

# Trends and patterns of childhood cancer incidence in the United States, 1995–2010

LI ZHU\*†, LINDA W. PICKLE, ZHAOHUI ZOU, AND JAMES CUCINELLI

*Background:* The American Cancer Society publishes *Cancer Facts & Figures* every year to provide estimated incidence rates and case counts for the current calendar year. Childhood cancer (children aged 0–14 with all cancer) incidence was also estimated and reported in *Cancer Facts & Figures*, but there is no report on patients in the 15–19 years of age or combined 0–19 years of age, or on major childhood cancers such as leukemia. There may be different effects of socioeconomic status and geography on childhood cancer in comparison to cancer in adults. Therefore, models used to predict cancer case counts for all patients may need to be modified to accurately evaluate and predict incidence rates of childhood cancers.

*Methods:* The original method was developed in 2007 to project the total cancer incidence counts reported in the *Cancer Facts & Figures*. This method requires first a spatiotemporal prediction across the extensive years of data since not all cancer registries have data for every year, and secondly a 4-year temporal projection from the most current data point to the current calendar year. Here we use this same process applied only to childhood cancer data. The generalized linear mixed effects model is applied to observed childhood cancer case counts reported to the North American Association of Central Cancer Registries over 1995–2006 to predict case counts and incidence rates for every U.S. state and the U.S. total (for patients aged 0 to 19 and for major cancers among childhood). Covariates included in the model are measures of income, education, housing, urban/rural status, health insurance coverage, smoking, obesity and cancer screening. Temporal trends and spatial distribution patterns are compared among childhood cancers for males and females.

*Results:* A total of 15,168 new cancer cases was projected for the 0–19 age group, with 10,032 childhood cancer cases younger than age 15, and 5,136 cases in the 15–19 age group, in the U.S. for 2010. There are more male cancer cases than female cases in all three age groups for most cancer sites. The order for the most common malignancies diagnosed in

children differs by age group and sex. The study also identifies differences in the geographic patterns at a small geographic area level by gender and temporal trends by gender and subsites of childhood cancer incidence.

*Conclusions:* The resulting set of predictions provides annual estimates for states that did not provide data at all, and projections ahead in time to the current calendar year for every U.S. state and the U.S. in total. These projections fill in the data gaps for recent years of diagnosis and state registries to provide complete count and rate estimates for childhood cancers for all states, regions, and the U.S. in total for the current calendar year.

**KEYWORDS AND PHRASES:** Childhood cancer, Incidence, Projection method, Joinpoint regression, Spatial pattern, Temporal trend.

## ABBREVIATIONS

NAACCR: North American Association of Central Cancer Registries

ACS: American Cancer Society

CFF: Cancer Facts & Figures

SEER: Surveillance, Epidemiology, and End Results

HSA: Health Service Area

BRFSS: Behavioral Risk Factor Surveillance System

CV: Coefficient of Variation.

## 1. INTRODUCTION

Cancer surveillance, i.e., an assessment of the number and type of cancer cases that occur in an area and in a specified population, provides critically important information needed for determining geographic variations in cancer rates and for planning cancer control intervention programs. However, these data are not available for every U.S. state for a sufficiently long period of time to provide stable estimates of their rates, case counts and trends – currently not all state registries are certified as high quality by the North American Association of Central Cancer Registries (NAACCR) and data are available for varying time periods from different states. Furthermore, data collection and processing require several years to complete so that information is lacking about the expected cancer burden in the near future, important information for cancer control planning. The American Cancer Society (ACS) publishes *Cancer Facts & Figures* (CFF) which provides estimated incidence rates and case counts for the current calendar year. The

\*Corresponding author.

†Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-8317, United States. Tel: +1 301 594 6546. Fax: +1 301 480 2046.

methods used to produce these estimates have changed over the years as more incidence data have become available and statistical models improved. Prior to 2007, the statistical projection method to produce the cancer incidence was an autoregressive quadratic time trend model [1], under the assumption that the pooled incidence-to-mortality ratios for nine cancer registries with high-quality data can be taken as constant across all states. Cancer registries have now been established in every state and territory in the U.S., and high-quality incidence data are available to most for several years. A new method was developed in 2007 to improve the ACS incidence projections by incorporating potential predictors and spatial and temporal variation of cancer occurrence [2], as well as delay in case reporting [3]. This method requires first a spatiotemporal prediction across the extensive years of data since not all cancer registries have data for every year, and secondly a 4-year temporal projection from the most current data point to the current calendar year. The 4-year projection is required here since the time lag between the data report and release is 4 years. For example, *Cancer Facts & Figures 2010* was produced based on incidence data up to a diagnosis year of 2006. ACS has adopted the method since 2007 to predict the number of new cancer cases since it produces more accurate estimates for years and areas where data are available for comparison [4]. CFF has estimated and reported childhood cancer for ages 0–14 for many years, but there is no report on patients for ages 15–19 or combined ages 0–19, or for some major childhood cancers such as leukemia. The effect of socio-economic and geographic factors may be different from the effect of these factors on adult cancers so that the models applied to adults for CFF may not be best for explaining childhood cancer patterns. Therefore, models used to predict cancer case counts for all patients may need to be modified to accurately evaluate and predict incidence rates of childhood cancers.

The purpose of this paper is to model observed childhood cancer case counts from 1995–2006 to predict annual case counts and incidence rates for every U.S. state and the U.S. total from 1995–2010 for ages 0–19 and major childhood cancers. The resulting set of predictions will provide: (1) annual estimates for states included in the NAACCR dataset for which year-to-year variation due to small populations is reduced with a comparison to the observed counts and rates, (2) annual estimates for states that did not provide any data, and (3) projections ahead in time to the current calendar year for every U.S. state and the total U.S. These projections fill in the data gaps in the cancer registry reporting to provide complete count and rate estimates for childhood cancers for all states, regions, and the U.S. in total for the current calendar year.

## 2. METHODS

### 2.1 Data source

To estimate the number of new cases in states without observed data, a dataset with good spatial coverage in all

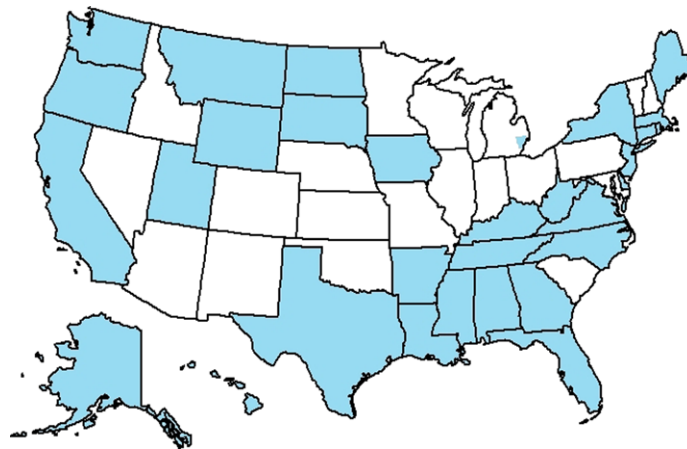


Figure 1. Registries Included in 1995–2006 Cancer Incidence Data for this Project Available from NAACCR.

regions of the United States is critical. We obtained an incidence database spanning from 1995 to 2006 from NAACCR. U.S. cancer registries reporting data to NAACCR participate in the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program [5] or the Centers for Disease Control and Prevention’s National Program of Cancer Registries [6], or both, and receive additional support from the state where they are located. Registries for 30 states and the Detroit metropolitan area (Figure 1) met NAACCR registry certification standards as providing complete, accurate, and timely data for at least 3 consecutive years between 1995 to 2006 and agreed to release county-level incidence data for this study. Not every registry included in this project has certified data for every year. Together, these registries cover 72% of the U.S. population. Models were initially applied to data at the county level to capture within-state variation of incidence counts. However, due to the rareness of childhood cancer, the data were aggregated to the Health Service Area (HSA) level to provide more stable estimates for new case counts. The HSA units were originally defined by the National Center for Health Statistics [7] to be groups of counties defined on the basis of where county residents aged 65 years and over obtained hospital care. The National Cancer Institute modified the original HSA definition so that any HSAs that crossed state or SEER Registry boundaries were split so that all counties from one HSA were in one state and/or SEER Registry [8]. There are 944 HSAs in the U.S according to the modified HSA definition.

