74)	1.34	(0.85, 2.11)	1.15	(0.96, 1.39)	0.98	(0.76, 1.26)
Birthweight	1.25	(0.68, 2.30)	0.91	(0.40, 2.06)	2.11	(1.55, 2.87)	1.65	(1.04, 2.60)
Birth Order	1.16	(0.78, 1.74)	1.21	(0.72, 2.03)	0.94	(0.74, 1.20)	1.11	(0.84, 1.45)
% White	0.99	(0.94, 1.03)	1.00	(0.93, 1.07)	1.05	(1.02, 1.08)	1.01	(0.97, 1.05)
% Hispanic	0.91	(0.85, 0.97)	0.90	(0.84, 0.97)	0.95	(0.92, 0.99)	0.99	(0.95, 1.03)
% Poverty	0.96	(0.79, 1.16)	1.08	(0.82, 1.43)	0.82	(0.74, 0.91)	0.90	(0.77, 1.06)
County-Level	Neuroblastoma				Retinoblastoma			
Characteristic	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Maternal Age	1.47	(1.22, 1.77)	1.16	(0.89, 1.52)	1.04	(0.79, 1.37)	1.06	(0.70, 1.59)
Birthweight	1.83	(1.32, 2.55)	1.07	(0.67, 1.70)	0.98	(0.63, 1.54)	0.91	(0.44, 1.87)
Birth Order	0.67	(0.53, 0.85)	0.81	(0.60, 1.10)	0.78	(0.54, 1.11)	0.62	(0.38, 1.00)
% White	1.01	(0.98, 1.04)	1.00	(0.96, 1.04)	0.98	(0.94, 1.01)	1.01	(0.95, 1.08)
% Hispanic	0.93	(0.89, 0.96)	0.94	(0.90, 0.98)	0.96	(0.91, 1.01)	0.94	(0.88, 1.00)
% Poverty	0.78	(0.70, 0.87)	0.90	(0.76, 1.07)	1.04	(0.91, 1.20)	1.20	(0.94, 1.53)
County-Level	Wilms' Tumor				Hepatoblastoma			
Characteristic	RR	95% CI	RR	95% CI	RR	95% CI	RR	95% CI
Maternal Age	1.35	(1.04, 1.75)	1.14	(0.81, 1.62)	1.12	(0.75, 1.67)	1.16	(0.63, 2.14)
Birthweight	1.61	(1.06, 2.46)	1.31	(0.70, 2.43)	1.52	(0.74, 3.10)	1.91	(0.65, 5.67)
Birth Order	0.93	(0.68, 1.28)	1.15	(0.79,1.68)	1.09	(0.64, 1.88)	1.43	(0.71, 2.90)
% White	1.00	(0.96, 1.04)	0.97	(0.92, 1.03)	1.03	(0.97, 1.09)	1.00	(0.90, 1.10)
% Hispanic	0.93	(0.88, 0.98)	0.95	(0.90, 1.01)	1.03	(0.96, 1.10)	1.06	(0.97, 1.16)
% Poverty	0.85	(0.74, 0.97)	0.90	(0.73, 1.12)	0.95	(0.76, 1.19)	1.00	(0.69, 1.46)

Notes: Rate ratios (RR) and 95% confidence intervals (CI) estimated from Poisson mixed models with a random intercept for repeated county-level measures over time. RR estimates a 5-year change in maternal age, 500-gram change in birthweight, 1-unit change in birth order, and 10% change in proportion white, Hispanic, and below the poverty line. Fully adjusted models control for year of diagnosis and all county-level characteristics simultaneously.

Figure 2-2 Temporal trends in childhood cancer incidence rates, ages 0 to 4 years, SEER 9 counties (N=194), 1975-2013



Notes: Annual incidence rates represent the cancer incidence rate among children, ages 0 to 4 years, within all 194 counties. Average annual percent change (AAPC) estimated from Poisson mixed models with a random intercept for repeated county-level measures over time.

Table 2-5 Crude and adjusted average annual percent change (AAPC) in childhood cancer incidence rates, ages 0 to 4 years, SEER 9 counties (N=194), 1975-2013

	Combined Cancers			ALL			AML			CNS			HB		
Model	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$
Crude	0.71	(0.55, 0.86)		0.78	(0.49, 1.07)		1.86	(1.13, 2.59)		1.31	(0.94, 1.67)		2.70	(1.68, 3.72)	
individually adjusted for:															
Maternal Age	0.32	(0.08, 0.56)	-55%	0.47	(0.02, 0.91)	-40%	1.48	(0.47, 2.49)	-20%	1.02	(0.50, 1.55)	-22%	2.48	(1.19, 3.78)	-8%
Birthweight	0.86	(0.69, 1.02)	21%	0.92	(0.61, 1.23)	18%	1.94	(1.17, 2.72)	4%	1.56	(1.18, 1.95)	20%	2.85	(1.80, 3.92)	6%
Birth Order	0.71	(0.55, 0.86)	0%	0.77	(0.48, 1.06)	0%	1.86	(1.12, 2.59)	0%	1.31	(0.94, 1.67)	0%	2.70	(1.68, 3.72)	0%
% White	0.80	(0.63, 0.96)	12%	0.98	(0.68, 1.28)	26%	1.80	(1.05, 2.56)	-3%	1.47	(1.09, 1.85)	12%	2.79	(1.76, 3.83)	3%
% Hispanic	0.86	(0.69, 1.03)	21%	0.72	(0.40, 1.03)	-8%	2.28	(1.49, 3.08)	23%	1.51	(1.11, 1.91)	15%	2.57	(1.50, 3.65)	-5%
% Poverty	0.83	(0.67, 0.99)	17%	0.89	(0.59, 1.19)	15%	1.90	(1.14, 2.66)	2%	1.47	(1.09, 1.85)	12%	2.74	(1.71, 3.78)	2%
Fully Adjusted	0.75	(0.48, 1.03)	6%	0.81	(0.31, 1.31)	4%	1.62	(0.38, 2.87)	-13%	1.69	(1.03, 2.35)	29%	2.36	(0.71, 4.04)	-12%

Notes: Average annual percent change (AAPC) estimated from Poisson mixed models with a random intercept for repeated county-level measures over time. Crude models specify year of diagnosis predicting county incidence rates. Individually adjusted models further specify each county-level characteristic, modeled one at a time. Fully adjusted models specify year of diagnosis and all county-level characteristics simultaneously. The % change compares crude to adjusted AAPCs. ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; CNS = combined central nervous system tumors; HB = hepatoblastoma.

Table 2-6 Sensitivity analysis by county size: crude and adjusted average annual percent change (AAPC) in childhood cancer incidence rates, ages 0 to 4 years, SEER 9 counties (N=194), 1975-2013

	Entire Sample (N=194)			Population < 10,000 (N=162)			Population < 5,000 (N=144)			Population < 1,000 (N=70)		
Model	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$
Crude	0.71	(0.55, 0.86)		0.69	(0.34, 1.03)		0.53	(0.09, 0.98)		0.83	(-0.15, 1.82)	
individually adjusted for:												
Maternal Age	0.32	(0.08, 0.56)	-55%	0.18	(-0.26, 0.63)	-73%	0.03	(-0.54, 0.60)	-94%	0.06	(-1.21, 1.34)	-93%
Birthweight	0.86	(0.69, 1.02)	21%	0.88	(0.52, 1.25)	28%	0.73	(0.25, 1.21)	36%	1.00	(-0.05, 2.07)	21%
Birth Order	0.71	(0.55, 0.86)	0%	0.68	(0.34, 1.03)	-1%	0.53	(0.09, 0.98)	0%	0.78	(-0.20, 1.77)	-6%
% White	0.80	(0.63, 0.96)	12%	0.80	(0.45, 1.15)	16%	0.60	(0.14, 1.07)	13%	0.75	(-0.41, 1.92)	-10%
% Hispanic	0.86	(0.69, 1.03)	21%	0.84	(0.48, 1.19)	21%	0.67	(0.22, 1.13)	26%	0.95	(-0.04, 1.95)	14%
% Poverty	0.83	(0.67, 0.99)	17%	0.72	(0.37, 1.06)	4%	0.53	(0.09, 0.97)	-1%	0.74	(-0.26, 1.75)	-11%
Fully Adjusted	0.75	(0.48, 1.03)	6%	0.58	(0.08, 1.08)	-16%	0.31	(-0.39, 1.01)	-42%	-0.41	(-2.12, 1.33)	-150%

Notes: Average annual percent change (AAPC) estimated from Poisson mixed models with a random intercept for repeated county-level measures over time. Crude models specify year of diagnosis predicting county incidence rates. Individually adjusted models further specify each county-level characteristic, modeled one at a time. Fully adjusted models specify year of diagnosis and all county-level characteristics simultaneously. The % change compares crude to adjusted AAPCs.

Table 2-7 Sensitivity analysis restricted to white children: crude and adjusted average annual percent change (AAPC) in childhood cancer incidence rates, ages 0 to 4 years, SEER 9 counties (N=194), 1975-2013

	Combined Cancers			ALL			AML			CNS			HB		
Model	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$	AAPC	95% CI	$\%_{change}$
Crude	0.73	(0.56, 0.90)		0.97	(0.66, 1.28)		1.52	(0.70, 2.35)		1.35	(0.95, 1.75)		2.83	(1.69, 3.98)	
individually adjusted for:															
Maternal Age	0.38	(0.14, 0.62)	-48%	0.81	(0.39, 1.23)	-16%	1.19	(0.09, 2.30)	-22%	1.19	(0.66, 1.72)	-12%	2.50	(1.09, 3.92)	-12%
Birthweight	0.83	(0.65, 1.01)	14%	1.00	(0.67, 1.32)	3%	1.71	(0.84, 2.58)	12%	1.52	(1.10, 1.94)	12%	2.92	(1.73, 4.13)	3%
Birth Order	0.73	(0.56, 0.90)	0%	0.97	(0.66, 1.28)	0%	1.52	(0.70, 2.35)	0%	1.35	(0.95, 1.75)	0%	2.83	(1.69, 3.99)	0%
Fully Adjusted	0.61	(0.34, 0.88)	-16%	0.93	(0.44, 1.42)	-4%	1.40	(0.13, 2.69)	-8%	1.54	(0.91, 2.18)	14%	2.54	(0.88, 4.23)	-10%

Notes: Average annual percent change (AAPC) estimated from Poisson mixed models with a random intercept for repeated county-level measures over time. Crude models specify year of diagnosis predicting county incidence rates. Individually adjusted models further specify each county-level characteristic, modeled one at a time. Fully adjusted models specify year of diagnosis and all county-level characteristics simultaneously. The % change compares crude to adjusted AAPCs. ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; CNS = combined central nervous system tumors; HB = hepatoblastoma.

Chapter 3

Manuscript 2: Socioeconomic status and childhood cancer incidence: a population-based multilevel analysis

Introduction

Socioeconomic status (SES) is consistently linked to a range of health outcomes in children and adults including many adulthood cancers.[44-46] SES may also be associated with childhood cancer incidence operating through various mechanisms at both the individual and area levels. Individual-level SES may influence incidence through mediators such as parental occupational exposures, dietary patterns, infectious agents, family reproductive decisions such as maternal age and family size, and birth outcomes such as birthweight.[3,63] SES may also operate as a mediator of other preceding demographic social determinants of health including foreign birthplace or race/ethnicity.[3,60] For example, a recent study reported lower risk of several cancers among Hispanic children of foreign born compared to US born mothers.[60] Area-level SES may independently influence risk through mediators such as environmental pollutants and toxins, infectious agents, and social norms regarding lifestyle behaviors.[50,64] Thus, in-depth knowledge of the association between SES and childhood cancer risk may provide etiologic insight into the role of environmental and behavioral exposures, particularly since the causes of childhood cancer are not well understood.

Empirical evidence of an association between SES and childhood cancer incidence remains limited and inconclusive. While ecologic findings from international

and within-country small area studies suggest an adverse association between higher SES and incidence of some childhood cancers,[50-54] individual-level studies largely report null associations.[43,55-59] However, there are several limitations to previous work conducted at the individual-level. First, studies have predominantly focused on childhood leukemia due to its higher incidence and suspected infectious etiology.[13,47,48] Given that different cancers likely have different etiologies,[5] it is important to investigate associations of SES with other non-leukemic cancers. Second, a case-control design is commonly used to test associations due to the rarity of childhood cancers. Because controls in studies requiring active participation tend to be higher SES than the source population of interest,[116] participation-based case-control studies can produce biased estimators. Third, established risk factors associated with SES, such as birthweight and maternal age,[3] have not been consistently controlled for across studies. This hinders cross-study comparisons and potentially obscures underlying mechanisms contributing to an SES association. Finally, no study, to our knowledge, has used multilevel methods to test for independent associations of SES at the individual and small area levels. Without a multilevel approach, it is possible that SES at one level is merely a proxy for SES at the other level, thus masking etiology. In this study, we addressed these limitations by leveraging registry data in a population-based case-cohort study. We assessed SES at both the individual and neighborhood level and accounted for established risk factors that may confound or mediate associations of SES and childhood cancer incidence.

Methods

Study Population

We ascertained cases, diagnosed ages 0 to 14 years, from the Minnesota Cancer Surveillance System (MCSS), which is estimated to have 99.7% cancer case completeness and 96.5% overall data accuracy.[117] We restricted our sample to cases born between 1989 (when residential addresses were first recorded on birth certificates) and 2014, with a linked Minnesota birth record (86%). Records were linked based on first and last name, date of birth, and social security number (when available) through probabilistic record linkage using LinkPlus software.[118,119] We then implemented a case-cohort study design in which we randomly sampled four controls per case, frequency matched on birth year, from the Minnesota birth registry without regard to case status.[120] To rule out children with higher penetrance genetic syndromes,[121,122] we excluded 20 cases and 5 controls with Down syndrome and 11 cases with multiple primary tumors, resulting in a final analytic sample of 2,947 cases and 11,907 controls (N=14,854).

Outcome

We assessed incidence of combined (all diagnoses) and individual childhood cancers based on the International Classification of Childhood Cancer, third edition.[113] We assessed incidence of acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), lymphomas, central nervous system (CNS) tumors, neuroblastoma, retinoblastoma, Wilms' tumor, hepatoblastoma, and rhabdomyosarcoma. Bone sarcomas and germ cell tumors were not assessed individually because these cancers are most common in adolescents.

Measures of SES

Individual-level SES – We used maternal education to measure individual-level SES. The validity of using maternal education as a measure of childhood SES has been demonstrated previously,[123,124] and is commonly used as a reliable measure of SES in US birth registry studies.[125] Between 1989 and 2010, maternal education was recorded on birth certificates as years of schooling, ranging from 0 to 17. Beginning in 2011, education was recorded as highest degree earned, ranging from (1) 8th grade or less to (8) doctorate or professional degree. We grouped education into four categories to ensure sufficient sample size and consistency across years: (1) < 12 years of schooling or less than high school diploma, (2) 12 years of schooling or high school diploma, (3) 13-15 years of schooling or some college or associates degree, and (4) 16+ years of schooling or bachelor's degree or higher. After confirming linearity (Figure 3-1, Panel A), we modeled education ordinally.

Neighborhood-level SES – Residential addresses were abstracted from birth records and geocoded to census tracts using normalized 2010 geographic boundaries.[126] We created a SES index derived from the first component score from a nationwide principal components analysis (PCA)[127] of 5 tract-level US Census variables: % poverty, % on welfare or public assistance, % of those aged 16+ unemployed, % female headed households with children, and % of those aged 25+ with less than a high school education.[128,129] A separate PCA was performed for each decade (1980 to 2010) using decennial census data, 1980-2000, and American Community Survey (ACS) data, 2005-2009, for 2010. We assigned values based on year of birth and linearly interpolated SES

scores for intercensal years. High internal consistency was observed for each decade of data (Cronbach's alpha range: .89–.92).[129] We confirmed linearity (Figure 3-1, Panel B) and standardized scores so that a one-unit change equates to one standard deviation; higher index values indicate higher neighborhood SES.

Statistical Analysis

We compared demographic, socioeconomic, and birth characteristics of cases and controls using Pearson's chi-square test for categorical measures and the two-sample t-test for continuous measures. To untangle associations of SES and cancer incidence (combined and by individual type), we tested logistic mixed models with a random intercept for clustering within census tracts. Intraclass correlations (ICC) estimated from unimputed bivariate logistic mixed models revealed minimal tract-level clustering across cancer types (ICC <0.08) except rhabdomyosarcoma (ICC = 0.29), though ICCs for rare outcomes may be unreliable.[130] We tested two sets of models, one specifying maternal education as the primary predictor ("A" models), the other specifying neighborhood SES as the primary predictor ("B" models). Model 1 tested bivariate (crude) associations between SES and cancer incidence. Model 2 tested these associations adjusted for maternal race/ethnicity (non-Hispanic white versus otherwise). Model 3 further adjusted for pregnancy-related factors previously associated with childhood cancer risk including maternal age (years),[3,24] birthweight (grams; values <350 grams considered implausible and recoded to missing),[3,111] gestational age (weeks; values <20 or >45 weeks considered implausible and recoded to missing),[3,28] birth order (first born versus higher),[3,29] birth year,[3] and sex.[3] Model 4 tested the independent association of SES at each level of

exposure by simultaneously fitting models with both maternal education and neighborhood SES, along with previously specified covariates.