The cancer site was coded according to the SEER Program recodes [9] in the same manner used for the SEER Cancer Statistics Review [10]. While these groups are not as appropriate for children as they are for adults, they are necessary in order to be able to compare incidence to mortality. The sites considered in this paper include “All” (all primary sites combined) and the eight most common childhood cancers: “Bone & Joint”, “Brain and Other Nervous

System” (Brain, Cranial Nerves Other Nervous System), “Hodgkin Lymphoma”, “Non-Hodgkin Lymphoma”, “Kidney and Renal Pelvis”, “Acute Lymphocytic Leukemia”, “Acute Myeloid Leukemia”, and “Soft Tissue including Heart”. Together, the eight cancer sites account for 77% and 71% of all primary cancer sites in the 0–14 and 0–19 age groups, respectively. Only primary malignancies are included. Race was grouped as White, Black, and Other. Age at diagnosis was coded as age groups < 1 year, 1–4 years, 5–9 years, 10–14 years, and 15–19 years. Unless otherwise specified, “childhood cancer” in this paper always refers to cancer in the 0–19 age group. Input to the models consisted of numbers of new cases stratified by site, sex, race, age group, HSA, and year of diagnosis. Similarly stratified populations were obtained from the SEER Program which are the Census Bureau estimates with small SEER modifications, which account for the intercensal and Vintage annual time series of July 1 population estimates [11]. Population characteristics for the HSA, including measures of income, education, housing, racial distribution, urban/rural status, availability of physicians and cancer screening facilities, health insurance coverage, cigarette smoking, obesity, cancer screening rates, and mortality rates were considered as potential predictors of incidence. While some behavior risk factors such as cigarette smoking and cancer screening behavior are more relevant to adult cancer than childhood cancer, they were included in the model as indicators of the general health attitude, knowledge and behavior of the community. The covariates were available for every U.S. county from a variety of sources, including the Census Bureau, Area Resource File [12], Behavioral Risk Factor Surveillance System<sup>1</sup> (BRFSS) [13] and the Mortality Data from the National Vital Statistics System [14]. The HSA-level covariate values were aggregated from the county-level values weighted by the population size. A total of 35 covariates and interactions were considered in the original model. After backward variable selection for each cancer site, the final number of covariates (including interactions) are between 11 and 34.

## 2.2 Statistical methods

The method was developed in 2007 and has been applied to produce the number of new cancer cases in CFF [4] and the United States Cancer Statistics [15]. The method applied statistical models of cancer incidence that incorporate potential predictors and spatial and temporal variation of cancer occurrence and, in a second step, then adjust for delay in case reporting. A hierarchical Poisson regression model was used to estimate the number of cases for all HSAs in the U.S. by their demographic and lifestyle profiles, based on the association of these profiles with cancer incidence in

the HSAs with observed data, as well as age, race, mortality rate, diagnosis year, and spatial and temporal random effects. Specifically, the number of new cancer cases in HSA  $i$ , age group  $j$ , and year  $t$ , denoted  $y_{ijt}$ , was assumed to be distributed as a Poisson random variable, with mean  $n_{ijt}\lambda_{ijt}$  where  $n_{ijt}$  is the corresponding population at risk and  $\lambda_{ijt}$  is the incidence rate. A log-linear rate structure was assumed with

$$\log(\lambda_{ijt}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{\delta}) = \alpha_r + f(a_j)\boldsymbol{\beta} + \log(m_{ijt})\gamma + \mathbf{X}_i\boldsymbol{\delta} + \theta_s + \phi_t$$

where  $\alpha_r$  is the intercept for census region,  $\alpha_j$  is the mid-point of age group,  $f()$  is a cubic function of age groups that accounts for non-linearity of some cancer rates among age groups,  $m_{ijt}$  is the age-specific mortality rate, and  $\mathbf{X}_i$  is the vector of socio-demographic and life style covariates.  $\theta_s$  and  $\phi_t$  account for spatial and temporal random effects, respectively.

This model was implemented using SAS PROC GLIMMIX procedure [16] with its optional spline-based approximation for spatial and temporal autocorrelation. The spatial projection component of the method, i.e. estimation of numbers of new cases in states without observed data, has been previously used to produce state- and county-level maps of estimated cancer incidence in 1999 [17]. The spatial projection was conducted at the HSA level to capture within-state variation of incidence counts. Results of the spatiotemporal models are site- and sex-specific smoothed annual estimates for registries that provided data and annual modeled estimates for registries with missing data. The assumed spatial and temporal autocorrelation plus covariates included in the model result in a sharing of information across areas that are similar in location, time, and socio-demographic characteristics. For example, the number of new cases in registries with no input data was estimated using several years of data from neighboring states and from other states and HSAs with similar socio-demographic and lifestyle profiles. Estimated numbers for a registry with a single missing year of data were based on observations from that registry before and after the missing time point, as well as from states that are neighbors or have similar characteristics. Model estimates were added over race, age and HSA to produce state-year-sex-cancer-specific estimates for the time span of the available incidence data. These estimated numbers were then adjusted to account for the delay expected in reporting cancer cases to the registry [3]. The delay adjustment factor varies by diagnosis year, cancer site, and age at diagnosis. The age-adjusted incidence rate (per 100,000 persons) was calculated as an average of the delay-adjusted age-specific incidence rates weighted by the standard populations [18]. The delay-adjusted incidence count or delay-and-age-adjusted incidence rates for 1995–2006 were then projected 4 years ahead to 2010 using a piecewise linear regression method called the Joinpoint program [19, 20] which

<sup>1</sup>During our study period, BRFSS follows a rule of not reporting or interpreting percentages from questions whose unweighted number of valid responses is less than 50. Hence for counties where BRFSS data are not available, we use state BRFSS data as surrogates in our study.

Table 1. Projected New Childhood Cancer Cases by Sex for Selected Cancers, U.S., 2010. Numbers in parentheses are percentages of each cancer in the age group/sex all sites total

	Age 0–14			Age 15–19			Age 0–19		
	Total	Female	Male	Total	Female	Male	Total	Female	Male
<b>All Sites</b>	<b>10032</b>	<b>4617</b>	<b>5415</b>	<b>5136</b>	<b>2331</b>	<b>2805</b>	<b>15168</b>	<b>6948</b>	<b>8220</b>
Bone & Joint	445 (4%)	217 (5%)	228 (4%)	373 (7%)	130 (6%)	243 (9%)	818 (5%)	347 (5%)	471 (6%)
Brain & Other Nervous	2063 (21%)	986 (21%)	1077 (20%)	475 (9%)	218 (9%)	257 (9%)	2538 (17%)	1204 (17%)	1334 (16%)
Kidney & Renal Pelvis	556 (6%)	324 (7%)	232 (4%)	47 (1%)	22 (1%)	25 (1%)	603 (4%)	346 (5%)	257 (3%)
Leukemia (Total)	3078 (31%)	1335 (29%)	1744 (32%)	690 (13%)	220 (9%)	469 (17%)	3768 (25%)	1555 (22%)	2213 (27%)
Acute Lymphocytic Leukemia	2574 (26%)	1022 (22%)	1552 (29%)	431 (8%)	121 (5%)	310 (11%)	3005 (20%)	1143 (16%)	1862 (23%)
Acute Myeloid Leukemia	385 (4%)	193 (4%)	192 (4%)	183 (4%)	86 (4%)	97 (3%)	568 (4%)	279 (4%)	289 (4%)
Lymphoma (Total)	937 (9%)	321 (7%)	616 (11%)	1256 (24%)	528 (23%)	728 (26%)	2193 (14%)	849 (12%)	1344 (16%)
Hodgkin Lymphoma	365 (4%)	140 (3%)	225 (4%)	772 (15%)	371 (16%)	401 (14%)	1137 (7%)	511 (7%)	626 (8%)
Non-Hodgkin Lymphoma	570 (6%)	178 (4%)	392 (7%)	453 (9%)	141 (6%)	312 (11%)	1023 (7%)	319 (5%)	704 (9%)
Soft Tissue Including Heart	600 (6%)	294 (6%)	306 (6%)	257 (5%)	121 (5%)	136 (5%)	857 (6%)	415 (6%)	442 (5%)
Other	2353 (23%)	1140 (25%)	1212 (22%)	2038 (40%)	1092 (47%)	947 (34%)	4391 (29%)	2232 (32%)	2159 (26%)

is currently used to describe trends in incidence and mortality for CFF, SEER Cancer Statistics Review, and Annual Report to the Nation on the Status of Cancer [21]. Model fit was assessed using the pseudo-likelihood statistics. To validate the proposed method, a total of nine states (California, Connecticut, Georgia, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, and Utah) was selected since the most recent data have been available for the 2010 year of diagnosis. The 4-year projection produced an estimated total of 3,898 childhood cancer cases in 2010 for the nine states, compared to the total of 3,745 observed cases. Overall, the projection produces an overestimate of 4.1% with similar scales in female (5.3%) and male (3.1%). Since the projection considers reporting delay while the observed case counts do not adjust for the delay factor, the overestimate is considered reasonable.

In summary, the projected incidence counts and rates of childhood cancer in 2010 for each sex/cancer site stratification were produced by the following steps:

- 1) Apply the spatiotemporal model to data from registries available from NAACCR from 1995 to 2006 to estimate the numbers of cases by sex, race, age group, and diagnosis year for every HSA in the U.S.
- 2) Sum these estimates over age group, race, and HSA to calculate estimated numbers by sex and year for every state and for the total U.S., then adjust these results to account for case reporting delay.

- 3) Sum the estimates from Step 1) over race and HSA to calculate either age-specific or age-adjusted rate (per 100,000 persons) for each sex/site/state stratification.
- 4) Sum the estimates from Step 1) over race and site to calculate the HSA-level age-adjusted rate for all cancer sites combined (further site stratification is not possible due to sparse data at this small area level).
- 5) Apply the Joinpoint Regression Program to project each state's or U.S. time series of the delay-adjusted estimated number of new cases or delay-and-age-adjusted rates to 2010.

### 3. RESULTS

#### 3.1 Overview

A total of 15,168 new cancer cases was projected for the 0–19 age group, with 10,032 childhood cancer cases younger than age 15, and 5,136 cases in the 15–19 age group, in the U.S. for 2010 (Table 1). In comparison, the CFF 2010 reported an estimated 10,700 new cases among children aged 0 to 14 years [22] in 2010. There are more male cancer cases than female cases in all three age groups for most cancer sites, except for kidney & renal pelvis with more female cases in 0–14 age group, and acute myeloid leukemia with approximately equal numbers of female and male cases in the 0–14 age group. In the 0–14 age group, the most common childhood cancer site is acute lymphocytic leukemia,

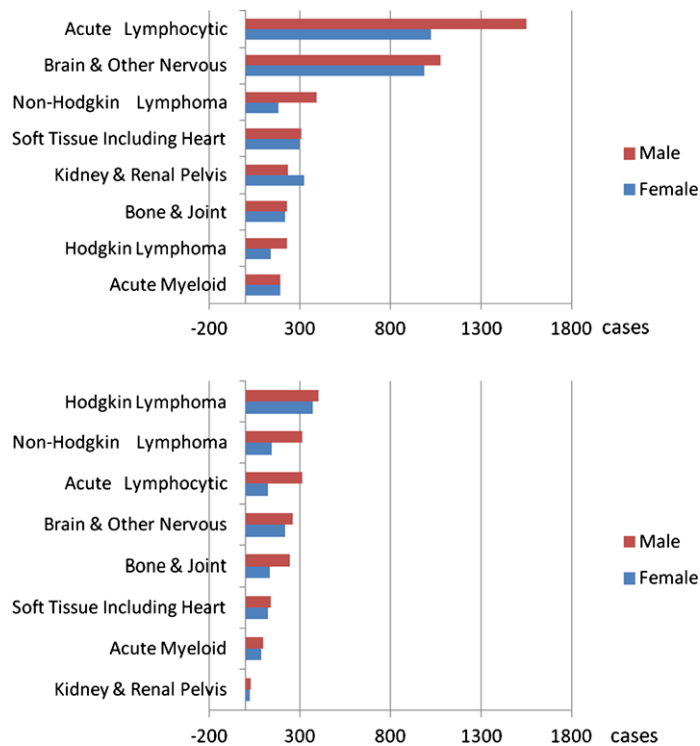


Figure 2. Projected New Childhood Cancer Cases by Sex and Age Group by Site in 2010 (Sorted by male new cases; Top: Age 0–14; Bottom: Age 15–19).

with a total of 2,574 new cases projected for the year 2010. Malignant tumors of the brain and other nervous system are the second most common cancer in this age groups. In the 15–19 age group, the most common cancer is Hodgkin lymphoma, which accounts for 15% of all cancer incidence counts in this age group, followed by malignant tumors of the brain and the other nervous system, non-Hodgkin lymphoma, and acute lymphocytic leukemia. What is unique about lymphoma is that the adolescent (age 15–19) cases account for the majority of all cases, especially for Hodgkin lymphoma among both sexes. The order for the most common malignancies diagnosed in children differs by age group and sex (Figure 2). In the 0–14 age group, in order of decreasing incidence for males are acute lymphocytic leukemia, brain & other nervous system cancer, non-Hodgkin lymphoma, cancer of soft tissue including heart, cancer of kidney & renal pelvis, cancer of bone & joint, Hodgkin lymphoma, and acute myeloid leukemia. Females have a slightly different order in the same age group. The most common cancer sites among males in the 15–19 age group are the lymphomas (Hodgkin lymphoma and non-Hodgkin lymphoma), followed by acute lymphocytic leukemia, cancer of brain & other nervous system, and cancer of bone and joint. Among females in the 15–19 age group, the most common cancers are Hodgkin lymphoma and cancer of brain & other nervous system, followed by non-Hodgkin lymphoma, cancer of

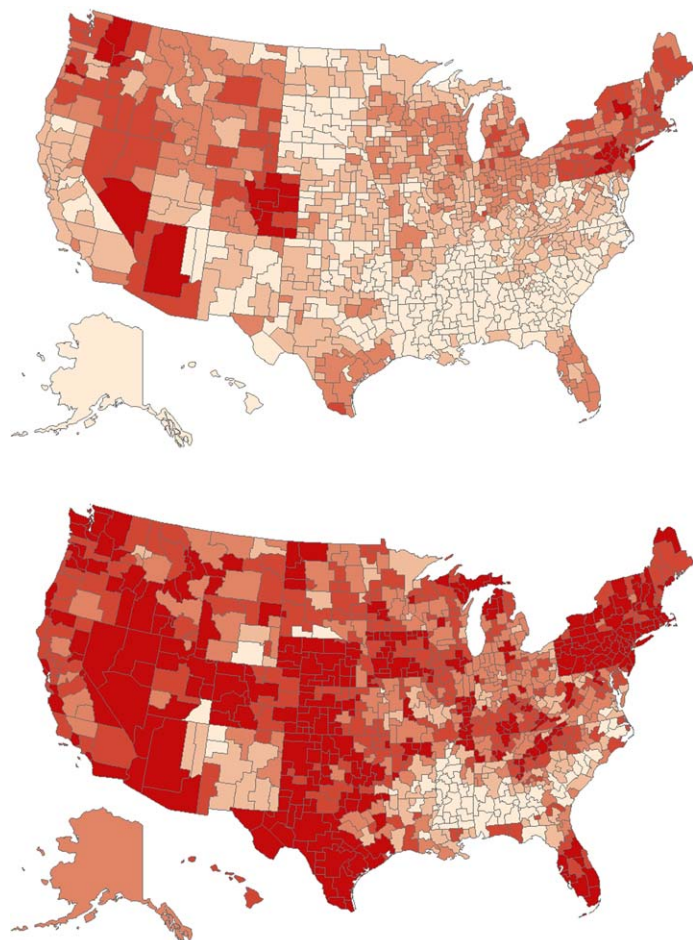


Figure 3. Age-adjusted Incidence Rates, Ages 0–19 of All Childhood Cancers at HSA Level, 1995–2006 (Top: Female; Bottom: Male; From lightest to darkest: 10.0–14.5, 14.6–15.5, 15.6–16.5, 16.6–17.5, 17.6–22.2 per 100,000).

soft tissue including heart, acute lymphocytic leukemia, and acute myeloid leukemia.

Table 2 shows the number of projected childhood cancer new cases for selected sites for all 50 States, the District of Columbia and the total U.S. in 2010. Not all states shown in the table provided data in this study, but through sharing information across areas that are similar in demographic and lifestyle profiles, as well as age, race, mortality rate, diagnosis year, and spatial and temporal random effects, the statistical models provided estimates for states with and without data. For states with relatively small populations, the estimated new cases of certain cancer sites less than 4 are denoted with \* to indicate unstable estimates. The following sections will focus on the spatial and temporal patterns at different geographic and time scales. The maps in Figures 3 through 5 show the delay-adjusted age-adjusted incidence rates of all childhood cancer, and the plots in Figures 6 and 7 show all and selected childhood cancer sites, respectively.