To further characterize the utility of covariate adjustment in studies of childhood cancer etiology, we tested additional models probing associations of SES and established risk factors. First, for cancers in which covariate adjustment attenuated SES effect estimates, we tested trivariate models of SES predicting cancer incidence, adjusting for each covariate, previously specified in Model 3, one at a time. We then calculated the percent change in estimated SES-cancer incidence associations between bivariate (Model 1) and trivariate models (% change = $(\beta_{bivar} - \beta_{trivar})/ \beta_{bivar}$). Second, to assess the utility of adjusting for SES in prediction models, we compared effect estimates of race/ethnicity, maternal age, birthweight, gestational age, and birth order from fully adjusted models (Model 4) to estimates from models specifying all covariates expect the two SES measures (Model 5). We used multiple imputation by chained equations to impute missing data.[131] Statistical significance was determined as $p < .05$ for a 2-sided hypothesis test. Multiple imputation and analyses were performed using Stata 14.2 (College Station, TX).[114]

Results

Descriptive characteristics of cases and controls are compared in Table 3-1. Cases were more likely to be male, higher birthweight, and exhibited slightly shorter gestation compared to controls. Mothers of cases were older, had higher education levels, and were more likely to be non-Hispanic white than controls. Cases had higher neighborhood SES

than controls. The two SES measures displayed moderate correlation (Spearman ρ=0.35). Additional variable correlations are available in Table 3-2.

In Table 3-3, we present associations of maternal education and childhood cancer incidence. In crude Model 1A, a one-step increase in maternal education (e.g. from high school graduate to some college) was associated with an 8% increase in risk of combined childhood cancers (OR, 1.08; 95% CI, 1.04, 1.13; p <.001). Statistically significant crude adverse associations with higher maternal education were also observed for ALL (OR, 1.10; 95% CI, 1.02, 1.19, p = .018), CNS tumors (OR, 1.12; 95% CI, 1.04, 1.21; p = .005), and neuroblastoma (OR, 1.15; 95% CI, 1.02, 1.30, p = .028). Crude associations were elevated and marginally statistically significant, for lymphomas (OR, 1.11; 95% CI, 0.99, 1.25; p = .066) and Wilms' tumor (OR, 1.15; 95% CI, 1.00, 1.32; p = .057). After adjusting for race/ethnicity (Model 2A) and pregnancy-related risk factors (Model 3A), adverse associations of higher SES were attenuated towards the null and no longer statistically significant. Further adjustment for neighborhood SES (Model 4A) did not substantively alter non-significant estimates. An elevated odds ratio was observed for maternal education predicting retinoblastoma incidence, though confidence intervals were wide (e.g. Model 4A: OR, 1.25; 95% CI, 0.94, 1.65; p = .12).

Conversely, we found a statistically significant protective association between higher SES and hepatoblastoma incidence. In crude Model 1A, a one-step increase in maternal education was associated with a 28% reduced risk of hepatoblastoma (OR, 0.72; 95% CI, 0.54, 0.94; p = .017). This association was robust to comprehensive covariate adjustment (Model 4A: OR, 0.70; 95% CI, 0.51, 0.98; p = .037). A protective, marginally

statistically significant, association was also observed between higher maternal education and rhabdomyosarcoma incidence (Model 4A: OR, 0.77; 95% CI, 0.59, 1.00; p = .054).

For cancers substantively impacted by covariate control (combined, ALL, lymphomas, CNS tumors, neuroblastoma, Wilms' tumor), we further investigated which of the established risk factors accounted for associations (Table 3-4). Adjustment for race/ethnicity reduced the estimated effect of maternal education on cancer incidence by more than 10% for combined and individual childhood cancers except lymphomas (6%); the largest reductions were observed for combined childhood cancers (30%), Wilms' tumor (46%), and CNS tumors (50%). Adjustment for maternal age reduced the estimated maternal education effect by >10% for combined childhood cancers (20%), ALL (18%), and lymphomas (20%); adjustment for birthweight reduced the effect by >10% for ALL (12%) and Wilms' tumor (13%); and adjustment for birth year reduced the effect by >10% for neuroblastoma (19%). Individual adjustment for gestational age, parity, and sex had little impact on maternal education effect estimates.

We then evaluated associations between neighborhood SES and childhood cancer incidence (Table 3-5). In crude models (Model 1B), higher neighborhood SES was adversely associated with incidence of combined childhood cancers (OR, 1.09; 95% CI, 1.04, 1.14; p <.001), ALL (OR, 1.10; 95% CI, 1.01, 1.20; p = .026), lymphomas (OR, 1.14; 95% CI, 1.01, 1.30; p = .037), and CNS tumors (OR, 1.18; 95% CI, 1.08, 1.29; p <.001). Neighborhood SES was no longer associated with incidence of combined or individual cancers in models adjusted for race/ethnicity and pregnancy-related risk factors (Models 3B), or in models further adjusted for maternal education (Model 4B).

Among cancers impacted by covariate control (combined, ALL, lymphomas, CNS tumors), adjustment for race/ethnicity reduced the estimated effect of neighborhood SES on cancer incidence by >10% for combined childhood cancers (39%), ALL (18%), and CNS tumors (49%); adjustment for maternal age reduced the effect by >10% for combined childhood cancers (14%), ALL (15%), and lymphomas (12%); and adjustment for birthweight reduced the effect for ALL by 10% (Table 3-6).

As illustrated in Table 3-7, effect estimates of established demographic and pregnancy-related risk factors predicting combined childhood cancer incidence were not considerably altered by adjustment for maternal education and neighborhood SES. Associations for individual types of childhood cancer are provided in Table 3-8.

Discussion

This is the first study to employ multilevel methods to examine associations between SES and childhood cancer incidence. We tested associations in a population-based sample using a registry-based case-cohort study design, thus minimizing multiple potential sources of bias. Through our analysis, we generated several findings. First, we found that higher SES at both levels (maternal education and census tract composite index) was consistently adversely associated with incidence of many childhood cancers in crude models. Second, these adverse associations were accounted for by established demographic and pregnancy-related risk factors. Third, associations between established risk factors and childhood cancer incidence were robust to adjustment for SES. Fourth, unlike other cancers, higher individual-level SES was significantly associated with lower risk of hepatoblastoma, even after comprehensive control of other cancer risk factors.

Higher SES, whether operationalized as maternal education at the individual-level or as an area-level index, was associated with higher risk of the most common childhood cancers (combined, ALL, lymphomas, and CNS tumors), with a similar pattern emerging for both SES measures. However, no statistically significant adverse associations between higher SES and incidence of combined or individual childhood cancers remained after accounting for established demographic and pregnancy-related risk factors. This suggests that crude associations of SES at either level primarily capture established risk factors not specified in the model. In particular, non-Hispanic white race/ethnicity accounted for a substantial portion of adverse associations of higher SES, especially for combined childhood cancers, CNS tumors, and Wilms' tumor. Older maternal age and higher birthweight also explained some of the adverse SES association, though to varying degrees across cancer types.

Further investigation of associations between established risk factors (e.g. race/ethnicity and birthweight) and childhood cancer incidence revealed that effect estimates were robust to adjustment for SES, suggesting that SES has limited impact on incidence of most childhood cancers beyond the social patterning of these exposures. Therefore, adjustment for proximal demographic and pregnancy-related risk factors is likely sufficient in analyses of childhood cancer etiology. This is reassuring, especially given the lack of socioeconomic data in medical records and cancer registries,[49] which are common data sources for childhood cancer research. However, it remains important to note that SES is a common prior cause of some of these more proximal risk factors including birthweight and maternal age. Therefore, these exposures could be considered

mediators and although adjusting for them (as we do here in Models 3-4) is a conservative approach and typical in the clinical epidemiologic literature, it may understate the role of SES in childhood cancer incidence.

We did identify a statistically significant protective association between higher maternal education and hepatoblastoma incidence that was independent of established risk factors. Though little is known about the etiology of hepatoblastoma, some studies have identified parental tobacco use as a potential risk factor,[132] which may contribute to a social patterning of incidence. Due to insufficient smoking data in our study, we look to future studies to explore this potential mechanism. We note that a similarly designed study using pooled data from five US state registries found no evidence of a protective association between higher maternal education and risk of hepatic tumors.[43] Therefore, additional research is needed to replicate our finding and to explore potential effect modifiers that may explain cross-place differences.

Limitations

This study has several limitations. First, besides education, we could not account for other dimensions of individual-level SES, such as household income or occupation. Nevertheless, given the lack of individual-level socioeconomic data in cancer registries,[49] our use of maternal education data from birth records improves upon many population-based cancer studies. Second, because our measures of SES were limited to a single exposure window at birth, although a common approach for operationalizing childhood SES, we cannot draw conclusions about SES later in childhood. Yet neighborhood SES at birth was strongly correlated with SES at diagnosis among cases (Pearson $\rho = 0.70$),

suggesting temporal stability, which has been reported previously.[112,133] Third, like many prior studies, we were limited by sample size due to the rarity of childhood cancers. This prevented us from testing more homogenized cancer subgroups or stratifying by age. It is notable however that some SES associations were strong enough to be detected with only a case sample of 50 (i.e. hepatoblastoma). Fourth, there is the potential for disease misclassification bias if children born in Minnesota subsequently moved out of state and developed cancer. However, given the rarity of childhood cancers and low out-migration rate among Minnesota youth,[134] this is not a major threat to validity. There is also the potential for selection bias if cases without a matching birth record were in fact born in Minnesota, and thus part of the source cohort. This may have occurred because of inconsistent (e.g. name changes) or missing data. We found that unmatched cases resided in lower SES neighborhoods at time of diagnosis than matched cases ($p < .001$), which may reflect differences in data quality by SES.[125] However, it may also reflect higher residential mobility among lower SES cases,[112] and thus a higher in-migration rate to Minnesota during childhood. Our 86% record linkage rate is comparable to prior registry-based studies of childhood cancer.[135-137] Finally, given the high proportion of non-Hispanic whites in our sample (82%), results may be less generalizable to more racially diverse populations. Despite these limitations, this study provides important insight into unpacking the association between SES and childhood cancer incidence.

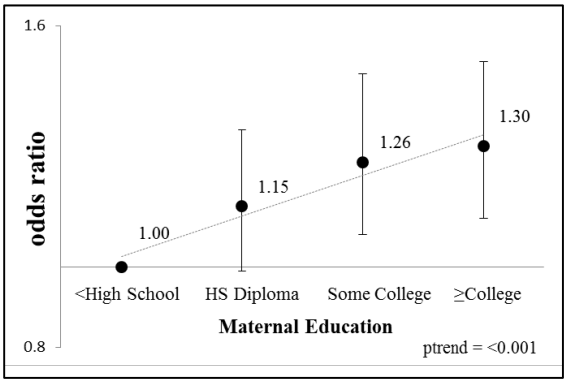
Conclusion

Results from this study suggest that SES has a limited impact on childhood cancer incidence, beyond the social patterning of established demographic and pregnancy-

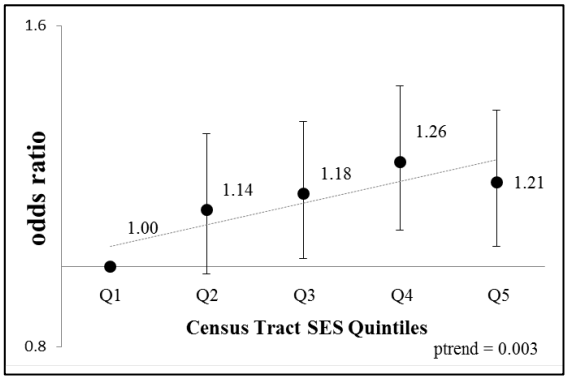
related risk factors including race/ethnicity, maternal age, and birthweight. It is reassuring that these socially patterned risk factors of childhood cancer incidence are already known and well described in the literature. However, given that these exposures only account for a small portion of the total disease burden,[3] more work is needed to better understand childhood cancer etiology. Unfortunately, while it is important to continue monitoring socioeconomic differences in risk to ensure health equity, our findings suggest that continued investigation of SES associations may generate limited new etiologic insight into childhood cancer incidence, at least for the more common cancers.

Figure 3-1 Crude associations between categorical SES and combined childhood cancer incidence; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

A. Individual-level SES



B. Neighborhood-level SES



Notes: Odds ratios estimated from bivariate logistic mixed models of categorical SES (maternal education/neighborhood deprivation) predicting combined childhood cancer incidence (all diagnoses). Reference = lowest SES category (<high school (HS)/Q1 neighborhood). Standardized census tract SES principal component scores categorized into quintiles; Q1 = low SES (high deprivation), Q5 = high SES (low deprivation).

Table 3-1 Characteristics of cases and controls; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

Variable	Controls (N=11,907)	Cases (N=2,947)	P	% Missing
Birth Characteristics				
Birth Year, mean (SD)	1999 (6.5)	1999 (6.5)	.87	0.0
Female, No. (%)	5,861 (49.2)	1,306 (44.3)	<.001	0.0
Birthweight, mean (SD), g	3,403 (585)	3,437 (608)	.005	0.1
Gestational Age, mean (SD), wks	38.9 (2.0)	38.8 (2.2)	.014	2.2
First Born, No. (%)	4,587 (38.7)	1,169 (39.8)	.27	0.5
Maternal Characteristics				
Age at Delivery, mean (SD), y	28.1 (5.8)	28.5 (5.7)	.002	0.5
Non-Hispanic White, No. (%)	9,631 (81.4)	2,496 (85.0)	<.001	0.6
Education, No. (%)			.002	2.4
< high school diploma	1,265 (10.9)	259 (9.0)		
high school diploma	3,439 (29.6)	808 (28.0)		
some college	3,106 (26.8)	806 (27.9)		
≥ bachelor's degree	3,799 (32.7)	1,013 (35.1)		
Neighborhood SES, mean (SD)	-0.02 (1.02)	0.06 (0.92)	<.001	0.0

Notes: P-values compare characteristics of cases versus controls estimated from Pearson's chi-square test for categorical measures and the two-sample t-test for continuous measures. SD = standard deviation; g = grams; wks = weeks; y = years; SES = socioeconomic status (higher values = higher SES).

Table 3-2 Correlation matrix; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

Variable	1	2	3	4	5	6	7	8	9
1. Birth Year	1								
2. Female	-0.02	1							
3. Mother NH White	-0.16	0.01	1						
4. Maternal Age	0.05	0.01	0.18	1					
5. Birthweight	-0.04	-0.09	0.11	0.08	1				
6. Gestational Age	-0.08	0.03	0.02	-0.03	0.63	1			
7. First Born	0.01	0.02	0.03	-0.32	-0.09	0.02	1		
8. Maternal Education	0.10	0.00	0.28	0.45	0.10	-0.02	0.02	1	
9. Neighborhood SES	-0.01	0.00	0.41	0.26	0.09	0.00	0.00	0.35	1

Notes: Pearson correlation estimated continuous/continuous associations; Spearman correlation estimated ordinal/continuous and ordinal/dichotomous associations; tetrachoric correlation estimated dichotomous/dichotomous associations; and biserial correlation estimated dichotomous/continuous associations. NH = non-Hispanic; SES = socioeconomic status (higher values = higher SES).

Table 3-3 Associations between maternal education and childhood cancer incidence; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

	All Childhood Cancers Combined N = 2,947			Acute Lymphoblastic Leukemia N = 673			Acute Myeloid Leukemia N = 112			Lymphomas N = 311			Central Nervous System Tumors N = 662		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Model 1A	1.08	(1.04, 1.13)	<.001	1.10	(1.02, 1.19)	.018	1.03	(0.86, 1.25)	.73	1.11	(0.99, 1.25)	.066	1.12	(1.04, 1.21)	.005
Model 2A	1.06	(1.01, 1.10)	.014	1.08	(1.00, 1.18)	.053	1.06	(0.87, 1.29)	.55	1.11	(0.98, 1.25)	.096	1.06	(0.97, 1.15)	.18
Model 3A	1.03	(0.98, 1.08)	.26	1.05	(0.96, 1.16)	.27	0.96	(0.78, 1.20)	.75	1.12	(0.98, 1.28)	.099	1.05	(0.96, 1.16)	.29
Model 4A	1.02	(0.97, 1.07)	.36	1.05	(0.95, 1.15)	.36	0.95	(0.76, 1.19)	.66	1.10	(0.96, 1.26)	.18	1.04	(0.95, 1.15)	.42

	Neuroblastoma N = 267			Retinoblastoma N = 80			Wilms' Tumor N = 210			Hepatoblastoma N = 50			Rhabdomyosarcoma N = 79		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Model 1A	1.15	(1.02, 1.30)	.028	1.15	(0.91, 1.44)	.23	1.15	(1.00, 1.32)	.057	0.72	(0.54, 0.94)	.017	0.91	(0.73, 1.14)	.41
Model 2A	1.13	(0.99, 1.29)	.064	1.21	(0.96, 1.53)	.11	1.08	(0.93, 1.25)	.32	0.74	(0.55, 0.98)	.038	0.85	(0.67, 1.08)	.19
Model 3A	1.07	(0.93, 1.24)	.34	1.27	(0.97, 1.68)	.088	1.02	(0.86, 1.20)	.85	0.69	(0.50, 0.96)	.028	0.78	(0.60, 1.01)	.061
Model 4A	1.09	(0.94, 1.26)	.28	1.25	(0.94, 1.65)	.12	1.02	(0.86, 1.21)	.83	0.70	(0.51, 0.98)	.037	0.77	(0.59, 1.00)	.054

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from logistic mixed models with a random intercept for clustering within census tracts. Model 1A: bivariate (crude). Model 2A: adjusted for maternal race/ethnicity. Model 3A: further adjusted for birth year, sex, maternal age, birthweight, gestational age, and birth order. Model 4A: further adjusted for neighborhood SES. Maternal education modeled as an ordinal variable; values range from 1 = <12 years, no high school diploma to 4 = 16+ years, advanced degree. Random intercepts dropped from Models 1A-3A for AML and Wilms' tumor due to non-convergence stemming from zero-value intraclass correlations. Control sample: N=11,907.

Table 3-4 Associations between maternal education and childhood cancer incidence estimated from bivariate and trivariate models; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

	All Cancers Combined N = 2,947				Acute Lymphoblastic Leukemia N = 673				Lymphomas N = 311			
Model	OR	95% CI	P	%Change	OR	95% CI	P	%Change	OR	95% CI	P	%Change
Bivariate	1.08	(1.04, 1.13)	<.001		1.10	(1.02, 1.19)	.018		1.11	(0.99, 1.25)	.066	
Trivariate *adjusted for:*												
Race/Ethnicity	1.06	(1.01, 1.10)	.014	30%	1.08	(1.00, 1.18)	.053	15%	1.11	(0.98, 1.25)	.096	6%
Maternal Age	1.06	(1.02, 1.11)	.008	20%	1.08	(0.99, 1.18)	.082	18%	1.09	(0.96, 1.24)	.19	20%
Birthweight	1.07	(1.03, 1.12)	.001	6%	1.09	(1.00, 1.18)	.039	12%	1.11	(0.99, 1.24)	.082	5%
Gestational Age	1.08	(1.04, 1.13)	<.001	-1%	1.10	(1.02, 1.19)	.017	0%	1.11	(0.99, 1.25)	.067	1%
First Born	1.08	(1.04, 1.13)	<.001	0%	1.10	(1.02, 1.19)	.018	0%	1.11	(0.99, 1.25)	.063	-1%
Sex	1.08	(1.04, 1.13)	<.001	-1%	1.10	(1.02, 1.19)	.017	-1%	1.11	(0.99, 1.25)	.063	-1%
Birth Year	1.08	(1.04, 1.13)	<.001	-1%	1.10	(1.02, 1.19)	.017	-1%	1.13	(1.01, 1.27)	.039	-13%
	Central Nervous System Tumors N = 662				Neuroblastoma N = 267				Wilms' Tumor N = 210			
Model	OR	95% CI	P	%Change	OR	95% CI	P	%Change	OR	95% CI	P	%Change
Bivariate	1.12	(1.04, 1.21)	.005		1.15	(1.02, 1.30)	.028		1.15	(1.00, 1.32)	.057	
Trivariate *adjusted for:*												
Race/Ethnicity	1.06	(0.97, 1.15)	.18	50%	1.13	(0.99, 1.29)	.064	12%	1.08	(0.93, 1.25)	.32	46%
Maternal Age	1.13	(1.03, 1.23)	.008	-5%	1.15	(1.00, 1.32)	.046	-2%	1.14	(0.97, 1.33)	.12	8%
Birthweight	1.11	(1.03, 1.20)	.010	8%	1.15	(1.02, 1.30)	.026	-2%	1.13	(0.98, 1.30)	.099	13%
Gestational Age	1.12	(1.03, 1.21)	.005	0%	1.15	(1.02, 1.30)	.026	-2%	1.15	(1.00, 1.33)	.055	-1%
First Born	1.12	(1.03, 1.21)	.005	1%	1.15	(1.01, 1.30)	.029	1%	1.15	(0.99, 1.32)	.059	1%
Sex	1.12	(1.04, 1.21)	.005	0%	1.15	(1.02, 1.30)	.026	-1%	1.15	(1.00, 1.32)	.057	0%
Birth Year	1.13	(1.04, 1.22)	.004	-4%	1.12	(0.99, 1.27)	.075	19%	1.13	(0.98, 1.31)	.083	9%

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from logistic mixed models with a random intercept for census tract clustering. Bivariate models equate to Model 1A in Table 3-3. Trivariate models adjusted for maternal race/ethnicity equate to Model 2A in Table 3-3. Random intercepts dropped from Wilms' tumor models due to non-convergence stemming from zero-value intraclass correlations. % change compares beta coefficients of maternal education predicting cancer incidence estimated from bivariate and trivariate models (i.e. βbivar-βtrivar/βbivar). Control sample: N=11,907.

Table 3-5 Associations between neighborhood SES and childhood cancer incidence; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

	All Childhood Cancers Combined N = 2,947			Acute Lymphoblastic Leukemia N = 673			Acute Myeloid Leukemia N = 112			Lymphomas N = 311			Central Nervous System Tumors N = 662		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Model 1B	1.09	(1.04, 1.14)	<.001	1.10	(1.01, 1.20)	.026	1.07	(0.88, 1.30)	.52	1.14	(1.01, 1.30)	.037	1.18	(1.08, 1.29)	<.001
Model 2B	1.05	(1.00, 1.11)	.035	1.08	(0.99, 1.19)	.095	1.12	(0.91, 1.39)	.30	1.15	(1.00, 1.31)	.052	1.09	(0.99, 1.20)	.091
Model 3B	1.04	(0.99, 1.09)	.16	1.06	(0.97, 1.17)	.22	1.07	(0.86, 1.33)	.54	1.15	(1.00, 1.32)	.058	1.09	(0.98, 1.20)	.11
Model 4B	1.03	(0.98, 1.09)	.21	1.05	(0.96, 1.16)	.29	1.08	(0.87, 1.34)	.49	1.13	(0.98, 1.30)	.10	1.08	(0.97, 1.19)	.15
	Neuroblastoma N = 267			Retinoblastoma N = 80			Wilms' Tumor N = 210			Hepatoblastoma N = 50			Rhabdomyosarcoma N = 79		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Model 1B	1.01	(0.89, 1.15)	.82	1.07	(0.85, 1.36)	.55	1.13	(0.96, 1.32)	.14	0.83	(0.66, 1.03)	.089	1.14	(0.87, 1.50)	.34
Model 2B	0.97	(0.84, 1.11)	.67	1.16	(0.90, 1.50)	.25	1.01	(0.85, 1.20)	.90	0.86	(0.67, 1.11)	.24	1.06	(0.79, 1.42)	.71
Model 3B	0.95	(0.82, 1.09)	.43	1.16	(0.90, 1.51)	.25	0.98	(0.83, 1.17)	.85	0.88	(0.68, 1.14)	.33	1.03	(0.76, 1.40)	.83
Model 4B	0.93	(0.81, 1.08)	.35	1.13	(0.87, 1.46)	.37	0.98	(0.82, 1.17)	.83	0.92	(0.70, 1.20)	.52	1.08	(0.79, 1.48)	.63

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from logistic mixed models with a random intercept for clustering within census tracts. Model 1B: bivariate (crude). Model 2B: adjusted for maternal race/ethnicity. Model 3B: further adjusted for birth year, sex, maternal age, birthweight, gestational age, and birth order. Model 4B: further adjusted for maternal education. Neighborhood SES modeled as a continuous variable (higher values = higher SES); values range from -7.3 to 1.5 (SD=1). Control sample: N=11,907.

Table 3-6 Associations between neighborhood SES and childhood cancer incidence estimated from bivariate and trivariate models; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

	All Cancers Combined N = 2,947				Acute Lymphoblastic Leukemia N = 673			
Model	OR	95% CI	P	$\%_{Change}$	OR	95% CI	P	$\%_{Change}$
Bivariate	1.09	(1.04, 1.14)	<.001		1.10	(1.01, 1.20)	.026	
Trivariate *adjusted for:*								
Race/Ethnicity	1.05	(1.00, 1.11)	.035	39%	1.08	(0.99, 1.19)	.095	18%
Maternal Age	1.08	(1.03, 1.13)	.002	14%	1.09	(0.99, 1.18)	.066	15%
Birthweight	1.08	(1.04, 1.13)	<.001	5%	1.09	(1.00, 1.19)	.045	10%
Gestational Age	1.09	(1.04, 1.14)	<.001	0%	1.10	(1.01, 1.20)	.026	0%
First Born	1.09	(1.04, 1.14)	<.001	0%	1.10	(1.01, 1.20)	.026	0%
Sex	1.09	(1.04, 1.14)	<.001	-1%	1.10	(1.01, 1.20)	.026	0%
Birth Year	1.09	(1.04, 1.14)	<.001	0%	1.10	(1.01, 1.20)	.026	0%
	Lymphomas N = 311				Central Nervous System Tumors N = 662			
Model	OR	95% CI	P	$\%_{Change}$	OR	95% CI	P	$\%_{Change}$
Bivariate	1.14	(1.01, 1.30)	.037		1.18	(1.08, 1.29)	<.001	
Trivariate *adjusted for:*								
Race/Ethnicity	1.15	(1.00, 1.31)	.052	0%	1.09	(0.99, 1.20)	.091	49%
Maternal Age	1.13	(0.99, 1.28)	.073	12%	1.18	(1.08, 1.30)	.001	0%
Birthweight	1.14	(1.00, 1.29)	.044	3%	1.17	(1.07, 1.28)	.001	4%
Gestational Age	1.15	(1.01, 1.30)	.037	0%	1.18	(1.08, 1.29)	<.001	0%
First Born	1.15	(1.01, 1.30)	.036	0%	1.18	(1.08, 1.29)	<.001	0%
Sex	1.15	(1.01, 1.30)	.035	-1%	1.18	(1.08, 1.29)	<.001	0%
Birth Year	1.15	(1.01, 1.30)	.036	-1%	1.18	(1.08, 1.29)	<.001	0%

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from logistic mixed models with a random intercept for clustering within census tracts. Bivariate models equate to Model 1B in Table 3-5. Trivariate models adjusted for maternal race/ethnicity equate to Model 2B in Table 3-5. % change compares beta coefficients of neighborhood SES predicting cancer incidence estimated from bivariate and trivariate models (i.e. βbivar-βtrivar/βbivar). Control sample: N=11,907.

Table 3-7 Associations between established risk factors and combined childhood cancer incidence, unadjusted and adjusted for SES; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014 (N=14,854)

	Model 5 (unadjusted for SES)			Model 4 (adjusted for SES)		
Predictor	OR	95% CI	P	OR	95% CI	P
Race/Ethnicity (NH white vs. otherwise)	1.23	(1.09, 1.38)	.001	1.17	(1.03, 1.33)	.015
Maternal Age (per 5-year increase)	1.05	(1.01, 1.09)	.012	1.04	(0.99, 1.08)	.111
Birthweight (per 500-gram increase)	1.11	(1.06, 1.16)	<.001	1.10	(1.06, 1.16)	<.001
Gestational Age (per 1-week increase)	0.94	(0.92, 0.97)	<.001	0.94	(0.92, 0.97)	<.001
First Born (first born vs. otherwise)	1.11	(1.02, 1.22)	.017	1.10	(1.01, 1.21)	.036

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from logistic mixed models with a random intercept for clustering within census tracts. Model 5: adjusted for maternal race/ethnicity, birth year, sex, maternal age, birthweight, gestational age, and birth order. Model 4: further adjusted for maternal education and neighborhood SES. NH = non-Hispanic; SES = socioeconomic status. Case sample: N=2,947; Control sample: N=11,907.

Table 3-8 Associations between established risk factors and childhood cancer incidence, unadjusted and adjusted for SES, by cancer type; registry-based case-cohort study, linked Minnesota birth and cancer records, 1989-2014

	Acute Lymphoblastic Leukemia (N=673)						Acute Myeloid Leukemia (N=112)						Lymphomas (N=311)					
	Model 5 (unadjusted for SES)			Model 4 (adjusted for SES)			Model 5 (unadjusted for SES)			Model 4 (adjusted for SES)			Model 5 (unadjusted for SES)			Model 4 (adjusted for SES)		
Predictor	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Race/Ethnicity	1.11	(0.89, 1.39)	.34	1.03	(0.81, 1.32)	.79	0.73	(0.46, 1.17)	.20	0.70	(0.42, 1.18)	.19	1.06	(0.77, 1.45)	.73	0.89	(0.64, 1.26)	.52
Maternal Age	1.05	(0.98, 1.13)	.17	1.02	(0.94, 1.11)	.56	1.16	(0.98, 1.38)	.093	1.16	(0.96, 1.41)	.12	1.08	(0.97, 1.20)	.17	1.02	(0.90, 1.15)	.79
Birthweight	1.20	(1.10, 1.30)	<.001	1.19	(1.09, 1.30)	<.001	1.23	(1.00, 1.52)	.048	1.23	(1.00, 1.52)	.048	0.97	(0.85, 1.10)	.61	0.96	(0.84, 1.09)	.50
Gestational Age	0.92	(0.88, 0.97)	<.001	0.92	(0.88, 0.97)	.002	0.99	(0.87, 1.13)	.88	0.99	(0.87, 1.13)	.90	1.04	(0.97, 1.13)	.27	1.05	(0.97, 1.13)	.23
First Born	1.08	(0.91, 1.28)	.37	1.06	(0.89, 1.26)	.53	1.00	(0.66, 1.51)	.99	1.00	(0.66, 1.53)	.99	0.92	(0.72, 1.19)	.54	0.88	(0.68, 1.14)	.33
	Central Nervous System Tumors (N=662)						Neuroblastoma (N=267)						Retinoblastoma (N=80)					
	Model 5			Model 4			Model 5			Model 4			Model 5			Model 4		
Predictor	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Race/Ethnicity	1.83	(1.41, 2.36)	<.001	1.66	(1.26, 2.19)	<.001	1.38	(0.97, 1.96)	.077	1.40	(0.95, 2.06)	.090	0.82	(0.48, 1.42)	.49	0.66	(0.37, 1.19)	.17
Maternal Age	1.03	(0.95, 1.10)	.49	1.00	(0.92, 1.08)	.94	1.04	(0.93, 1.17)	.49	1.02	(0.89, 1.16)	.77	0.93	(0.75, 1.14)	.48	0.82	(0.64, 1.06)	.13
Birthweight	1.07	(0.98, 1.17)	.12	1.07	(0.98, 1.17)	.15	1.06	(0.92, 1.22)	.40	1.06	(0.92, 1.21)	.43	1.06	(0.83, 1.37)	.63	1.05	(0.81, 1.35)	.72
Gestational Age	1.00	(0.95, 1.05)	.94	1.00	(0.95, 1.05)	.99	0.94	(0.87, 1.01)	.092	0.94	(0.87, 1.01)	.094	1.01	(0.87, 1.17)	.92	1.01	(0.87, 1.18)	.88
First Born	1.18	(0.99, 1.39)	.062	1.15	(0.97, 1.37)	.11	1.19	(0.91, 1.55)	.20	1.17	(0.89, 1.53)	.26	1.06	(0.66, 1.71)	.81	0.96	(0.59, 1.57)	.87
	Wilms' Tumor (N=210)						Hepatoblastoma (N=50)						Rhabdomyosarcoma (N=79)					
	Model 5			Model 4			Model 5			Model 4			Model 5			Model 4		
Predictor	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Race/Ethnicity	2.04	(1.25, 3.31)	.004	2.05	(1.22, 3.46)	.007	0.76	(0.39, 1.48)	.42	1.05	(0.49, 2.24)	.91	1.41	(0.68, 2.94)	.36	1.54	(0.70, 3.38)	.28
Maternal Age	1.03	(0.90, 1.18)	.63	1.03	(0.89, 1.20)	.71	1.04	(0.81, 1.35)	.73	1.21	(0.92, 1.59)	.17	1.13	(0.91, 1.39)	.26	1.23	(0.98, 1.54)	.078
Birthweight	1.31	(1.12, 1.53)	.001	1.31	(1.12, 1.53)	.001	0.82	(0.59, 1.13)	.23	0.84	(0.61, 1.16)	.30	1.26	(0.98, 1.62)	.075	1.28	(0.99, 1.65)	.058
Gestational Age	0.88	(0.81, 0.96)	.005	0.88	(0.81, 0.96)	.005	0.85	(0.75, 0.97)	.017	0.85	(0.75, 0.97)	.014	0.87	(0.76, 0.99)	.039	0.86	(0.76, 0.99)	.036
First Born	1.18	(0.87, 1.61)	.28	1.18	(0.86, 1.61)	.30	1.12	(0.61, 2.06)	.70	1.26	(0.68, 2.35)	.46	1.40	(0.86, 2.29)	.17	1.52	(0.93, 2.50)	.098

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from logistic mixed models with a random intercept for clustering within census tracts. Model 5: adjusted for maternal race/ethnicity, birth year, sex, maternal age, birthweight, gestational age, and birth order. Model 4: further adjusted for maternal education and neighborhood SES. Race/ethnicity ORs compare non-Hispanic white mother versus otherwise; maternal age ORs compare 5-year increase; birthweight ORs compare 500-gram increase; gestational age ORs compare 1-week increase; and first born ORs compare first live birth versus otherwise. Control sample: N=11,907.