Table 2. Projected New Childhood Cancer Cases (age 0–19) for Selected Cancer Sites by State, U.S., 2010 (\* denotes 3 or fewer new cases.)

State	All Sites		Bone & Joint	Brain & Other Nervous	Kidney & Renal Pelvis	Leukemia	Lymphoma	Soft Tissue Including Heart
	Female	Male						
Alabama	95	100	10	35	9	46	27	12
Alaska	15	19	*	6	*	9	5	*
Arizona	180	215	25	64	12	89	49	17
Arkansas	62	68	7	25	6	32	19	9
California	886	1041	103	286	60	549	271	117
Colorado	126	141	16	42	10	59	39	14
Connecticut	77	102	9	31	6	42	33	10
Delaware	21	23	*	8	*	9	7	*
District Of Columbia	11	9	*	4	*	4	4	*
Florida	410	486	50	149	34	211	121	48
Georgia	221	261	28	81	23	122	63	33
Hawaii	21	29	*	9	*	13	5	*
Idaho	40	46	6	17	4	19	11	4
Illinois	297	362	34	99	30	161	88	33
Indiana	151	168	14	52	16	80	44	18
Iowa	64	86	9	26	5	34	19	8
Kansas	56	69	9	26	5	33	18	6
Kentucky	91	108	10	33	8	60	24	12
Louisiana	73	94	8	31	8	43	29	10
Maine	28	32	*	9	*	13	12	*
Maryland	120	149	15	37	11	61	40	18
Massachusetts	139	159	16	51	11	69	56	16
Michigan	234	257	22	76	22	119	72	25
Minnesota	115	125	16	46	11	59	40	16
Mississippi	58	66	6	23	7	29	16	8
Missouri	125	148	20	51	11	64	39	13
Montana	21	23	*	9	*	9	7	*
Nebraska	40	51	6	19	4	21	11	5
Nevada	68	81	9	26	5	33	18	7
New Hampshire	32	36	*	10	*	15	13	4
New Jersey	213	257	24	81	18	108	77	27
New Mexico	44	48	6	17	4	25	14	6
New York	439	517	56	173	35	217	161	45
North Carolina	194	232	25	73	19	104	63	24
North Dakota	12	19	*	5	*	10	4	*
Ohio	255	275	25	89	25	124	79	29
Oklahoma	76	123	9	31	12	59	26	12
Oregon	83	94	8	31	6	45	28	9
Pennsylvania	304	339	34	109	24	130	109	30
Rhode Island	26	24	*	6	*	12	9	*
South Carolina	91	99	11	35	9	45	26	11
South Dakota	18	20	*	7	*	9	5	*
Tennessee	129	149	16	48	12	66	37	18
Texas	618	799	69	242	55	407	172	79
Utah	79	103	9	33	7	41	22	11
Vermont	12	16	*	5	*	6	6	*
Virginia	171	201	20	61	14	87	56	23
Washington	142	165	14	56	12	78	46	19
West Virginia	33	37	4	14	*	16	10	4
Wisconsin	122	135	11	41	12	65	37	21
Wyoming	12	13	*	5	*	5	4	*
U.S. Total	6948	8220	818	2538	603	3769	2193	857

### 3.2 Spatial distribution pattern

Figure 3 shows the female and male age-adjusted incidence rate of all childhood cancer at the HSA level for the entire period 1995–2006. The two maps use the same cut-points, the 6-quantiles of the pooled male and female rates, to compare the rates between female and male patterns. The age-adjusted incidence rates range from 10.0 to 22.2 per 100,000 population at the HSA level with male rates higher than female across all areas in the U.S. High-rate areas for females are mainly in the Northeast U.S., the Mountain States and the North Pacific Region (mainly the state of Washington). High-rate areas for males include the above mentioned regions as well as Florida and the West Central U.S., spreading from Nebraska and Iowa through the western part of Kansas and Oklahoma to the western and southern part of Texas. It is worth noting though that these maps reveal distribution patterns at the HSA level for the whole study period, some of the rates are based on small numbers and so these patterns need to be interpreted with caution.

To investigate the variation in statistical stability of rates, the coefficient of variation (CV) is calculated and mapped for all HSAs. CV is defined as the ratio of standard deviation of the estimated age-adjusted incidence rate to the estimated age-adjusted incidence rate at each HSA. CV is a measure of stability of a statistical estimate, with larger values indicating less stable estimates. The National Vital Statistics Report has used fewer than 20 deaths (equivalent to a CV of 22.4% or more) as a benchmark to indicate unreliable death rates, with rates based on fewer cases considered to be statistically unreliable [23]. Since childhood cancer is a very rare disease, we relaxed the criterion to fewer than 10 new childhood cancer cases (equivalent to a CV of 31.6% or more) to indicate the very unstable estimates in this study. Figure 4 shows the CV in all HSAs for female and male age-adjusted incidence rates. The darkest areas are HSAs with the most unstable estimates (or less than 10 new cases), and they are mainly in the less populated West and Midwest Regions.

### 3.3 Temporal trend

While HSA level maps reveal small area spatial distribution patterns, the HSA level data are too sparse to examine temporal trends (with many more unstable estimates than shown in Figure 4). As a result, we aggregated the number of HSA-level new cases to the State level and mapped the State level age-adjusted incidence rate by year. Shown in Figure 5 are State level female and male age-adjusted incidence rates for the beginning year 1995, year 2000, the last year 2006 when observed data are available, and the 4-year projection to year 2010. The same cutpoints, the 6-quantiles of the pooled male and female rates, are used here to compare the rates between females and males. The age-adjusted childhood all cancer incidence rates range from 10.9 to 20.9

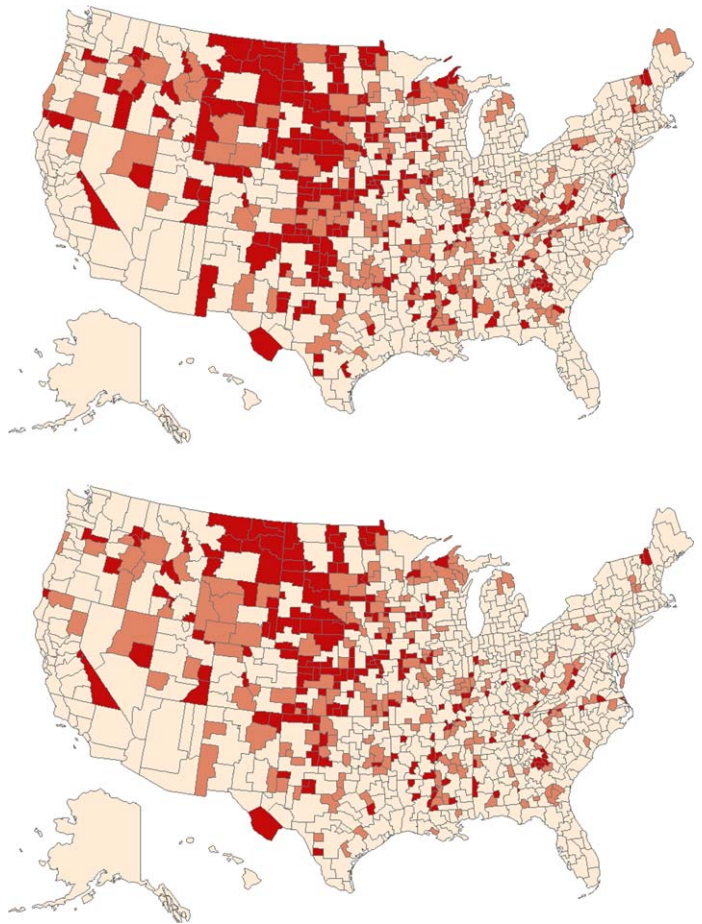


Figure 4. Coefficient of Variation on Age-adjusted Incidence Rate of Childhood All Cancer at HSA Level (Top: Female; Bottom: Male; From lightest to darkest: 1.8%–22.4%, 22.5–31.6%, 31.7%–300%).

per 100,000 population. It can be seen from the maps that male rates are always higher than female rates, and that the rates have been increasing over the past 15 years for both males and females. These maps show the same high-rate areas as the HSA maps (Figure 3) in the Northeast U.S., the Mountain States, Pacific States as well as Florida and the West Central U.S. spreading from Iowa through Kansas and Oklahoma to Texas.