Chapter 4

Manuscript 3: Does socioeconomic status explain black-white racial disparities in childhood cancer survival?

Introduction

Despite improvements over the last four decades in cancer survival among the US pediatric population, marked racial disparities persist.[4,67] Compared to non-Hispanic white (white) children, non-Hispanic black (black) children experience lower survival from combined (all diagnoses)[67] and many individual cancers including leukemias,[79-81] lymphomas,[82-84] CNS tumors,[85] and extracranial solid tumors.[86-88,108] The underlying mechanisms accounting for these differences are not well understood, and may vary by cancer. As outlined in Figure 4-1, both biological and socioeconomic pathways have been proposed in the literature.[89,90,106] It is suggested that underlying genetic variations associated with ancestry may lead to differences in tumor biology and pharmacogenetics for some childhood cancers.[90] For example, compared to whites, black patients with neuroblastoma are more likely to have characteristics of high-risk disease, including older age at presentation and unfavorable tumor histology.[86] A higher proportion of unfavorable tumor characteristics among black compared to white patients has also been documented for acute lymphoblastic leukemia (ALL),[92-94] Hodgkin lymphoma (HL),[95] rhabdomyosarcoma,[96] and central nervous system (CNS) tumors.[85] However, race is a socially constructed taxonomy that is not synonymous with ancestry.[91] Race is highly correlated with socioeconomic status (SES),[103,104] especially in the United States where embedded institutionalized racism continues to place racial and ethnic minorities at high risk of low SES.[105] Evidence is emerging of a positive association between higher SES

and childhood cancer survival.[85,87,106-108] Therefore, SES may also contribute to racial survival differences through mechanisms such as treatment adherence and access to quality health care, including enrollment into clinical trials.[90]

Disentangling the relative role of social versus biological pathways in explaining survival differences by race is important for practice and intervention implications. If underlying socioeconomic factors explain survival differences by race (e.g. worse healthcare coverage among blacks compared to whites results in inferior treatment), then interventions addressing social and economic barriers to care are warranted to address racial disparities. On the other hand, if SES does not account for survival differences by race, then biological mechanisms, such as tumor biology and pharmacogenetics, must be considered. In this case, investment in developing novel and personalized drug therapies may be appropriate. Moreover, if both social and biological factors contribute to survival differences, then a multipronged intervention approach may ultimately be needed.

Formal mediation methods are required to empirically test the underlying pathways (i.e. mediators) contributing to survival differences by race. Through mediation analysis, the proportion of the race-survival association explained by a given pathway (e.g. SES) can be estimated to ultimately provide insight into which factors should be targeted in interventions addressing survival disparities. However, to date, mediation methods have not been used in the childhood cancer literature. In this study, we conducted a formal mediation analysis to determine whether SES mediates the association of race (black versus white) and childhood cancer survival and, if so, to what degree. We also tested prognostic factors, including tumor stage and treatment, as other potential mediators. We used population-based data representative of the US pediatric

cancer population, and we assessed survival from combined and individual childhood cancers to determine if mediation differs across tumor groups.

Methods

Study Population

We obtained population-based data from the Surveillance, Epidemiology, and End Results (SEER) 18 database, excluding the Alaska Native Tumor Registry.[138] We restricted to black and white cases, ages 0 to 19 years, with microscopically confirmed first primary malignancies. The main predictor of interest was race/ethnicity, based on provider's report in the medical record.[139] We restricted to black and white cases because of the magnitude of the disparity and small case counts among other racial and ethnic groups.[140] SES data were available in SEER for years 2000 to 2012. Therefore, we further restricted to cases diagnosed January 1, 2000 to December 31, 2011, followed through December 31, 2012. This allowed for at least one year of follow-up. We excluded 37 cases staged with *in situ* tumors, 428 cases with missing or zero months of follow-up time, and 568 cases with missing SES data. Our final analytic sample consisted of 27,741 cases (4,529 blacks; 23,212 whites). We assessed overall survival from combined and individual cancers (where $N \geq 200$ cases per group) based on the International Classification of Childhood Cancer, third edition.[113] Overall survival was pre-calculated in SEER as months from date of diagnosis to date of death from any cause, or censored at date of last contact.

Potential Mediators

Socioeconomic Status – We measured SES, at date of cancer diagnosis, using the validated census tract-level composite index available in the restricted SEER database.[141]

As described previously,[142] the index was constructed using factor analysis of 2000 decennial census data and 2005-2009 American Community Survey (ACS) data.[141] Seven measures of SES, specified by Yost *et al*. (2001), were included in the index: proportion employed in working-class occupations, proportion aged 16+ unemployed, education index,[143] median household income, proportion below 200% poverty level, median rent, and median house value.[144] Addresses at diagnosis were geocoded to census tracts, 2000 geographic boundaries. 2000 census values were assigned to cases diagnosed 2000-2003; 2005-2009 ACS values were assigned to cases diagnosed 2004-2011.[141] The index is available in SEER as a five-level variable categorized into quintiles (Q1, low SES; Q5, high SES). In a secondary analysis, we tested individual-level health insurance status as an alternative measure of SES defined as (1) private, (2) any Medicaid, (3) insured, no specifics, (4) uninsured, and (5) unknown.

Treatment and Stage – First-course cancer-directed surgery (surgery) was defined as (1) surgery performed versus (0) otherwise. Radiation therapy (radiation) was defined as (1) beam radiation, combination of beam radiation with implant or isotopes, radiation not otherwise specified, radioactive implants, or radioisotopes versus (0) otherwise. Distal stage at diagnosis (distal stage) was defined as (1) distant site(s)/nodes involved versus (0) otherwise, based on SEER summary staging classifications.[145] Surgery and stage do not apply to leukemias.

Statistical Analysis

We compared characteristics of black and white cases using Pearson's chi-square test for categorical measures and the two sample t-test for continuous measures. For each cancer type, we estimated black-white mortality hazard ratios (total effects) from

multivariable Cox proportional hazards regression adjusted for age group at diagnosis (<1, 1-4, 5-9, 10-14, 15-19 years), year of diagnosis (modeled linearly), and sex. For cancers with a statistically significant black-white mortality hazard ratio, we used the inverse odds weighting (IOW) method to test for mediation.[146,147] IOW is a semiparametric weight-based approach that overcomes many limitations of traditional parametric mediation methods.[148] For example, IOW is appropriate for any functional form (rather than just linear models), can test multiple mediators simultaneously (as opposed to testing them one by one), and is valid even in the presence of exposure-mediator interactions.[149] We tested for mediation by (1) SES, (2) treatment and stage, and (3) SES, treatment, and stage. For cancers in which the census tract SES index significantly mediated the race-survival association, we further tested for mediation by health insurance status. This secondary analysis was restricted to cases diagnosed 2007-2011 (N=11,225), when health insurance data became available in SEER.

Implementation of the IOW method has been described in detail previously.[149] Briefly, we estimated the total effect of race on survival by specifying an unweighted multivariable Cox proportional hazards model adjusted for age group, sex, and year of diagnosis. The total effect provides an estimate of the overall association between race and survival, without specifying any mediating pathways (so it captures all possible mediators).[150] We then estimated the (natural)[150] direct effect of race on survival by specifying a multivariable Cox proportional hazards model adjusted for age group, sex, and year of diagnosis, weighted by the IOW weight. Weight specification renders the exposure and mediator(s) independent, thus "blocking" the indirect pathway through the tested mediator(s).[151] Therefore, the direct effect provides an estimate of the race-survival

association that remains *after* accounting for the pathway through the tested mediator(s).[150] To create the IOW weights operationally, we first recovered the predicted odds of exposure (i.e. black race) for each subject from a multivariable logistic regression model specifying mediator(s), age group, sex, and year of diagnosis. We then took the inverse of the predicted odds to create the IOW weight for whites; blacks were assigned a weight of one. Black race was selected as the reference population to minimize extreme weighting values. Finally, we estimated the (natural)[150] indirect effect of race on survival operating through the tested mediator(s) by subtracting the direct (log hazard ratio (β)) from the total effect, and bootstrapping to obtain standard errors (500 replications). A significant indirect effect provides statistical evidence that the tested variable(s) does in fact mediate the race-survival association. To quantify the mediation magnitude, we calculated the percent change between the total and direct effect $((\beta_{total} - \beta_{direct})/\beta_{total})$, which estimates the proportion of the race-survival association explained by the tested mediator(s). Statistical significance was determined as $p < .05$ for a 2-sided hypothesis test. Within each set of IOW results, we adjusted p-values for multiple comparisons using the Benjamini-Hochberg method,[152] implemented in R 3.1.3.[153] All other analyses were performed using Stata 14.2 (College Station, TX).[114]

Results

Characteristics of combined childhood cancer cases are provided in Table 4-1. Average follow-up time was 70.3 months (standard deviation, 44.1; range, 1-155); 15.9% of cases died by end of follow-up. Black cases were younger, more likely to be female, and more likely to have distal stage disease at diagnosis compared to whites. Blacks were more likely to undergo radiation, while whites were more likely to undergo surgery. SES

was significantly lower among blacks compared to whites; 40.1% of blacks were in the lowest SES quintile (Q1) compared to 11.2% of whites. Comparisons by individual cancer type are provided in Table 4-2.

Table 4-3 presents hazard ratios comparing all-cause mortality in black and white childhood cancer cases, adjusted for age group, sex, and year of diagnosis (total effects). Blacks had 71% higher hazard of death compared to whites for combined childhood cancers (Mortality HR, 1.71; 95% CI, 1.59, 1.83). Among individual cancers, blacks had significantly higher mortality compared to whites for 8 of the 12 cancers, including ALL, acute myeloid leukemia (AML), HL, non-Hodgkin lymphoma (NHL), astrocytomas, non-astrocytoma CNS tumors (other CNS), neuroblastoma, and non-rhabdomyosarcoma soft tissue sarcomas (NRSTS). Mortality hazard ratios for race were not statistically significant for rhabdomyosarcoma, Wilms' tumor (WT), osteosarcoma, or germ cell tumors (GCT). Survival curves by race are provided in Figure 4-2; similar patterns emerged on the absolute scale.

IOW results for mediation by SES are presented in Table 4-4. We determined SES to be a significant mediator of the race-survival association if the percent change between the total and direct effect was ≥ 10%, and if the indirect effect of race on survival operating through SES was statistically significant. Based on these criteria, SES was identified as a significant mediator of the race-survival association for combined childhood cancers (indirect effect HR (iHR), 1.12; 95% CI, 1.08, 1.15; $p < .001$; 20% change), ALL (iHR, 1.18; 95% CI, 1.08, 1.29; $p < .001$; 44%), and AML (iHR, 1.16; 95% CI, 1.04, 1.29; $p = .017$; 28%). The percent change was >10% for HL (11%), NHL

(32%), astrocytomas (12%), other CNS (18%), and neuroblastoma (22%), though indirect effects were non-significant.

As shown in Table 4-5, when we substituted individual-level health insurance status for tract-level SES, comparable indirect effects were estimated for combined childhood cancers (iHR, 1.10; 95% CI, 1.05, 1.15; p <.001; 19% change), ALL (iHR, 1.19; 95% CI, 0.96, 1.46; p = .14; 37%), and AML (iHR, 1.06; 95% CI, 0.92, 1.22; p = .45; 13%), though power was reduced.

IOW results for treatment and stage are shown in Table 4-6. Treatment and stage significantly mediated the race-survival association for astrocytomas (iHR, 1.19; 95% CI, 1.01, 1.40, p = .044; 27%) and neuroblastoma (iHR, 1.16; 95% CI, 1.04, 1.30; p = .015; 32%). The percent change was >10% for NRSTS (14%), but the indirect effect was non-significant (p = .59). No additional statistically significant indirect effects emerged when we specified SES, treatment, and stage as simultaneous mediators (Table 4-7).

Discussion

This is the first study to use formal mediation methods to unpack childhood cancer survival disparities by race. In our analysis, we replicated findings that whites have a significant survival advantage over blacks for many childhood cancers including leukemias,[79-81] lymphomas,[82-84] CNS tumors,[85] neuroblastoma,[86] and NRSTS.[88] As with prior studies, we found no significant survival differences for rhabdomyosarcoma, WT, osteosarcoma, or GCT.[67,87] Thus, race does not appear to be uniformly associated with survival across all types of childhood cancer.

For cancers with significant survival differences by race, we used novel methods to test the hypothesis of whether SES accounts for the disparity. In IOW analysis, SES

(operationalized as an area-based composite index) was found to significantly mediate the race-survival association for combined childhood cancers and leukemias, accounting for between 20 and 44% of the disparity. We found comparable results when we used a secondary measure of SES operationalized at the individual-level, health insurance status. Though power was reduced (health insurance data only available 2007-2011), these secondary findings reinforce the role of SES in survival from childhood leukemias, particularly ALL.

Prior literature suggests that the strong association between SES and ALL survival may be explained by differences in treatment adherence.[90] ALL treatment requires a prolonged maintenance phase composed of oral administration of antimetabolites to prevent relapse.[90] Low SES families may confront multiple barriers to maintaining therapy, including low health literacy, poor patient-provider communication, and economic constraints, such as inadequate health insurance coverage.[90] This is supported by a 2012 study that found lower treatment adherence among children with ALL living in a single-mother household versus a two-parent household.[154] Treatment differences may also explain the association between SES and AML survival. Studies among adult AML patients have reported positive associations between higher SES and receipt of chemotherapy and hematopoietic cell transplantation (HCT).[155,156] Further, a study by Knight *et al.* found that stress-related gene expression profiles associated with SES may directly influence HCT outcomes.[157] Additional research exploring socioeconomic differences in treatment protocol and adherence are needed in the pediatric population to ultimately inform interventions addressing survival disparities.

SES was not identified as a significant mediator of the race-survival association for childhood solid tumors. However, treatment and stage were found to mediate the race-survival association for astrocytomas and neuroblastoma. We tested whether SES (modeled ordinally) was associated with treatment and stage, but found no significant associations among astrocytoma cases, and found only an association between surgery and SES among neuroblastoma cases (Table 4-8). Yet, surgery was not associated with race among neuroblastoma cases (Table 4-2). Taken together, these findings provide some indirect evidence that non-social factors, such as tumor biology, may contribute to survival differences by race for these cancers. Prior studies have reported a greater proportion of high-risk disease among black compared to white children diagnosed with CNS tumors[85] and neuroblastoma.[86]

We did not identify any significant mediators for the other types of solid tumors with survival differences by race (HL, NHL, other CNS, and NRSTS). For HL, this may be due to high overall survival in both racial groups (5-year survival black vs. white: 94% vs. 96%), and thus a relatively small disparity on the absolute scale. This may also explain why we found no survival differences for WT or GCT, which both have very high survival rates (Figure 4-2). For NHL, other CNS, and NRSTS, significant mediating effects may have been obscured by tumor heterogeneity (e.g. differences in cellular morphology, gene expression, or metastatic potential) within cancer groups. Further stratification into relevant subgroups may be required to detect mediators for these cancers. For example, because we observed a marginally significant indirect effect of SES for NHL cases, further tumor stratification may reveal even stronger SES associations for specific subtypes of NHL.