To illustrate the temporal changes of incidence rates in more detail, more plots are included to show the spatiotemporal model-estimated and joinpoint-predicted age-adjusted (Figure 6) and age-specific (Figure 7) incidence rates for selected childhood cancer sites. The horizontal axis in each plot is time from 1995 to 2010. For the first 12 years (1995–2006), the predicted rates (solid line) are plotted along with the estimated rates (symbols around the solid line). The period from 2007 to 2010 does not have model-estimated points, and the solid line is extended to 2010 from the last line segment in the Joinpoint result. Each solid line is labeled

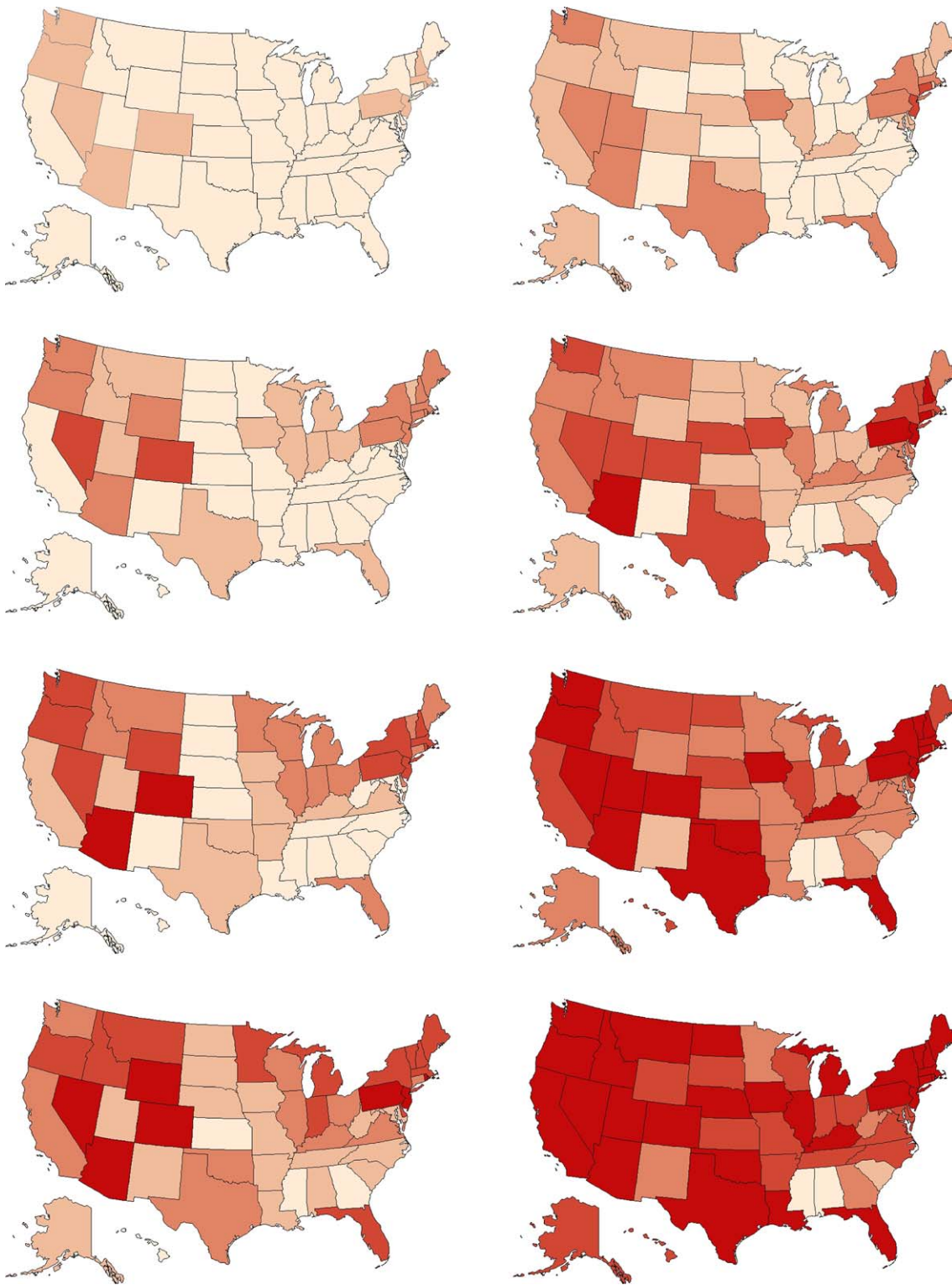


Figure 5. Predicted Age-adjusted Incidence Rates of All Childhood Cancers at State Level (Left: Female, Right: Male; Top to Bottom: 1995, 2000, 2006, 2010; From lightest to darkest: 10.9–15.5, 15.6–16.5, 16.6–17.5, 17.6–18.5, 18.6–20.9 per 100,000).



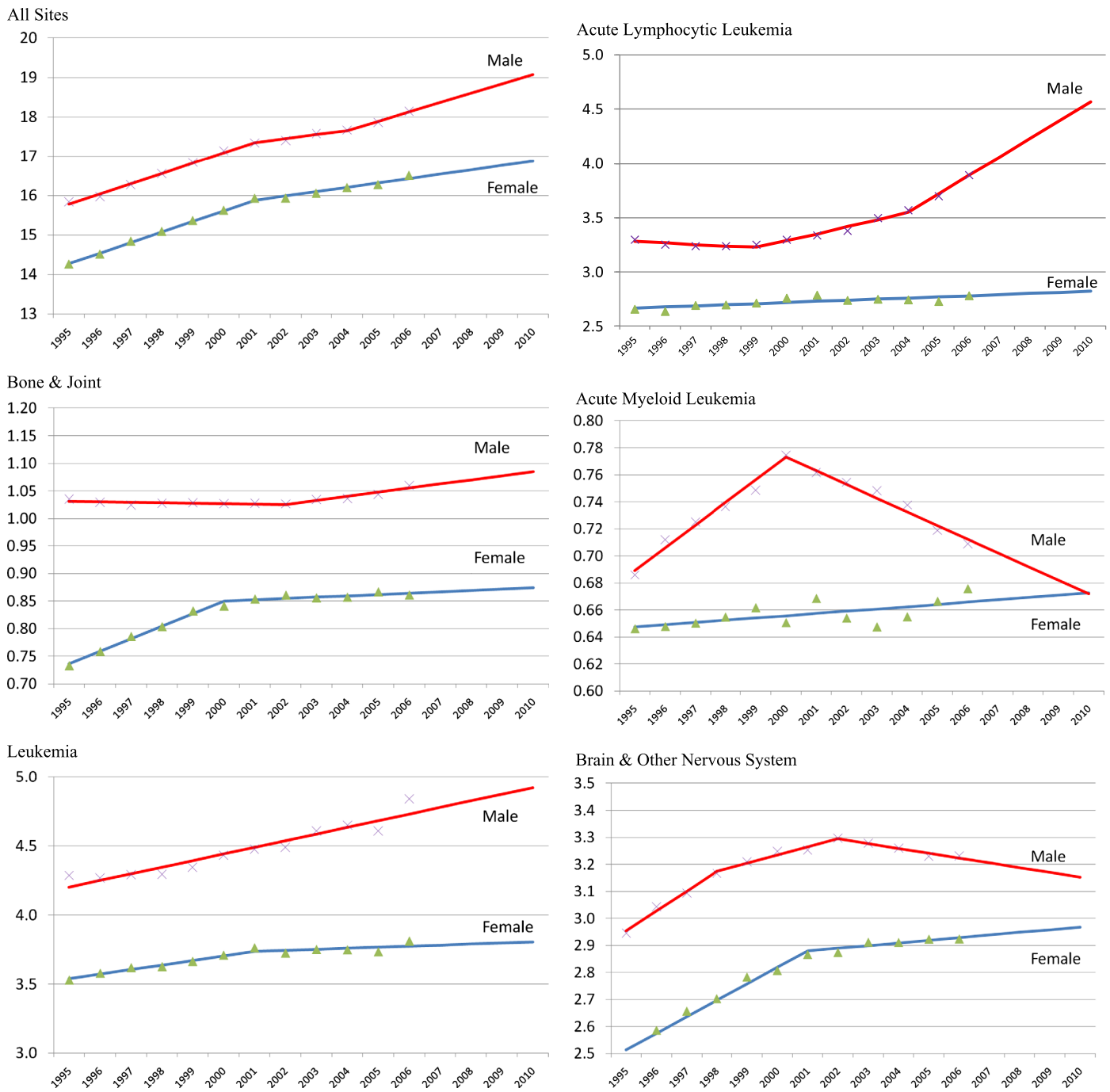


Figure 6. Joinpoint-predicted (solid lines) and Spatiotemporal Model Estimated (symbols around the lines) Age-adjusted Incidence Rates per 100,000 for Selected Childhood Cancers by Sex, 1995–2010.

with gender (Figure 6) or age group (Figure 7). Presented here are the age-adjusted or age-specific rates only, without consideration of the variations or the confidence intervals of the rates. For some groups with rates close to each other, it is expected to see overlapping of the confidence intervals.