Limitations

There are several limitations to this study. First, SEER's tract-level SES index was only available for years 2000-2012, which restricted sample size and follow-up time. This prevented us from testing more homogenized cancer subgroups or stratifying by age. Second, because individual-level health insurance data were only available for a subset of our sample (diagnosed 2007-2011), we relied on an area-based index as our primary measure of SES. Though use of this measure improves upon many prior population-based cancer studies, tract-level SES is still a proxy for individual-level SES.[158,159] Measures of individual-level SES may reveal stronger associations between SES and survival,[107] especially since area-level and individual-level SES may have independent effects on health. Given the limitations of our SES measure, we caution readers against interpreting weak or null SES findings as evidence of a biological effect, especially since nondifferential measurement error can bias effects towards the null.[160] Third, we lacked geographic variables to explore potential spatial variations in survival. Fourth, the dearth of clinical data in SEER limited our ability to account for therapeutic and biological factors, such as cytogenetic or molecular features. We also lacked detailed diagnostic and staging information specific to childhood cancers,[161] which may help inform whether entry into care contributes to survival differences. We acknowledge that richer data may be available in clinic-based studies, but note that these studies can be hampered by issues of internal and external validity.[90,116] Fifth, there is the potential for differential loss to follow-up by race or SES, though a prior study reported no significant differences in follow-up between black and white SEER cases (all ages and cancers, 2000-2008).[162]

Finally, because we only considered disparities between black and white children, additional research is needed for other racial and ethnic groups.

Conclusion

Through the application of formal mediation methods, we estimated that SES accounts for 20% of the racial disparity in survival for combined childhood cancers, and nearly half of the disparity for ALL specifically. These findings suggest that childhood cancer survival differences by race could theoretically be addressed through initiatives that reduce social and economic barriers to effective care. Such efforts may include expanded healthcare coverage, improved patient care coordination, increased health literacy, provider education, and supplementation of transportation and childcare costs during treatment. However, because SES did not fully account for survival disparities in our analysis, we cannot rule out other possible explanations (including tumor biology) for survival differences by race, especially for astrocytomas and neuroblastoma. Therefore, a multipronged intervention approach that addresses socioeconomic barriers to care and invests in novel and personalized treatment regimens may ultimately be needed to fully eliminate childhood cancer survival differences.

Figure 4-1 Conceptual model of mediating pathways between race and childhood cancer survival

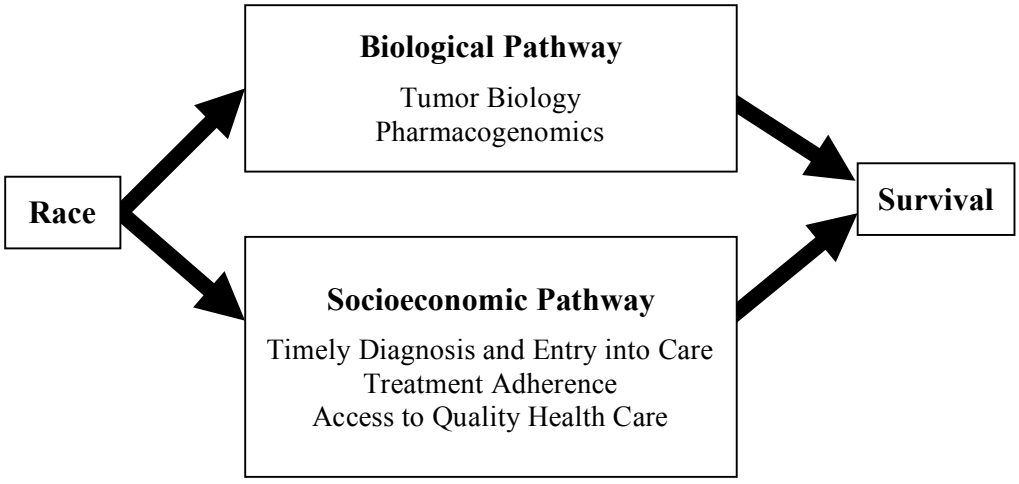


Table 4-1 Characteristics of combined childhood cancer cases, ages 0-19 years, overall and by race, SEER 18 registries, 2000-2011 diagnoses (N=27,741)

Variable	Entire Sample (N=27,741)	Black Cases (N=4,529)	White Cases (N=23,212)	P
Follow-Up Months, mean (SD)	70.3 (44.1)	63.9 (43.4)	71.61 (44.1)	<.001
All-Cause Mortality, No. (%)	4,414 (15.9)	1,043 (23.0)	3,371 (14.5)	<.001
Age at Diagnosis, years, mean (SD)	10.0 (6.4)	9.7 (6.3)	10.1 (6.5)	<.001
Age Group at Diagnosis, years, No. (%)				<.001
<1	1,682 (6.1)	298 (6.6)	1,384 (6.0)	
1-4	6,502 (23.4)	1,048 (23.1)	5,454 (23.5)	
5-9	4,443 (16.0)	794 (17.5)	3,649 (15.7)	
10-14	5,534 (20.0)	1,029 (22.7)	4,505 (19.4)	
15-19	9,580 (34.5)	1,360 (30.0)	8,220 (35.4)	
Female, No. (%)	12,767 (46.0)	2,157 (47.6)	10,610 (45.7)	.018
Radiation Therapy, No. (%)	7,252 (26.1)	1,252 (27.6)	6,000 (25.9)	.012
First-Course Cancer-Directed Surgery, No. (%)	16,275 (58.7)	2,581 (57.0)	13,694 (59.0)	.012
Distal Stage at Diagnosis, No. (%)	11,053 (39.8)	1,898 (41.9)	9,155 (39.4)	.002
Neighborhood SES, No. (%)				<.001
Q1 (lowest SES)	4,423 (15.9)	1,816 (40.1)	2,607 (11.2)	
Q2	4,821 (17.4)	1,025 (22.6)	3,796 (16.4)	
Q3	5,344 (19.3)	799 (17.6)	4,545 (19.6)	
Q4	5,997 (21.6)	568 (12.5)	5,429 (23.4)	
Q5 (highest SES)	7,156 (25.8)	321 (7.1)	6,835 (29.5)	

Notes: P-values compare characteristics of black versus white cases estimated from Pearson's chi-square test for categorical measures and the two-sample t-test for continuous measures. SD = standard deviation; SES = socioeconomic status.

Table 4-2 Characteristics of childhood cancer cases by cancer type, ages 0-19 years, overall and by race, SEER 18 registries, 2000-2011 diagnoses

		Survival Months	All-Cause Mortality	Age at Diagnosis	Female	Radiation	Surgery	Distal	Tract-Level SES Index %				
Cancer	N	Mean (SD)	%	Mean (SD)	%	%	%	%	Q1	Q2	Q3	Q4	Q5
Acute Lymphoblastic Leukemia													
Black	634	68.2 (41.9)	15.9	7.5 (5.4)	41.5	14.4	n/a	n/a	38.0	23.3	17.4	14.0	7.3
White	4,357	74.0 (43.1)	10.1	6.6 (5.2)	42.9	10.8	n/a	n/a	11.5	16.8	19.2	23.4	29.1
p-value		.001	<.001	<.001	.50	.008			<.001				
Acute Myeloid Leukemia													
Black	253	48.5 (43.0)	45.9	8.7 (6.4)	46.6	10.7	n/a	n/a	40.7	20.6	16.6	13.4	8.7
White	965	58.1 (45.1)	30.4	9.3 (6.8)	47.4	11.2	n/a	n/a	14.0	16.5	18.8	24.9	25.9
p-value		.002	<.001	.16	.84	.82			<.001				
Hodgkin Lymphoma													
Black	384	73.4 (40.1)	7.6	14.4 (3.9)	48.4	50.8	47.4	41.4	41.7	26.3	15.1	10.2	6.8
White	1,947	79.5 (41.7)	4.9	15.4 (3.4)	48.1	49.8	48.5	31.1	10.9	16.1	19.2	24.2	29.6
p-value		.009	.033	<.001	.91	.73	.68	<.001	<.001				
Non-Hodgkin Lymphoma													
Black	343	63.8 (42.0)	19.2	12.8 (4.7)	41.1	18.1	30.6	44.6	40.5	24.5	17.5	11.1	6.4
White	1,169	72.5 (45.0)	13.2	12.8 (5.0)	36.4	15.7	34.6	45.0	11.6	15.1	18.8	25.2	29.3
p-value		.002	.005	.95	.12	.29	.17	.90	<.001				
Astrocytomas													
Black	360	62.6 (44.9)	26.4	8.9 (5.5)	49.2	23.9	82.5	1.7	36.7	19.4	20.8	13.9	9.2
White	2,080	72.0 (45.5)	15.1	9.3 (5.6)	47.0	19.8	86.5	1.8	11.5	17.3	19.2	23.9	28.2
p-value		<.001	<.001	.17	.44	.076	.045	.83	<.001				
Other CNS Tumors													
Black	326	56.4 (45.3)	40.8	7.4 (5.6)	50.6	54.9	87.1	13.8	44.8	20.6	16.0	14.1	4.6
White	1,718	63.3 (45.4)	29.3	7.9 (5.9)	40.9	57.2	89.8	13.4	11.4	16.0	20.3	23.5	29.0
p-value		.012	<.001	.11	.001	.45	.16	.84	<.001				
Neuroblastoma													
Black	264	56.6 (39.4)	30.7	3.2 (3.8)	50.0	29.9	79.6	57.6	39.4	22.0	20.5	12.5	5.7
White	1,214	67.1 (43.9)	19.1	2.4 (3.2)	47.5	20.9	79.4	46.8	12.4	17.6	21.1	21.8	27.1
p-value		<.001	<.001	<.001	.47	.002	.96	.001	<.001				
Non-Rhabdomyosarcoma STS													
Black	296	63.9 (45.2)	26.7	12.1 (5.7)	49.3	32.8	84.8	14.2	39.9	21.3	18.6	13.9	6.4
White	974	69.2 (44.8)	20.1	12.0 (5.9)	45.2	32.6	86.7	13.5	8.1	18.3	20.2	23.3	30.1
p-value		.077	.016	.88	.21	.94	.42	.75	<.001				
Rhabdomyosarcoma													
Black	208	54.8 (41.3)	36.5	8.5 (5.7)	40.9	61.5	61.5	27.9	35.6	26.9	17.3	13.0	7.2
White	656	62.0 (44.2)	31.6	8.0 (5.8)	41.5	66.9	56.9	29.6	11.3	17.2	17.8	23.8	29.9
p-value		.040	.18	.25	.88	.15	.23	.64	<.001				
Wilms' Tumor													
Black	254	72.0 (44.7)	7.5	3.6 (2.5)	55.1	42.5	96.9	28.7	42.9	22.1	15.0	11.8	8.3
White	789	72.7 (43.1)	7.4	3.5 (3.3)	51.7	44.1	96.7	22.6	12.8	18.4	19.4	22.8	26.6
p-value		.82	.95	.67	.34	.66	.91	.045	<.001				
Osteosarcoma													
Black	217	63.8 (43.9)	30.4	13.5 (4.0)	43.8	1.8	86.6	20.7	39.6	25.8	16.6	10.6	7.4
White	619	62.3 (42.0)	33.0	13.3 (3.7)	45.6	3.6	90.3	17.6	8.2	18.4	21.7	24.7	27.0
p-value		.65	.49	.58	.65	.21	.13	.31	<.001				
Germ Cell Tumors													
Black	231	76.7 (43.4)	7.4	12.4 (6.5)	67.5	16.5	89.2	20.8	42.4	21.7	16.0	11.7	8.2
White	1,549	77.6 (43.1)	5.7	14.0 (6.0)	31.0	13.7	90.1	14.5	10.8	15.7	20.2	24.0	29.4
p-value		.77	.31	<.001	<.001	.26	.66	.013	<.001				

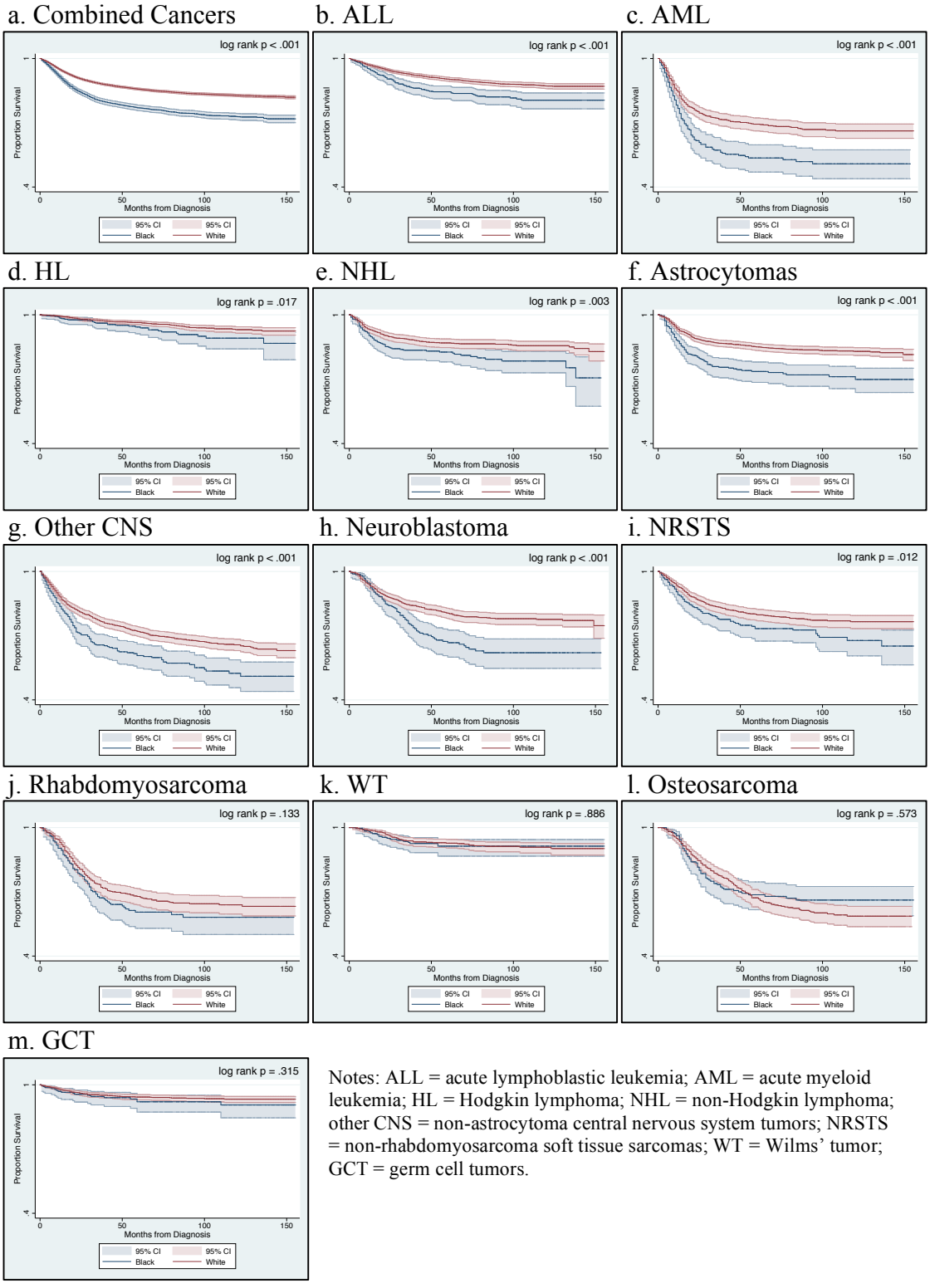
Notes: P-values compare characteristics among black versus white cases estimated from Pearson's chi-square test for categorical measures and the two-sample t-test for continuous measures. Surgery and stage are not applicable (n/a) to leukemias. SD = standard deviation; SES = socioeconomic status; CNS = central nervous system; STS = soft tissue sarcomas.

Table 4-3 Comparison of all-cause mortality in black versus white cases of combined and individual childhood cancers, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=27,741)

Cancer	Sample Size			Mortality Hazard Ratio (black vs. white)	95% CI	P
	Entire Sample	Black Cases	White Cases			
Combined Cancers	27,741	4,529	23,212	1.71	(1.59, 1.83)	<.001
Acute Lymphoblastic Leukemia (ALL)	4,991	634	4,357	1.47	(1.18, 1.82)	.001
Acute Myeloid Leukemia (AML)	1,218	253	965	1.69	(1.37, 2.10)	<.001
Hodgkin Lymphoma (HL)	2,331	384	1,947	1.74	(1.15, 2.65)	.009
Non-Hodgkin Lymphoma (NHL)	1,512	343	1,169	1.60	(1.20, 2.14)	.001
Astrocytomas	2,440	360	2,080	1.91	(1.52, 2.41)	<.001
Non-Astrocytoma CNS Tumors (other CNS)	2,044	326	1,718	1.53	(1.26, 1.86)	<.001
Neuroblastoma	1,478	264	1,214	1.59	(1.23, 2.06)	<.001
Non-Rhabdomyosarcoma STS (NRSTS)	1,270	296	974	1.44	(1.11, 1.87)	.006
Rhabdomyosarcoma	864	208	656	1.28	(0.98, 1.67)	.070
Wilms' tumor (WT)	1,043	254	789	1.03	(0.61, 1.74)	.90
Osteosarcoma	836	217	619	0.88	(0.67, 1.17)	.39
Germ Cell Tumors (GCT)	1,780	231	1,549	1.21	(0.71, 2.08)	.49

Notes: Mortality hazard ratios compare all-cause mortality in black versus white cases estimated from multivariable Cox proportional hazards regression models adjusted for age category, sex, and year of diagnosis. CNS = central nervous system; STS = soft tissue sarcomas.

Figure 4-2 Kaplan-Meier survival curves by race, combined and individual childhood cancers, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=27,741)



Notes: ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; HL = Hodgkin lymphoma; NHL = non-Hodgkin lymphoma; other CNS = non-astrocytoma central nervous system tumors; NRSTS = non-rhabdomyosarcoma soft tissue sarcomas; WT = Wilms' tumor; GCT = germ cell tumors.

Table 4-4 Mediating effect of SES on the race-survival association for combined and individual childhood cancers, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=27,741)

	Total Effect of black race on survival through all mediating pathways			**Direct Effect** of black race on survival after blocking SES pathway			**Indirect Effect** of black race on survival operating through SES pathway			**Percent Change** from total to direct effect
Cancer	Mortality HR	95% CI	P	Mortality HR	95% CI	P	Mortality HR	95% CI	P	%
Combined	1.71	(1.59, 1.83)	<.001	1.53	(1.42, 1.65)	<.001	1.12	(1.08, 1.15)	<.001	20%
ALL	1.47	(1.18, 1.82)	.002	1.24	(0.98, 1.57)	.095	1.18	(1.08, 1.29)	<.001	44%
AML	1.69	(1.38, 2.09)	<.001	1.46	(1.16, 1.85)	.002	1.16	(1.04, 1.29)	.017	28%
HL	1.74	(1.11, 2.74)	.024	1.64	(0.97, 2.78)	.087	1.06	(0.81, 1.40)	.68	11%
NHL	1.60	(1.18, 2.17)	.005	1.37	(0.98, 1.94)	.087	1.16	(0.99, 1.37)	.087	32%
Astrocytomas	1.91	(1.54, 2.38)	<.001	1.77	(1.40, 2.25)	<.001	1.08	(0.98, 1.19)	.14	12%
Other CNS	1.53	(1.25, 1.87)	<.001	1.42	(1.13, 1.79)	.006	1.08	(0.96, 1.20)	.20	18%
Neuroblastoma	1.59	(1.24, 2.04)	<.001	1.44	(1.09, 1.89)	.017	1.11	(0.97, 1.27)	.15	22%
NRSTS	1.44	(1.10, 1.88)	.014	1.43	(1.02, 2.00)	.059	1.01	(0.84, 1.21)	.91	3%

Notes: Mortality hazard ratios (HR) compare all-cause mortality in black versus white cases. Inverse odds weighting method used to test for mediation. Models adjusted for age category, sex, and year of diagnosis. P-values adjusted for multiple comparisons using the Benjamini-Hochberg method. Combined = all cancer diagnoses; ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; HL = Hodgkin lymphoma; NHL = non-Hodgkin lymphoma; other CNS = non-astrocytoma central nervous system tumors; NRSTS = non-rhabdomyosarcoma soft tissue sarcomas.

Table 4-5 Mediating effect of health insurance status on the race-survival association for combined and childhood leukemias, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=11,225)

	Total Effect of black race on survival through all mediating pathways			**Direct Effect** of black race on survival after blocking health insurance pathway			**Indirect Effect** of black race on survival operating through health insurance pathway			**Percent Change** from total to direct effect
Cancer	Mortality HR	95% CI	P	Mortality HR	95% CI	P	Mortality HR	95% CI	P	%
Combined	1.67	(1.47, 1.88)	<.001	1.52	(1.33, 1.73)	<.001	1.10	(1.05, 1.15)	<.001	19%
ALL	1.59	(1.02, 2.49)	.074	1.34	(0.83, 2.18)	.26	1.19	(0.96, 1.46)	.14	37%
AML	1.54	(1.03, 2.29)	.074	1.45	(0.95, 2.23)	.13	1.06	(0.92, 1.22)	.45	13%

Notes: Mortality hazard ratios (HR) compare all-cause mortality in black versus white cases. Inverse odds weighting method used to test for mediation. Models adjusted for age category, sex, and year of diagnosis. P-values adjusted for multiple comparisons using the Benjamini-Hochberg method. Combined = all cancer diagnoses; ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia.

Table 4-6 Mediating effect of treatment and stage on the race-survival association for combined and individual childhood cancers, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=27,741)

	Total Effect of black race on survival through all mediating pathways			**Direct Effect** of black race on survival after blocking treatment and stage pathway			**Indirect Effect** of black race on survival operating through treatment and stage pathway			**Percent Change** from total to direct effect
Cancer	Mortality HR	95% CI	P	Mortality HR	95% CI	P	Mortality HR	95% CI	P	%
Combined	1.71	(1.59, 1.83)	<.001	1.65	(1.54, 1.77)	<.001	1.03	(1.02, 1.05)	<.001	6%
ALL	1.47	(1.18, 1.82)	.002	1.44	(1.15, 1.79)	.002	1.02	(0.99, 1.05)	.20	5%
AML	1.69	(1.38, 2.09)	<.001	1.72	(1.39, 2.13)	<.001	0.98	(0.95, 1.02)	.45	-3%
HL	1.74	(1.11, 2.74)	.021	1.67	(1.06, 2.63)	.038	1.05	(0.97, 1.13)	.28	8%
NHL	1.60	(1.18, 2.17)	.004	1.61	(1.18, 2.19)	.006	1.00	(0.93, 1.07)	.92	-1%
Astrocytomas	1.91	(1.54, 2.38)	<.001	1.61	(1.34, 1.93)	<.001	1.19	(1.01, 1.40)	.044	27%
Other CNS	1.53	(1.25, 1.87)	<.001	1.51	(1.23, 1.86)	<.001	1.01	(0.96, 1.07)	.69	3%
Neuroblastoma	1.59	(1.24, 2.04)	<.001	1.37	(1.09, 1.74)	.014	1.16	(1.04, 1.30)	.015	32%
NRSTS	1.44	(1.10, 1.88)	.013	1.37	(1.08, 1.74)	.015	1.05	(0.89, 1.24)	.59	14%

Notes: Mortality hazard ratios (HR) compare all-cause mortality in black versus white cases. Inverse odds weighting method used to test for mediation. Models adjusted for age category, sex, and year of diagnosis. Surgery and stage do not apply to leukemias. P-values adjusted for multiple comparisons using the Benjamini-Hochberg method. Combined = all cancer diagnoses; ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; HL = Hodgkin lymphoma; NHL = non-Hodgkin lymphoma; other CNS = non-astrocytoma central nervous system tumors; NRSTS = non-rhabdomyosarcoma soft tissue sarcomas.

Table 4-7 Mediating effect of SES, treatment, and stage on the race-survival association for combined and individual childhood cancers, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=27,741)

	Total Effect of black race on survival through all mediating pathways			**Direct Effect** of black race on survival after blocking SES, treatment, and stage pathways			**Indirect Effect** of black race on survival operating through SES, treatment, and stage pathways			**Percent Change** from total to direct effect
Cancer	Mortality HR	95% CI	P	Mortality HR	95% CI	P	Mortality HR	95% CI	P	%
Combined	1.71	(1.59, 1.83)	<.001	1.50	(1.39, 1.62)	<.001	1.14	(1.10, 1.18)	<.001	24%
ALL	1.47	(1.18, 1.82)	.002	1.23	(0.97, 1.57)	.11	1.19	(1.08, 1.30)	<.001	45%
AML	1.69	(1.38, 2.09)	<.001	1.47	(1.16, 1.86)	.002	1.15	(1.03, 1.29)	.022	27%
HL	1.74	(1.11, 2.74)	.023	1.55	(0.90, 2.64)	.13	1.13	(0.84, 1.52)	.44	22%
NHL	1.60	(1.18, 2.17)	.005	1.37	(0.97, 1.94)	.10	1.17	(0.98, 1.40)	.11	33%
Astrocytomas	1.91	(1.54, 2.38)	<.001	1.54	(1.26, 1.89)	<.001	1.24	(1.04, 1.48)	.023	33%
Other CNS	1.53	(1.25, 1.87)	<.001	1.40	(1.11, 1.76)	.010	1.10	(0.97, 1.23)	.14	21%
Neuroblastoma	1.59	(1.24, 2.04)	<.001	1.22	(0.95, 1.58)	.14	1.30	(1.09, 1.55)	.006	57%
NRSTS	1.44	(1.10, 1.88)	.013	1.37	(1.01, 1.85)	.057	1.05	(0.85, 1.30)	.64	14%

Notes: Mortality hazard ratios (HR) compare all-cause mortality in black versus white cases. Inverse odds weighting method used to test for mediation. Models adjusted for age category, sex, and year of diagnosis. Surgery and stage do not apply to leukemias. P-values adjusted for multiple comparisons using the Benjamini-Hochberg method. Combined = all cancer diagnoses; ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; HL = Hodgkin lymphoma; NHL = non-Hodgkin lymphoma; other CNS = non-astrocytoma central nervous system tumors; NRSTS = non-rhabdomyosarcoma soft tissue sarcomas.

Table 4-8 Associations of SES predicting treatment and stage for combined and individual childhood cancers, ages 0 to 19 years, SEER 18 registries, 2000-2011 diagnoses (N=27,741)

	Radiation			Surgery			Distal Stage		
Cancer	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Combined	0.98	(0.96, 1.00)	.10	1.03	(1.01, 1.05)	<.001	0.98	(0.96, 1.00)	.022
ALL	0.97	(0.91, 1.04)	.45	n/a	n/a	n/a	n/a	n/a	n/a
AML	0.98	(0.86, 1.11)	.73	n/a	n/a	n/a	n/a	n/a	n/a
HL	1.03	(0.97, 1.09)	.40	1.06	(0.99, 1.13)	.076	1.03	(0.96, 1.10)	.43
NHL	0.98	(0.88, 1.09)	.71	0.94	(0.87, 1.02)	.15	0.96	(0.89, 1.04)	.33
Astrocytomas	0.95	(0.88, 1.02)	.15	1.01	(0.93, 1.10)	.86	0.95	(0.76, 1.18)	.62
Other CNS	1.00	(0.93, 1.07)	.94	1.02	(0.91, 1.13)	.78	1.02	(0.92, 1.12)	.75
Neuroblastoma	1.01	(0.92, 1.11)	.89	1.15	(1.05, 1.27)	.004	1.02	(0.94, 1.10)	.64
NRSTS	1.02	(0.93, 1.12)	.65	1.16	(1.02, 1.31)	.020	1.00	(0.88, 1.13)	.99

Notes: Odds ratios (OR) and 95% confidence intervals (CI) estimated from multivariable logistic regression models of SES (modeled ordinally) predicting each prognostic factor (tested separately) adjusted for age category, sex, year of diagnosis, and race. Surgery and stage not applicable (n/a) to leukemias. Combined = all cancer diagnoses; ALL = acute lymphoblastic leukemia; AML = acute myeloid leukemia; HL = Hodgkin lymphoma; NHL = non-Hodgkin lymphoma; other CNS = non-astrocytoma central nervous system tumors; NRSTS = non-rhabdomyosarcoma soft tissue sarcomas.

Chapter 5

Discussion

Key Findings

A fundamental goal of childhood cancer research is to identify non-genetic, socially patterned, risk factors of incidence and survival. This dissertation contributed to this aim by investigating three key questions in the literature: do pregnancy-related exposures contribute to the rise in childhood cancer incidence over time; is there an association between SES and childhood cancer incidence; and what is the underlying role of social versus biological factors in explaining black-white racial survival disparities? Each of these questions was addressed using robust population-based data and advanced statistical methods. Findings emerged in each of the three investigations that provide insight for future childhood cancer research.

In *manuscript* 1, we confirmed increasing temporal trends in incidence rates of combined cancers, ALL, AML, CNS tumors, and hepatoblastoma among children, 0 to 4 years of age, in the United States. We found preliminary evidence of a temporal association between increasing county-level average maternal age and county-level childhood cancer incidence rates. Specifically, we showed through descriptive analysis that the county-level temporal trend towards older average maternal age aligns with rising county-level childhood cancer incidence rates. Further, in regression analysis, we found that adjustment for only county-level average maternal age substantially attenuated the estimated average annual percent change in incidence of combined and individual childhood cancers. However, the estimated effect of maternal age on cancer incidence

trends was much weaker in models fully adjusted for other county-level pregnancy-related and sociodemographic characteristics, which suggests that temporal associations of maternal age and childhood cancer incidence may be confounded or counteracted by other risk factors. We found no evidence that county-level trends in birthweight, birth order, or sociodemographic factors are associated with the rise in childhood cancer incidence over time. Overall, we did not find conclusive evidence in support of our hypothesis that temporal trends in established pregnancy-related risk factors account for the rise in childhood cancer incidence over time.

In *manuscript 2*, we found an adverse crude association between higher SES and incidence of combined childhood cancers, ALL, lymphomas, CNS tumors, neuroblastoma, and Wilms' tumor. However, adjustment for established demographic and pregnancy-related risk factors substantially attenuated these adverse associations towards the null. In fact, no statistically significant adverse associations between higher SES and childhood cancer incidence remained in fully adjusted models. We identified non-Hispanic white race/ethnicity, older maternal age, and higher birthweight as the primary factors contributing to adverse associations between higher SES and childhood cancer incidence. Results were comparable whether SES was measured at the individual-level or neighborhood-level. These findings suggest that, beyond the social patterning of established demographic and pregnancy-related risk factors, SES has a limited impact on childhood cancer incidence, at least for the most common cancers.

In *manuscript 3*, we replicated prior findings that non-Hispanic white children have a significant survival advantage over non-Hispanic black children for combined

childhood cancers, leukemias, lymphomas, CNS tumors, neuroblastoma, and non-rhabdomyosarcoma STS. Through formal mediation analysis, we demonstrated that SES explains a portion of the black-white racial disparity in childhood cancer survival, particularly for childhood leukemias. Specifically, we estimated that SES accounts for 20% of the disparity for combined childhood cancers, and nearly half of the disparity for ALL. We also found that prognostic factors, including treatment and tumor stage at diagnosis, account for some of the survival difference by race, particularly for astrocytomas and neuroblastoma. These prognostic factors were not associated with SES among astrocytoma cases. Only surgery was associated with SES among neuroblastoma cases, though surgery was not associated with race among these cases. Thus, findings from this analysis suggest that both social and biological factors may contribute to survival differences by race, though to varying degrees across types of childhood cancer.

Future Directions

These findings provide direction for future childhood cancer research. Pregnancy-related risk factors of childhood cancer incidence are well documented in the literature, and were further confirmed in our *manuscript 2* analysis. However, we did not find conclusive evidence linking these exposures to rising childhood cancer incidence rates over time. In *manuscript 1*, comprehensive adjustment for county-level pregnancy-related and sociodemographic characteristics had little impact on the estimated average annual percent change in combined and individual childhood cancer incidence in the United States. Therefore, additional time series research is needed to better understand the cause of rising childhood cancer incidence rates over time.

Given the limitations of our *manuscript 1* study, further research with more robust data is needed to confirm temporal associations between pregnancy-related exposures and childhood cancer incidence. For example, because our study only consisted of 194 counties, future time series studies should test temporal associations in larger samples. Further, because we used data aggregated at the county-level, future studies should test temporal associations using smaller units of observation, such as census tracts. Stronger temporal associations may emerge in ecologic studies conducted at finer levels of observation in which there is less heterogeneity in exposure and covariate levels within groups.[61] We found some evidence for this in the *manuscript 1* sensitivity analysis testing temporal associations within subsets of our sample restricted by county size, which revealed that maternal age accounted for a much greater proportion of the annual trend in childhood cancer incidence among smaller counties. Birth and cancer registry data linked at the individual-level could also be used to assess temporal trends in childhood cancer incidence. For example, studies could test whether cancer incidence rates increase over time when restricted to the population of children born to older mothers (e.g. > 35 years). Future studies should also develop and test alternative hypotheses for the temporal rise in childhood cancer incidence rates over time. Because we could not conclusively link pregnancy-related exposures to increasing childhood cancer incidence rates over time, it is possible that other, potentially unidentified, risk factors account for rising rates.

In terms of developing new research directions, findings from *manuscript 2*, which tested SES patterns in childhood cancer incidence, are somewhat unsatisfying. Results from this study suggest that associations between SES and childhood cancer

incidence are mainly attributed to the social patterning of established demographic and pregnancy-related risk factors of incidence that are already well-documented in the literature. Thus, findings from this study provide limited new insight into potential childhood cancer risk factors. However, it is informative that we were able to account for adverse associations between higher SES and childhood cancer incidence, and that we found little evidence of an inverse association between SES and incidence of the most common childhood cancers. The lack of a robust association between SES and childhood cancer incidence indirectly suggests that other suspected socially patterned exposures, such as environmental pollutants, occupational exposures, and health behaviors, may contribute little to etiology. Therefore, future childhood cancer research should prioritize testing novel risk factors of incidence, rather than continuing to rehash previously hypothesized exogenous exposures that, as of yet, have not been firmly linked to childhood cancer risk.