Figure 6 shows an overall increase and relatively parallel female and male lines for all the childhood cancer sites

combined, cancer of bone & joint, leukemia, cancer of brain & other nervous system and Non-Hodgkin lymphoma. The female age-adjusted incidence rate for acute lymphocytic leukemia is very stable during the study period, but the male incidence rate has experienced a significant increase starting 2003. For acute myeloid leukemia, the female rate has also been stable while the male rate increased from 1995 to 2000

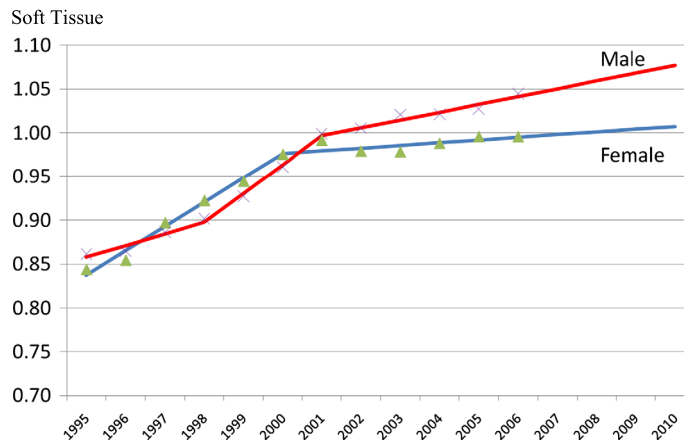
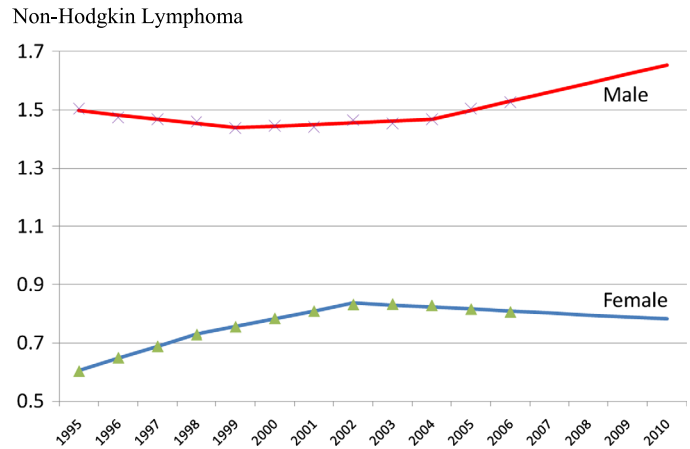
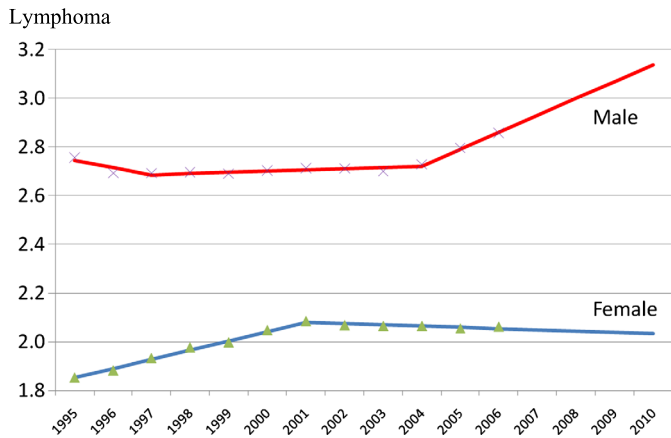
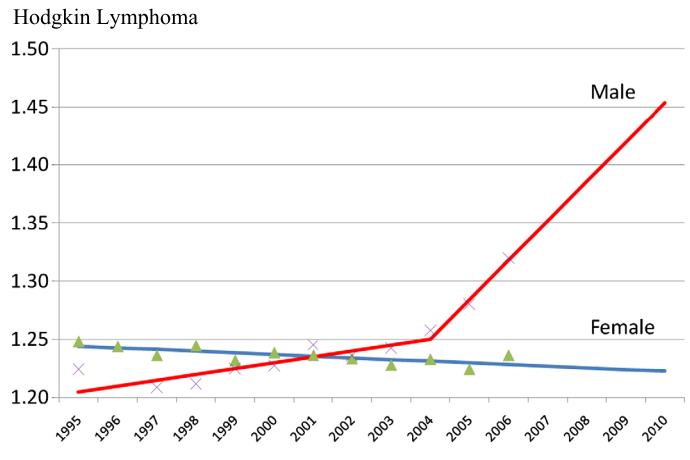
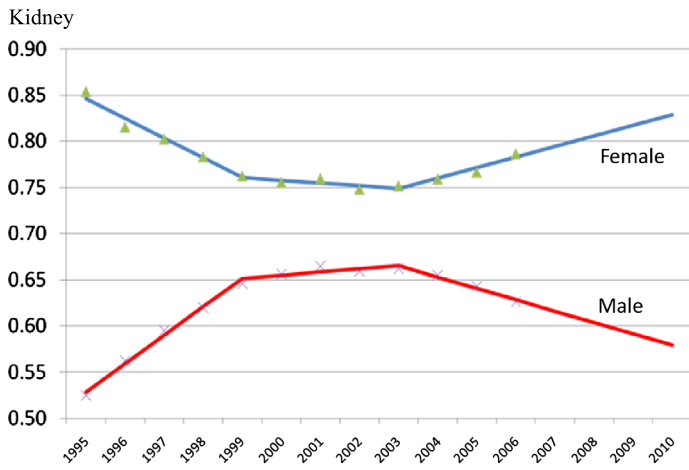


Figure 6. (Continued.)

and started dropping to the same level as the female rate in 2010. Unlike all other cancer sites, kidney cancer is unique in that female rates are higher than the corresponding male rates. Before 1999, the male kidney cancer rate increased and the female kidney cancer rate decreased; between 1999 and 2003, the rates of both genders were quite stable; then starting 2003, rates have been diverging so that the female rates have increased and male rates have decreased.

Lymphoma and Hodgkin lymphoma have similar trends in that the incidence rates were stable for female across the study period, but male incidence rates have started increasing from 2004. The incidence rates for cancer of soft tissue including heart were increasing and very similar for males and females in the earlier period until 2000–2001. After 2001 the rate of increase has slowed down, more so for females, resulting in a higher projected rate for males in 2010.