Though we found limited evidence of an association between SES and childhood cancer incidence in *manuscript 2*, findings from *manuscript 3* suggest that higher SES is associated with improved childhood cancer survival. However, additional research is needed to follow-up on our initial findings. First, because we were confined to using an area-based index as our primary measure of SES, future studies should test more direct measures of individual-level SES, such as parental education or household income. Such research may reveal even stronger associations between SES and childhood cancer survival, given that area-level and individual-level SES may have independent effects on health. Second, additional research is needed to identify underlying pathways through

which SES influences survival, such as differences in treatment protocol and adherence. This line of research may directly inform intervention efforts addressing childhood cancer survival disparities. Finally, because SES did not fully account for survival differences in our study, future studies should explore the potential role of non-social factors in explaining childhood cancer survival differences by race. Given the limitations of SEER data, we were not equipped to account for factors such as tumor biology and treatment efficacy in our analysis. Thus, studies with more extensive data on tumor and therapeutic characteristics, as well as socioeconomic indicators, are needed to further tease apart the role of social versus biological factors in childhood cancer outcomes.

Strengths and Limitations

A primary strength of this dissertation was the use of population-based registry data in each of the three investigations. This reduced several potential sources of bias that can arise in epidemiologic studies including recall bias, interviewer bias, and selection bias. Selection bias is of particular concern when studying socioeconomic exposures in participation-based case-control studies. This is because controls in studies requiring active participation tend to be higher SES than the source population of interest.[116] Therefore, the use of registry data is a notable strength in *manuscript 2* specifically.

A second key strength of this dissertation was the use of innovative and robust datasets to test study hypotheses. We combined multiple existing data sources, some of which are not readily available for public use, through record linkage. In *manuscript 1*, we received special access to birth registry data with county-level identifiers from the National Center for Health Statistics (NCHS). NCHS provided us with birth data from all

fifty states for years 1970 to 2013. This allowed us to explore temporal associations by linking birth data to SEER cancer registry data at the county-level. To further enrich this dataset, we linked county-level sociodemographic data from the US Census Bureau. In *manuscript 2*, we collaborated with the Minnesota Department of Health to link Minnesota birth registry data to the Minnesota Cancer Surveillance System for years 1989 to 2014. This allowed us to overcome the dearth of individual-level socioeconomic data in cancer registries by merging variables from birth records. We also linked census tract socioeconomic data from the US Census Bureau, which allowed us to test associations of SES at multiple levels of exposure. Access to the Minnesota birth registry also provided us with an enumerated source cohort from which we could randomly sample controls. In *manuscript 3*, we were granted access to the census tract SES index, available in the restricted SEER database, for years 2000 to 2012. By using this variable, we improved upon prior cancer studies that relied on county-level measures, available in the public-use SEER dataset, to approximate individual-level SES.[49]

A third strength of this dissertation was the use of advanced statistical methods that are not commonly employed in the childhood cancer literature. In *manuscript 1*, we used Poisson mixed models to test temporal associations of childhood cancer incidence in time series data. In *manuscript 2*, we used logistic mixed models to test multilevel associations of SES and childhood cancer incidence. In *manuscript 3*, we used the semiparametric inverse odds weighting (IOW) method to test for mediation of associations between race and childhood cancer survival. Few studies in the childhood

cancer literature currently employ multilevel methods to test etiologic associations, and no prior study has used the IOW method to test for mediation.

Although we improved upon past research through the use of population-based registry data, there were notable limitations to the data in each of our three investigations. In *manuscript 1*, we were confined to conducting an ecologic analysis at the county-level because we lacked individual-level or smaller area (e.g. census tract) identifiers to merge birth and caner data. In *manuscript 2*, we could only test individual-level measures of SES available in birth records between 1989 and 2014, which restricted us to parental education. We could not assess other dimensions of SES at the individual-level, such as household income or parental occupation. In *manuscript 3*, we could not test SES at the individual-level, except in a secondary analysis of health insurance status, which was confined to cases diagnosed 2007-2011 (when health insurance data became available in SEER). We were also restricted by limited SEER data on tumor and therapeutic characteristics, which may be essential for fully understanding survival differences by race. Given these noted limitations, a key takeaway from this dissertation is the need for more robust data collection in population-based datasets, such as US cancer registries. In particular, there is a need for more widespread collection of socioeconomic data in these types of data sources. Only so much progress can be made to identify and understand social determinants of health with existing population-based registry data, which are often used to study rare outcomes, such as childhood cancers.

Conclusions

Though many unknowns remain in the childhood cancer literature, this dissertation offers insight into the role of non-genetic, socially patterned, risk factors of incidence and survival. We confirmed the role of socially patterned demographic and pregnancy-related risk factors in childhood cancer incidence, though we were unable to conclusively link these factors to trends in childhood cancer incidence over time. Indirectly, our findings suggest that other socially patterned exposures, such as air pollutants and occupational toxins, may contribute little to childhood cancer risk, at least for the most common cancers. Conversely, we found evidence that SES is associated with childhood cancer survival, with the strongest association emerging for childhood leukemias. Therefore, though continued study of socioeconomic exposures may generate limited new insight into childhood cancer etiology, this line of research may be fruitful for understanding and, ultimately, addressing differences in childhood cancer outcomes.

Bibliography

1. National Center for Health Statistics. Health, United States, 2010: with special feature on death and dying. Hyattsville, MD. 2011.

2. American Cancer Society. *Cancer Facts & Figures 2017*. Atlanta: American Cancer Society; 2017.

3. Spector LG, Pankratz N, Marcotte EL. Genetic and nongenetic risk factors for childhood cancer. *Pediatr Clin North Am.* 2015;62(1):11-25.

4. Howlader N, Noone A-M, Krapcho M, et al. SEER Cancer Statistics Review (CSR) 1975-2010. *Bethesda: National Cancer Institute.* 2014.

5. Ries LAG, Smith MA, Gurney J, et al. Cancer incidence and survival among children and adolescents: United States SEER Program 1975-1995, National Cancer Institute, SEER Program. NIH Pub. No. 99-4649. Bethesda MD; 1999.

6. Hasle H. Pattern of malignant disorders in individuals with Down's syndrome. *Lancet Oncol.* 2001;2(7):429-436.

7. Hasle H, Clemmensen IH, Mikkelsen M. Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet.* 2000;355(9199):165-169.

8. Stiller CA. Epidemiology and genetics of childhood cancer. *Oncogene.* 2004;23(38):6429-6444.

9. Else T. Association of adrenocortical carcinoma with familial cancer susceptibility syndromes. *Mol Cell Endocrinol.* 2012;351(1):66-70.

10. Xu H, Yang W, Perez-Andreu V, et al. Novel susceptibility variants at 10p12. 31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst.* 2013:djt042.

11. Nasca PC. Infectious Agents and Cancer. In: Nasca PC, Pastides H, eds. *Fundamentals of cancer epidemiology*: Jones & Bartlett Learning; 2001:286-333.

12. McLaughlin CC. Childhood Cancer. In: Nasca PC, Pastides H, eds. *Fundamentals of cancer epidemiology*: Jones & Bartlett Learning; 2001:443-483.

13. McNally RJ, Eden TO. An infectious aetiology for childhood acute leukaemia: a review of the evidence. *Br J Haematol.* 2004;127(3):243-263.

14. Greaves M. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia.* 1988;2(2):120-125.

15. Greaves M. Aetiology of acute leukaemia. *Lancet.* 1997;349(9048):344-349.

16. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. *Blood.* 2003;102(7):2321-2333.

17. Gale KB, Ford AM, Repp R, et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proceedings of the National Academy of Sciences.* 1997;94(25):13950-13954.

18. Wiemels J, Cazzaniga G, Daniotti M, et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet.* 1999;354(9189):1499-1503.

19. Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. *Lancet.* 1988;332(8624):1323-1327.

20. Kinlen L. Epidemiological evidence for an infective basis in childhood leukaemia. *The Journal of the Royal Society for the Promotion of Health.* 1996;116(6):393-399.

21. Kinlen LJ. An examination, with a meta-analysis, of studies of childhood leukaemia in relation to population mixing. *Br J Cancer.* 2012;107(7):1163-1168.

22. Smith M. Considerations on a possible viral etiology for B-precursor acute lymphoblastic leukemia of childhood. *J Immunother.* 1997;20(2):89-100.

23. Schmiegelow K, Vestergaard T, Nielsen S, Hjalgrim H. Etiology of common childhood acute lymphoblastic leukemia: the adrenal hypothesis. *Leukemia.* 2008;22(12):2137-2141.

24. Johnson KJ, Carozza SE, Chow EJ, et al. Parental age and risk of childhood cancer: a pooled analysis. *Epidemiology.* 2009;20(4):475-483.

25. Wilding M. Can we define maternal age as a genetic disease? *Facts Views Vis Obgyn.* 2014;6(2):105-108.

26. Hargreave M, Jensen A, Deltour I, Brinton LA, Andersen KK, Kjaer SK. Increased risk for cancer among offspring of women with fertility problems. *Int J Cancer.* 2013;133(5):1180-1186.

27. Spector LG, Puumala SE, Carozza SE, et al. Cancer risk among children with very low birth weights. *Pediatrics.* 2009;124(1):96-104.

28. Callan AC, Milne E. Involvement of the IGF system in fetal growth and childhood cancer: an overview of potential mechanisms. *Cancer Causes Control.* 2009;20(10):1783-1798.

29. Von Behren J, Spector LG, Mueller BA, et al. Birth order and risk of childhood cancer: a pooled analysis from five US States. *Int J Cancer.* 2011;128(11):2709-2716.

30. Kaijser M, Lichtenstein P, Granath F, Erlandsson G, Cnattingius S, Ekbom A. In utero exposures and breast cancer: a study of opposite-sexed twins. *J Natl Cancer Inst.* 2001;93(1):60-62.

31. English PB, Goldberg DE, Wolff C, Smith D. Parental and birth characteristics in relation to testicular cancer risk among males born between 1960 and 1995 in California (United States). *Cancer Causes Control.* 2003;14(9):815-825.

32. Bernstein L, Depue RH, Ross RK, Judd HL, Pike MC, Henderson BE. Higher maternal levels of free estradiol in first compared to second pregnancy: early gestational differences. *J Natl Cancer Inst.* 1986;76(6):1035-1039.

33. Panagiotopoulou K, Katsouyanni K, Petridou E, Garas Y, Tzonou A, Trichopoulos D. Maternal age, parity, and pregnancy estrogens. *Cancer Causes Control.* 1990;1(2):119-124.

34. Maccoby EE, Doering CH, Jacklin CN, Kraemer H. Concentrations of sex hormones in umbilical-cord blood: their relation to sex and birth order of infants. *Child Dev.* 1979;50(3):632-642.

35. Latino-Martel P, Chan DS, Druesne-Pecollo N, Barrandon E, Hercberg S, Norat T. Maternal alcohol consumption during pregnancy and risk of childhood leukemia: systematic review and meta-analysis. *Cancer Epidemiology and Prevention Biomarkers.* 2010;19(5):1238-1260.

36. Cheng J, Su H, Zhu R, et al. Maternal coffee consumption during pregnancy and risk of childhood acute leukemia: a metaanalysis. *Am J Obstet Gynecol.* 2014;210(2):151-e151.

37. Klimentopoulou A, Antonopoulos CN, Papadopoulou C, et al. Maternal smoking during pregnancy and risk for childhood leukemia: A nationwide case–control study in greece and meta-analysis. *Pediatr Blood Cancer.* 2012;58(3):344-351.

38. Goh Y, Bollano E, Einarson T, Koren G. Prenatal multivitamin supplementation and rates of pediatric cancers: a meta-analysis. *Clin Pharmacol Ther.* 2007;81(5):685-691.

39. Wigle DT, Turner MC, Krewski D. A systematic review and meta-analysis of childhood leukemia and parental occupational pesticide exposure. *Environ Health Perspect.* 2009;117(10):1505.

40. Van Maele-Fabry G, Lantin A-C, Hoet P, Lison D. Residential exposure to pesticides and childhood leukaemia: a systematic review and meta-analysis. *Environ Int.* 2011;37(1):280-291.

41. Turner MC, Wigle DT, Krewski D. Residential pesticides and childhood leukemia: a systematic review and meta-analysis. *Ciencia & saude coletiva.* 2011;16(3):1915-1931.

42. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: one size does not fit all. *JAMA.* 2005;294(22):2879-2888.

43. Carozza SE, Puumala SE, Chow EJ, et al. Parental educational attainment as an indicator of socioeconomic status and risk of childhood cancers. *Br J Cancer.* 2010;103(1):136-142.

44. Starfield B, Riley AW, Witt WP, Robertson J. Social class gradients in health during adolescence. *J Epidemiol Community Health.* 2002;56(5):354-361.

45. Starfield B, Robertson J, Riley AW. Social class gradients and health in childhood. *Ambul Pediatr.* 2002;2(4):238-246.

46. Sommer I, Griebler U, Mahlknecht P, et al. Socioeconomic inequalities in non-communicable diseases and their risk factors: an overview of systematic reviews. *BMC Public Health.* 2015;15(1):914.

47. Poole C, Greenland S, Luetters C, Kelsey JL, Mezei G. Socioeconomic status and childhood leukaemia: a review. *Int J Epidemiol.* 2006;35(2):370-384.

48. Adam M, Rebholz CE, Egger M, Zwahlen M, Kuehni CE. Childhood leukaemia and socioeconomic status: what is the evidence? *Radiat Prot Dosimetry.* 2008;132(2):246-254.

49. Krieger N. Socioeconomic data in cancer registries. *Am J Epidemiol.* 2001;91(1):156-156.

50. Pan IJ, Daniels JL, Zhu K. Poverty and childhood cancer incidence in the United States. *Cancer Causes Control.* 2010;21(7):1139-1145.

51. Kamihara J, Ma C, Fuentes Alabi SL, et al. Socioeconomic status and global variations in the incidence of neuroblastoma: call for support of population-based cancer registries in low-middle–income countries. *Pediatr Blood Cancer.* 2017;64(2):321-323.

52. Stiller C, Parkin D. Geographic and ethnic variations in the incidence of childhood cancer. *Br Med Bull.* 1996;52(4):682-703.

53. Parkin D, Kramarova E, Draper G, et al. *International incidence of childhood cancer, Volume II.* IARC scientific publications; 1998.

54. Alston RD, Rowan S, Eden TO, Moran A, Birch JM. Cancer incidence patterns by region and socioeconomic deprivation in teenagers and young adults in England. *Br J Cancer.* 2007;96(11):1760-1766.

55. Adam M, Kuehni CE, Spoerri A, et al. Socioeconomic status and childhood leukemia incidence in Switzerland. *Front Oncol.* 2015;5(139).

56. Kollerud RDR, Blaasaas KG, Claussen B. Poverty and the risk of leukemia and cancer in the central nervous system in children: a cohort study in a high-income country. *Scand J Public Health.* 2015;43(7):736-743.

57. Raaschou-Nielsen O, Obel J, Dalton S, Tjønneland A, Hansen J. Socioeconomic status and risk of childhood leukaemia in Denmark. *Scand J Public Health.* 2004;32(4):279-286.

58. Dockerty JD, Draper G, Vincent T, Rowan SD, Bunch KJ. Case-control study of parental age, parity and socioeconomic level in relation to childhood cancers. *Int J Epidemiol.* 2001;30(6):1428-1437.

59. Keegan TJ, Bunch KJ, Vincent TJ, et al. Case-control study of paternal occupation and social class with risk of childhood central nervous system tumours in Great Britain, 1962–2006. *Br J Cancer.* 2013;108(9):1907-1914.

60. Heck JE, Park AS, Contreras ZA, et al. Risk of childhood cancer by maternal birthplace: a test of the Hispanic paradox. *JAMA Pediatr.* 2016;170(6):585-592.

61. Morgenstern H. Ecologic Studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*: Lippincott Williams & Wilkins; 2008:511-531.

62. Hamilton BE, Martin JA, Osterman M, Curtin S, Matthews T. Births: Final Data for 2014. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System.* 2015;64(12):1-64.

63. Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic disparities in adverse birth outcomes: a systematic review. *Am J Prev Med.* 2010;39(3):263-272.

64. Ash M, Fetter TR. Who Lives on the Wrong Side of the Environmental Tracks? Evidence from the EPA's Risk-Screening Environmental Indicators Model. *Social Science Quarterly.* 2004;85(2):441-462.

65. van Gelder MM, Reefhuis J, Caton AR, Werler MM, Druschel CM, Roeleveld N. Characteristics of pregnant illicit drug users and associations between cannabis

use and perinatal outcome in a population-based study. *Drug Alcohol Depend.* 2010;109(1):243-247.

66. de Graaf JP, Steegers EA, Bonsel GJ. Inequalities in perinatal and maternal health. *Curr Opin Obstet Gynecol.* 2013;25(2):98-108.

67. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin.* 2014;64(2):83-103.

68. Linet MS, Ries LA, Smith MA, Tarone RE, Devesa SS. Cancer surveillance series: recent trends in childhood cancer incidence and mortality in the United States. *J Natl Cancer Inst.* 1999;91(12):1051-1058.

69. Kroll M, Stiller C, Richards S, Mitchell C, Carpenter L. Evidence for under-diagnosis of childhood acute lymphoblastic leukaemia in poorer communities within Great Britain. *Br J Cancer.* 2012;106(9):1556-1559.

70. Adamson P, Law G, Roman E. Assessment of trends in childhood cancer incidence. *Lancet.* 2005;365(9461):753.

71. Mangano JJ. A rise in the incidence of childhood cancer in the United States. *Int J Health Serv.* 1999;29(2):393-408.

72. Linabery AM, Ross JA. Trends in childhood cancer incidence in the US (1992–2004). *Cancer.* 2008;112(2):416-432.

73. Bandi P, Dranger E, Hampton JM, Trentham-Dietz A. Trends in childhood cancer incidence in Wisconsin, 1980-1999. *WMJ-MADISON-.* 2006;105(7):30.

74. United States Environmental Protection Agency. Air quality national summary. 2017; https://www.epa.gov/air-trends/air-quality-national-summary. Accessed May 31, 2017.

75. Oken E. Secular trends in birthweight. *Recent Advances in Growth Research: Nutritional, Molecular and Endocrine Perspectives.* Vol 71: Karger Publishers; 2013:103-114.

76. Morisaki N, Esplin MS, Varner MW, Henry E, Oken E. Declines in birth weight and fetal growth independent of gestational length. *Obstet Gynecol.* 2013;121(1):51.

77. Maule MM, Merletti F, Pastore G, Magnani C, Richiardi L. Effects of maternal age and cohort of birth on incidence time trends of childhood acute lymphoblastic leukemia. *Cancer Epidemiology Biomarkers & Prevention.* 2007;16(2):347-351.

78. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Research

Data, Nov 2015 Sub (1973-2013) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969-2014 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2016, based on the November 2015 submission.

79. Bhatia S. Influence of race and socioeconomic status on outcome of children treated for childhood acute lymphoblastic leukemia. *Curr Opin Pediatr.* 2004;16(1):9-14.

80. Hossain MJ, Xie L, Caywood EH. Prognostic factors of childhood and adolescent acute myeloid leukemia (AML) survival: Evidence from four decades of US population data. *Cancer Epidemiol.* 2015;39(5):720-726.

81. Abrahão R, Keogh RH, Lichtensztajn DY, et al. Predictors of early death and survival among children, adolescents and young adults with acute myeloid leukaemia in California, 1988–2011: a population-based study. *Br J Haematol.* 2016;173(2):292-302.

82. Grubb W, Neboori H, Diaz A, Li H, Kwon D, Panoff J. Racial and ethnic disparities in the pediatric Hodgkin lymphoma population. *Pediatr Blood Cancer.* 2016;63(3):428-435.

83. Kent EE, Breen N, Lewis DR, de Moor JS, Smith AW, Seibel NL. US trends in survival disparities among adolescents and young adults with non-Hodgkin lymphoma. *Cancer Causes Control.* 2015;26(8):1153-1162.

84. Keegan TH, Clarke CA, Chang ET, Shema SJ, Glaser SL. Disparities in survival after Hodgkin lymphoma: a population-based study. *Cancer Causes Control.* 2009;20(10):1881-1892.

85. Austin MT, Hamilton E, Zebda D, et al. Health disparities and impact on outcomes in children with primary central nervous system solid tumors. *J Neurosurg Pediatr.* 2016;18(5):585-593.

86. Henderson TO, Bhatia S, Pinto N, et al. Racial and ethnic disparities in risk and survival in children with neuroblastoma: a Children's Oncology Group study. *J Clin Oncol.* 2010;29(1):76-82.

87. Johnson KA, Aplenc R, Bagatell R. Survival by race among children with extracranial solid tumors in the United States between 1985 and 2005. *Pediatr Blood Cancer.* 2011;56(3):425-431.

88. Waxweiler TV, Rusthoven CG, Proper MS, et al. Non-Rhabdomyosarcoma Soft Tissue Sarcomas in Children: A Surveillance, Epidemiology, and End Results Analysis Validating COG Risk Stratifications. *Int J Radiat Oncol Biol Phys.* 2015;92(2):339-348.

89. Furlow B. Ethnic disparities in childhood cancer survival: Biology, socioeconomics, or both? *Cancer Therapy Advisor. 2013*.

90. Bhatia S. Disparities in cancer outcomes: lessons learned from children with cancer. *Pediatr Blood Cancer.* 2011;56(6):994-1002.

91. Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nat Genet.* 2004;36:S28-S33.

92. Kadan-Lottick NS, Ness KK, Bhatia S, Gurney JG. Survival variability by race and ethnicity in childhood acute lymphoblastic leukemia. *JAMA.* 2003;290(15):2008-2014.

93. Bhatia S, Sather HN, Heerema NA, Trigg ME, Gaynon PS, Robison LL. Racial and ethnic differences in survival of children with acute lymphoblastic leukemia. *Blood.* 2002;100(6):1957-1964.

94. Pollock BH, DeBaun MR, Camitta BM, et al. Racial differences in the survival of childhood B-precursor acute lymphoblastic leukemia: a Pediatric Oncology Group Study. *J Clin Oncol.* 2000;18(4):813-813.

95. Metzger ML, Castellino SM, Hudson MM, et al. Effect of race on the outcome of pediatric patients with Hodgkin's lymphoma. *J Clin Oncol.* 2008;26(8):1282-1288.

96. Baker KS, Anderson JR, Lobe TE, et al. Children from ethnic minorities have benefited equally as other children from contemporary therapy for rhabdomyosarcoma: a report from the Intergroup Rhabdomyosarcoma Study Group. *J Clin Oncol.* 2002;20(22):4428-4433.

97. Relling MV, Hancock ML, Rivera GK, et al. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J Natl Cancer Inst.* 1999;91(23):2001-2008.

98. Collie-Duguid E, Sludden J, Li T, McLeod H. The frequency and distribution of thiopurine methyltransferase alleles in Caucasian and Asian populations. *Pharmacogenet Genomics.* 1999;9(1):37-42.

99. Hon YY, Fessing MY, Pui C-H, Relling MV, Krynetski EY, Evans WE. Polymorphism of the thiopurine S-methyltransferase gene in African-Americans. *Hum Mol Genet.* 1999;8(2):371-376.

100. McLeod HL, Lin JS, Scott EP, Pui CH, Evans WE. Thiopurine methyltransferase activity in American white subjects and black subjects. *Clin Pharmacol Ther.* 1994;55(1):15-20.

101. McLeod HL, Pritchard SC, Githang J, et al. Ethnic differences in thiopurine methyltransferase pharmacogenetics: evidence for allele specificity in Caucasian and Kenyan individuals. *Pharmacogenet Genomics.* 1999;9(6):773-776.

102. Cooper SC, Ford LT, Berg JD, Lewis MJ. Ethnic variation of thiopurine S-methyltransferase activity: a large, prospective population study. 2008.

103. Williams DR, Collins C. US socioeconomic and racial differences in health: patterns and explanations. *Annu Rev Sociol.* 1995;21(1):349-386.

104. LaVeist TA. Disentangling race and socioeconomic status: a key to understanding health inequalities. *J Urban Health.* 2005;82:iii26-iii34.

105. Jones CP. Levels of racism: a theoretic framework and a gardener's tale. *Am J Public Health.* 2000;90(8):1212.

106. Gupta S, Wilejto M, Pole JD, Guttmann A, Sung L. Low socioeconomic status is associated with worse survival in children with cancer: a systematic review. *PLoS One.* 2014;9(2):e89482.

107. Petridou ET, Sergentanis T, Perlepe C, et al. Socioeconomic disparities in survival from childhood leukemia in the United States and globally: a meta-analysis. *Ann Oncol.* 2014;26(3):589-597.

108. Austin MT, Nguyen H, Eberth JM, et al. Health disparities are important determinants of outcome for children with solid tumor malignancies. *J Pediatr Surg.* 2015;50(1):161-166.

109. Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL. Trends in obesity among adults in the United States, 2005 to 2014. *JAMA.* 2016;315(21):2284-2291.

110. Ogden CL, Carroll MD, Lawman HG, et al. Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. *JAMA.* 2016;315(21):2292-2299.

111. Samuelsen SO, Bakketeig LS, Tretli S, Johannesen TB, Magnus P. Birth weight and childhood cancer. *Epidemiology.* 2009;20(4):484-487.

112. Urayama KY, Von Behren J, Reynolds P, Hertz A, Does M, Buffler PA. Factors associated with residential mobility in children with leukemia: implications for assigning exposures. *Ann Epidemiol.* 2009;19(11):834-840.

113. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International classification of childhood cancer, third edition. *Cancer.* 2005;103(7):1457-1467.

114. StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

115. Kehm RD, Spector LG, Poynter JN, Vock DM, Osypuk TL. Socioeconomic status and childhood cancer incidence: a population-based multilevel analysis. *Under Review*.

116. Puumala SE, Spector LG, Robison LL, et al. Comparability and representativeness of control groups in a case-control study of infant leukemia: a report from the Children's Oncology Group. *Am J Epidemiol.* 2009;170(3):379-387.

117. Minnesota Cancer Surveillance System. *Cancer in Minnesota, 1988-2009.* St. Paul, Minnesota: Minnesota Department of Health;2012.

118. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med.* 1995;14(5-7):491-498.

119. Registry Plus, a suite of publicly available software programs for collecting and processing cancer registry data. Atlanta (GA): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 2015. Available at: http://www.cdc.gov/cancer/npcr/. Accessed 11/22/2016.

120. Rothman KJ, Greenland S, Lash TL. Case-Control Studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*: Lippincott Williams & Wilkins; 2008:111-127.

121. Maloney KW, Taub JW, Ravindranath Y, Roberts I, Vyas P. Down syndrome preleukemia and leukemia. *Pediatr Clin North Am.* 2015;62(1):121-137.

122. Cybulski C, Nazarali S, Narod SA. Multiple primary cancers as a guide to heritability. *Int J Cancer.* 2014;135(8):1756-1763.

123. Hauser RM. Measuring socioeconomic status in studies of child development. *Child Dev.* 1994;65(6):1541-1545.

124. Zill N. Parental schooling & children's health. *Public Health Rep.* 1996;111(1):34-43.

125. Northam S, Knapp TR. The reliability and validity of birth certificates. *J Obstet Gynecol Neonatal Nurs.* 2006;35(1):3-12.

126. Goldberg, D. W., Kohler, B., Kosary, C. (year). The Texas A&M, NAACCR, NCI Geocoding Service. Available online at http://geo.naaccr.org.

127. Tabachnick BG, Fidell LS. Principal Components and Factor Analysis. *Using multivariate statistics*. 3 ed. Northridge, California: California State University, Harper Collins College; 1996:635-708.

128. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health.* 2006;83(6):1041-1062.

129. Osypuk TL, Kehm R, Misra DP. Where We Used to Live: Validating Retrospective Measures of Childhood Neighborhood Context for Life Course Epidemiologic Studies. *PLoS One.* 2015;10(4):e0124635.

130. Pagel C, Prost A, Lewycka S, et al. Intracluster correlation coefficients and coefficients of variation for perinatal outcomes from five cluster-randomised controlled trials in low and middle-income countries: results and methodological implications. *Trials.* 2011;12(1):151.

131. Royston P, White IR. Multiple imputation by chained equations (MICE): implementation in Stata. *J Stat Softw.* 2011;45(4):1-20.

132. Spector LG, Birch J. The epidemiology of hepatoblastoma. *Pediatr Blood Cancer.* 2012;59(5):776-779.

133. Osypuk TL, Slaughter-Acey JC, Kehm RD, Misra DP. Life-course social mobility and reduced risk of adverse birth outcomes. *Am J Prev Med.* 2016;51(6):975-982.

134. Minnesota State Demographic Center. Minnesota on the move: Migration patterns & implications. 2015; https://mn.gov/bms-stat/assets/mn-on-the-move-migration-report-msdc-jan2015.pdf. Accessed 6/27/2016.

135. Puumala SE, Soler JT, Johnson KJ, Spector LG. Birth characteristics and Wilms tumor in Minnesota. *Int J Cancer.* 2008;122(6):1368-1373.

136. McLaughlin CC, Baptiste MS, Schymura MJ, Nasca PC, Zdeb MS. Maternal and infant birth characteristics and hepatoblastoma. *Am J Epidemiol.* 2006;163(9):818-828.

137. Reynolds P, Von Behren J, Elkin EP. Birth characteristics and leukemia in young children. *Am J Epidemiol.* 2002;155(7):603-613.

138. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (2000-2012), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released March 2016, based on the November 2014 submission.

139. Division of Cancer Prevention and Control, Centers for Disease Control and Prevention. Interpreting race and ethnicity in cancer data. *United States Cancer Statistics* 2016.

140. Armstrong GT, Pan Z, Ness KK, Srivastava D, Robison LL. Temporal trends in cause-specific late mortality among 5-year survivors of childhood cancer. *J Clin Oncol.* 2010;28(7):1224-1231.

141. Yu M, Tatalovich Z, Gibson JT, Cronin KA. Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. *Cancer Causes Control.* 2014;25(1):81-92.

142. Kish JK, Yu M, Percy-Laurry A, Altekruse SF. Racial and ethnic disparities in cancer survival by neighborhood socioeconomic status in Surveillance, Epidemiology, and End Results (SEER) Registries. *J Natl Cancer Inst Monogr.* 2014;2014(49):236-243.

143. Liu L, Deapen D, Bernstein L. Socioeconomic status and cancer of females. *Cancer Causes Control.* 1988;9:369-380.

144. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control.* 2001;12(8):703-711.

145. Fritz AG, RHIT C, Hurlbut AA, et al. SEER Summary Staging Manual-2000 Codes and Coding Instructions. 2001.

146. Tchetgen Tchetgen EJ, Shpitser I. Semiparametric estimation of models for natural direct and indirect effects. Berkeley, CA: bepress; 2011. (Harvard University Biostatistics Working Paper Series. Working Paper 129.

147. Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Stat Med.* 2013;32(26):4567-4580.

148. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51(6):1173.

149. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen EJT. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. *Am J Epidemiol.* 2015;181(5):349-356.

150. Pearl J. Direct and indirect effects. *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*. San Francisco, CA: Morgan Kaufmann 2001:411-420.

151. Nguyen TT, Tchetgen EJT, Kawachi I, Gilman SE, Walter S, Glymour MM. Comparing alternative effect decomposition methods: the role of literacy in mediating educational effects on mortality. *Epidemiology.* 2016;27(5):670-676.

152. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*. 1995:289-300.

153. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www/R-project.org/. Accessed May 24, 2017.

154. Bhatia S, Landier W, Shangguan M, et al. Nonadherence to oral mercaptopurine and risk of relapse in Hispanic and non-Hispanic white children with acute lymphoblastic leukemia: a report from the children's oncology group. *J Clin Oncol.* 2012:JCO. 2011.2038. 9924.

155. Jabo B, Morgan JW, Martinez ME, Ghamsary M, Wieduwilt MJ. Sociodemographic disparities in chemotherapy and hematopoietic cell transplantation utilization among adult acute lymphoblastic and acute myeloid leukemia patients. *PLoS One.* 2017;12(4):e0174760.

156. Mitchell JM, Conklin EA. Factors affecting receipt of expensive cancer treatments and mortality: evidence from stem cell transplantation for leukemia and lymphoma. *Health Serv Res.* 2015;50(1):197-216.

157. Knight JM, Rizzo JD, Logan BR, et al. Low socioeconomic status, adverse gene expression profiles, and clinical outcomes in hematopoietic stem cell transplant recipients. *Clin Cancer Res.* 2016;22(1):69-78.

158. Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *Am J Public Health.* 1992;82(5):703-710.

159. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. *Am J Epidemiol.* 1998;148(5):475-486.

160. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern epidemiology*. 3 ed: Lippincott Williams & Wilkins; 2008.

161. Gupta S, Aitken JF, Bartels U, et al. Paediatric cancer stage in population-based cancer registries: the Toronto consensus principles and guidelines. *Lancet Oncol.* 2016;17(4):e163-e172.

162. Pinheiro PS, Morris CR, Liu L, Bungum TJ, Altekruse SF. The impact of follow-up type and missed deaths on population-based cancer survival studies for Hispanics and Asians. *J Natl Cancer Inst Monogr.* 2014;2014(49):210.