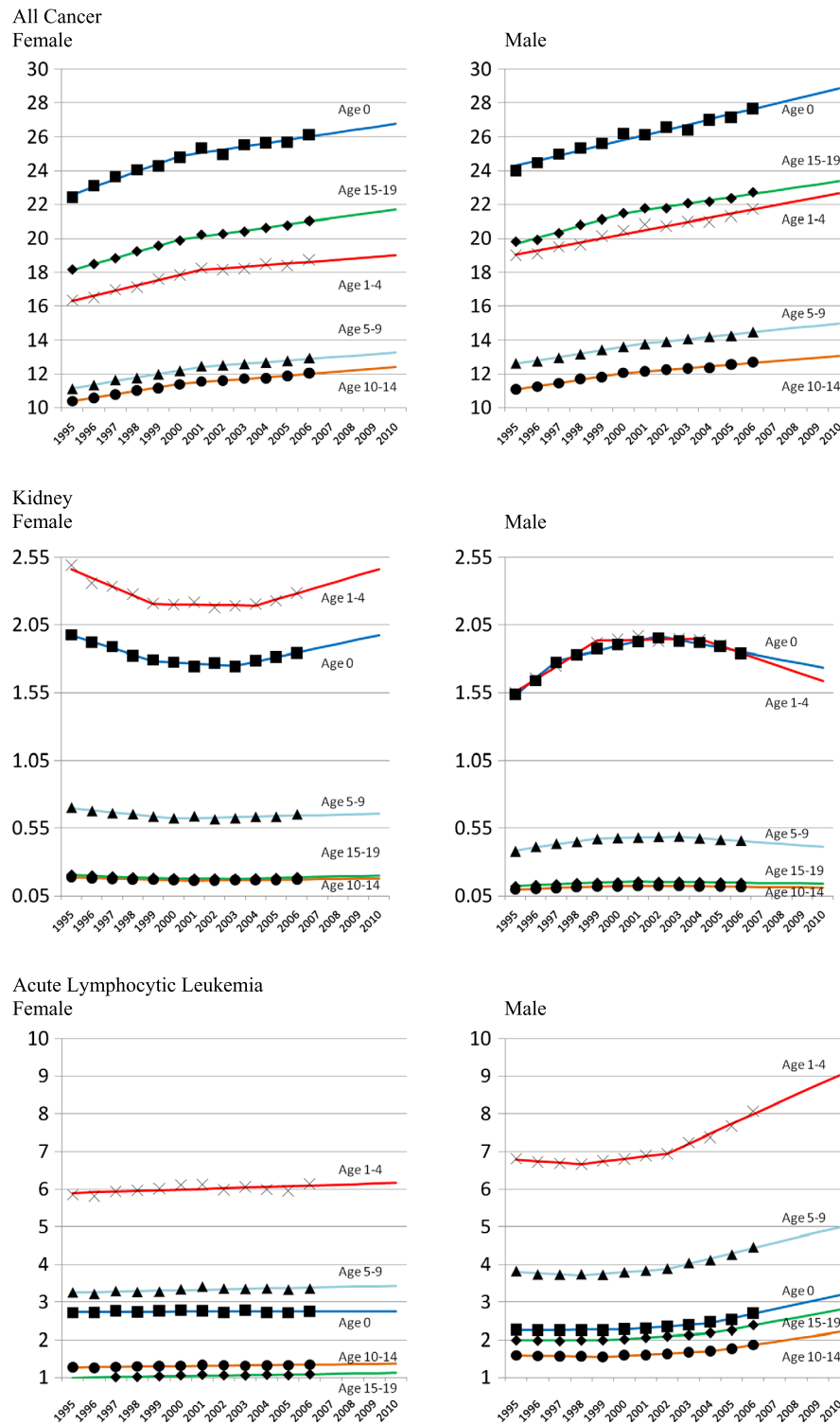
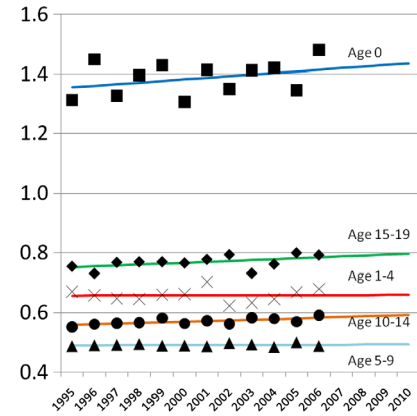


Figure 7. Joinpoint-predicted (solid lines) and Spatiotemporal Model Estimated (symbols around the lines) Age-specific Incidence Rates per 100,000 for Selected Childhood Cancers by Sex and Age Group, 1995–2010.

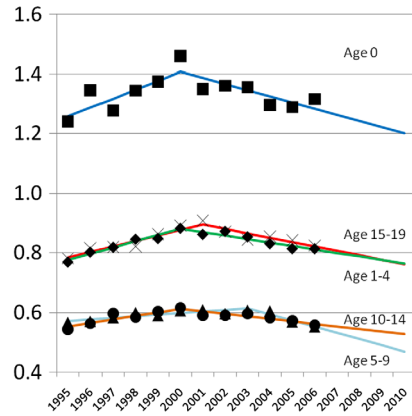
Figure 7 compares the temporal trend of age-specific incidence rates for selected childhood cancer sites between males and females and the 5 age groups. For all cancer sites combined, the youngest (age <1 year) and the oldest (age

15–19) age groups have the highest rates, followed by the 1–4, 5–9, and the 10–14 age groups in descending order. As noted before, female kidney cancer rates are higher than male rates, and that is true across all age groups. For the

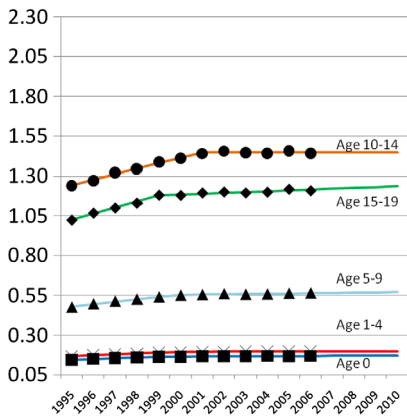
Acute Myeloid Leukemia  
Female



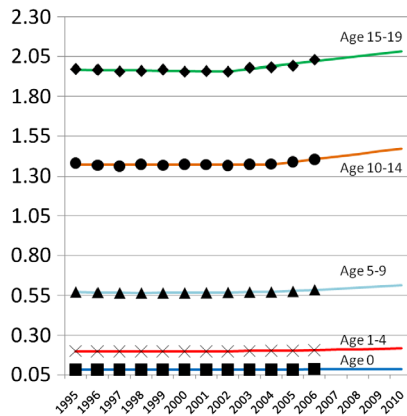
Male



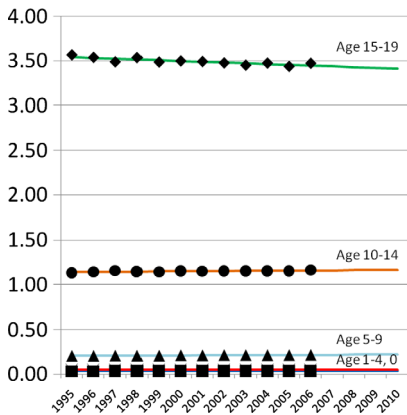
Bone & Joint  
Female



Male



Hodgkin Lymphoma  
Female



Male

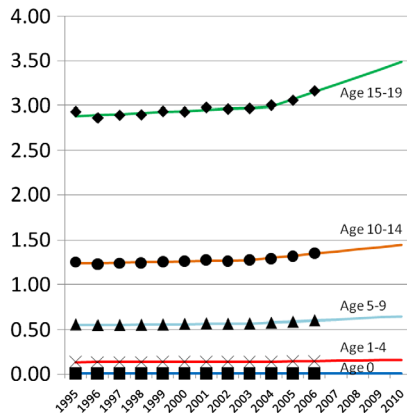


Figure 7. (Continued.)

two age groups (<1 and 1–4) with the highest rates, females and males have different temporal change patterns – female rates decreased first and then moved up, but the male rate increased first then started going down. The age

group with the highest rate for acute lymphocytic leukemia is ages 1–4. Across all the age groups, female acute lymphocytic leukemia rates have been stable and the male rates started increasing from 2002, the same trend as noted in

the age-adjusted rate. Acute myeloid leukemia also shows the same trend as the age-adjusted rate, i.e., that female rates have been stable but the male rates increased first and then dropped beginning in 2000. The bone & joint cancer rates have been stable during the study period. The age groups with the highest rates are the 15–19 (male) and the 10–14 (female) groups, with a much higher rate in the male 15–19 group. The rates for Hodgkin lymphoma have stayed stable for females during the study period and for most male groups. For the male age 15–19 group the rates started increasing in 2004. It is notable that the rate of the age 15–19 group is more than double in males and triple in females the rates of the age 10–14 group, the group with the second highest rate.

#### 4. DISCUSSION

This paper presents the first study in the U.S. that links childhood cancer incidence with mortality as well as socioeconomic status measures, behavioral risk factors, medical resource measures, geographic location and time through a sophisticated statistical model. This new method is an improvement over the old method used for cancer incidence projection. The old method projects the total number of case counts to the current year by applying the annual age-specific rates in 9 oldest SEER areas to the most current year. The new method takes into account potential covariates and spatial and temporal variation of cancer occurrence, as well as delay in reporting. Estimates produced in this new method are in close proximity to the cancer incidence estimate for age 0–14 as reported in CFF. Taking a further step than the CFF, this paper describes the spatiotemporal patterns and trends by age groups, sex, and major childhood cancer sites. The results found in this study confirm the slightly rising trend in all childhood cancer sites combined [24] and the variation in incidence with different age groups and gender [25]. Previous efforts have been made to investigate the geographic or temporal patterns of childhood cancer mortality [26–28] and incidence [29–31]. Presented here is a unique study that combined data from the National Program of Cancer Registries and the SEER cancer registries at fine resolutions in both geographic and temporal scales. This is the first effort to investigate U.S. childhood cancer incidence rates for small area geographic patterns and temporal trends, and projection ahead in time. The research findings fill in the data gaps in the cancer registry reporting to provide a complete coverage of the whole U.S., and provide a scientific basis to allocate resources in cancer prevention and control efforts at the national, state, and local level. The rarity of childhood cancer might lead to unstable estimates due to small case counts. This paper quantified and presented the instability of small area estimates along with the estimates themselves so that cautions may be taken when interpreting these estimates.

One limitation of this study is that death certificates often may not indicate the specific subsites of childhood cancers such as acute lymphocytic leukemia or acute myeloid leukemia. Therefore the mortality rates included as predictors in the regression model may be somewhat inaccurate. One way to bypass this problem could be to use mortality from the broader group leukemia to predict the incidence in the subsites of leukemia. A detailed inspection into the model results finds that results are consistent across the broader group and the subtypes of leukemia, hence the mortality data are kept in the current format as they are in producing the CFF incidence estimates. Questions may also arise for the accuracy of the model estimates. The original method paper by Pickle et al. [2] and related research (unpublished work) by the authors compares the estimated and the observed number of cases for adult cancers and found that they are in close proximity to each other. Validation in certain SEER registries also shows that the predicted case counts are close to the observed counts for childhood cancer. Another limitation is that the incidence data do not cover all the states in the U.S., since not all registries that met NAACCR registry certification standards grant permissions for us to use their county-level incidence data.

*Received 11 January 2013*

#### REFERENCES

- [1] WINGO, P., et al. (1998). Using cancer registry and vital statistics data to estimate the number of new cancer cases and deaths in the United States for the upcoming year. *Journal of Registry Management* **25**, 43–51.
- [2] PICKLE, L. W., et al. (2007). A new method of estimating United States and state-level cancer incidence counts for the current calendar year. *CA Cancer J Clin* **57**(1), 30–42.
- [3] National Cancer Institute. *SEER 9 Delay Model*. [cited 2011 08/25/2011]; Available from: <http://surveillance.cancer.gov/delay/model.html>.
- [4] American Cancer Society *Cancer Facts & Figures 2007*. 2007.
- [5] National Cancer Institute. *Surveillance Epidemiology and End Results (SEER) Program*. 2011 [cited 2010 January 4]; Available from: <http://seer.cancer.gov/>.
- [6] Centers for Disease Control and Prevention. *National Program of Cancer Registries (NPCR)*. 2010 [cited 2010 January 4]; Available from: <http://www.cdc.gov/cancer/npcr/>.
- [7] MAKUC, D. M., et al. (1991). Health service areas for the United States. *Vital Health Stat* **2**(112), 1–102.
- [8] Surveillance Epidemiology and End Results (SEER) Program, N.C.I. *Health Service Areas (HSA)*. 2010 [cited 2010 January 4]; Available from: <http://seer.cancer.gov/seerstat/variables/countyattribs/hsa.html>.
- [9] Surveillance Epidemiology and End Results (SEER) Program, N.C.I. *SEER incidence site recode*. [cited 2010 January 4]; Available from: <http://seer.cancer.gov/siterecode/index.html>.
- [10] HOWLADER, N., et al. (2011). *SEER Cancer Statistics Review, 1975–2008*.
- [11] Surveillance Epidemiology and End Results (SEER) Program, N.C.I. *Population Estimates Used in NCI'S SEER\*Stat Software*. 2010 [cited 2010 January 4]; Available from: <http://seer.cancer.gov/popdata/methods.html>.

- [12] Health Resources and Services Administration, *Area Resource File (ARF), National County-level Health Resource Information Database*, 2011.
- [13] Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System*. [cited 2011 August 25]; Available from: [http://www.cdc.gov/brfss/technical\\_infodata/index.htm](http://www.cdc.gov/brfss/technical_infodata/index.htm).
- [14] Centers for Disease Control and Prevention. *Mortality Data*. [cited 2011 August 25]; Available from: <http://www.cdc.gov/nchs/deaths.htm>.
- [15] U.S. Cancer Statistics Working Group *United States Cancer Statistics: 1999–2007 Incidence and Mortality Web-based Report*. 2010.
- [16] SAS Institute. *SAS/STAT GLIMMIX Procedure, June 2006*. 2006 [cited 2010 January 20]; Available from: <http://support.sas.com/rnd/app/papers/glimmix.pdf>.
- [17] PICKLE, L., FEUER, E. J. and EDWARDS, B. K. (2003). *U.S. Predicted Cancer Incidence, 1999: Complete maps by county and state from spatial projection models*, National Cancer Institute: Bethesda, MD.
- [18] Surveillance Epidemiology and End Results (SEER) Program, N.C.I. *Standard Populations (Millions) for Age-Adjustment*. [cited 2011 August 25]; Available from: <http://seer.cancer.gov/stdpopulations/index.html>.
- [19] KIM, H. J., et al. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med* **19**(3), 335–351.
- [20] National Cancer Institute. *Joinpoint Regression Program*. [cited 2011 08/25/2011]; Available from: <http://surveillance.cancer.gov/joinpoint/>.
- [21] KOHLER, B. A., et al. (2011). Annual report to the nation on the status of cancer, 1975–2007, featuring tumors of the brain and other nervous system. *J Natl Cancer Inst* **103**(9), 714–736.
- [22] American Cancer Society, *Cancer Facts & Figures 2010*. 2010.
- [23] ARIAS, E., et al. (2003). Deaths: Final data for 2001. *Natl Vital Stat Rep* **52**(3), 1–115.
- [24] ROSS, J. and SPECTOR, L. (2006). *Cancers in Children*, in *Cancer Epidemiology and Prevention (3rd Edition)*, D. Schottenfeld and J.J. Fraumeni, eds., Oxford University Press: New York, NY.
- [25] RIES, L., et al. (1999). *Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975–1995*.
- [26] BOSETTI, C., et al. (2010). Childhood cancer mortality in Europe, 1970–2007. *Eur J Cancer* **46**(2), 384–394.
- [27] Centers for Disease Control and Prevention (2007). *Trends in Childhood Cancer Mortality – United States, 1990–2004*. MMWR Morbidity and Mortality Weekly Report, **56**(48), 1257–1261.
- [28] KVASNICKA, J., et al. (1993). Temporal and regional trends of cancer mortality in West Germany. *Radiat Environ Biophys* **32**(2), 163–182.
- [29] THOMPSON, J. A., CAROZZA S. E., and ZHU L. (2008). Geographic risk modeling of childhood cancer relative to county-level crops, hazardous air pollutants and population density characteristics in Texas. *Environ Health* **7**, 45.
- [30] LI, J., et al., (2008). Cancer incidence among children and adolescents in the United States, 2001–2003. *Pediatrics* **121**(6), e1470–7.
- [31] HJALMARS, U., et al. (1999). Increased incidence rates but no space-time clustering of childhood astrocytoma in Sweden, 1973–1992: A population-based study of pediatric brain tumors. *Cancer* **85**(9), 2077–2090.

Li Zhu  
 Surveillance Research Program  
 Division of Cancer Control and Population Sciences  
 National Cancer Institute  
 National Institutes of Health  
 Bethesda, Maryland 20892-8317  
 United States  
 E-mail address: [li.zhu@nih.gov](mailto:li.zhu@nih.gov)

Linda W. Pickle  
 StatNet Consulting LLC  
 20203 Goshen Road, No. 189, Gaithersburg  
 Maryland 20879  
 United States  
 E-mail address: [linda@statnetconsulting.com](mailto:linda@statnetconsulting.com)

Zhaohui Zou  
 Information Management Services, Inc  
 12501 Prosperity Drive, Suite 200  
 Silver Spring, Maryland 20904  
 United States  
 E-mail address: [zouj@imsweb.com](mailto:zouj@imsweb.com)

James Cucinelli  
 Information Management Services, Inc  
 12501 Prosperity Drive, Suite 200  
 Silver Spring, Maryland 20904  
 United States  
 E-mail address: [cucinellij@imsweb.com](mailto:cucinellij@imsweb.com)