Trends in Colorectal Cancer Incidence and Survival in Iowa SEER Data: The Timing of It All

Rachel Carroll, Shanshan Zhao

Abstract

Colorectal cancer is common worldwide. Our results indicated that both overall and older onset colorectal cancer incidence began to decline in the early 2000s, whereas younger onset incidences decreased until the late 1980s but then increased steeply through the 2000s. These and other results suggested that regular colorectal screenings could reduce incidence and mortality in people under 50.

Background: Colorectal cancer (CRC) is common worldwide, with 140,250 diagnoses and 50,630 deaths estimated for the United States in 2018. Guidelines current to the most recent individuals in our analysis suggested regular screenings beginning at age 50 have reduced the incidence of CRC. However, the incidence continues to rise among those under 50. Less is known about survival following CRC diagnosis, but research has suggested that younger cases may also have worse survival. However, we hypothesize that younger individuals are generally healthier with fewer comorbidities, leading to the potential for better survival following diagnosis. **Materials and Methods:** We utilized the Surveillance, Epidemiology, and End Results data to estimate and assess both spatial and temporal variation in age-specific colorectal cancer incidence and survival in Iowa. **Results:** Both overall and older-onset colorectal cancer incidence began to decline in the early 2000s, whereas younger-onset incidences decreased until the late 1980s but then increased steeply through the 2000s. The risk for those younger than 50 years of age first exceeded the risk for those 50 years or older in 2007. Survival times did increase for overall CRC, older-onset CRC, and young-onset CRC throughout the study period, with young-onset CRC increasing at a higher rate. The spatial variation assessment indicated that the survival was positively associated with several variables of interest, most notably disparities including better access to healthcare and higher sociodemographic status. **Conclusion:** In conclusion, results suggest that regular colorectal screenings could reduce incidence and mortality in people under 50.

Clinical Colorectal Cancer, Vol. 18, No. 2, e261-74
Keywords: Colorectal cancer, Incidence, Risk, Spatial and temporal modeling, Survival

Introduction

Colorectal cancer (CRC) is one of the most diagnosed cancers in the United States. The American Cancer Society estimated that 140,250 individuals will be diagnosed with CRC in 2018.[1] Strong recommendations were made by the American Cancer Society in 1996, suggesting that all adults over the age of 50 should submit to regular screenings for CRC every 10 years. This change has resulted in an overall decline in CRC incidence.[2] However, recent studies have suggested an increase in incidence remains for individuals less than 50 years of age.[2-5] Specifically, the American Cancer Society stated that incidence rates for CRC among individuals older than 50 years declined by 3.8% annually from 2005 to 2010, whereas CRC rates for individuals younger than 50 years increased by 1.4% over the same time frame.[1] For these analyses, we consider CRC diagnosis prior to the age of 50 as young-onset CRC and CRC diagnosis after the age of 50 to be older-onset CRC.

Survival following CRC diagnosis in different age groups has not been explored as thoroughly, although studies suggest that treatment improvements lead to better survival overall.[6,7] The American Cancer Society estimated that, in 2018, 50,630 individuals' underlying cause of death will be CRC but did not offer age-specific mortality information.[1] Further, age—group-specific survival following CRC diagnosis has not been assessed much with respect to changes over calendar time.

In this article, we utilized information related to CRC incidence and survival following diagnosis recorded in the Iowa Surveillance Epidemiology and End Results (SEER) registry between the years

National Institute of Environmental Health Sciences, Durham, NC

Submitted: Sep 19, 2018; Revised: Dec 1, 2018; Accepted: Dec 6, 2018; Epub: Dec 22, 2018

Address for correspondence: Rachel Carroll, PhD, National Institute of Environmental Health Sciences, DPHS, 111 TW Alexander Dr, Durham, NC 27702
E-mail contact: rachel.carroll@nih.gov

1533-0028/$ - see frontmatter
https://doi.org/10.1016/j.clcc.2018.12.001

1973 and 2013, with the goal to understand the differences to younger and older individuals separately. To assess incidence of CRC, we employed a Bayesian Poisson Knorr-Held model,[8,9] which produced estimates for the temporal trend after accounting for several aggregated, county-level risk factors as well as the underlying spatial and spatio-temporal variation. Similarly, temporal trends in survival following CRC diagnosis were assessed after adjusting for individual-level risk factors as well as the underlying spatial and spatio-temporal variation via a spatio-temporal Bayesian accelerated failure time (AFT) model.[10-12]

Previous work has demonstrated improvements in the modeling of rare diseases by borrowing information from related, more common diseases.[13,14] Although CRC as a whole is not a rare disease, young-onset CRC is. And, because young- and older-onset CRC do have similarities in terms of risk factors, we believed that a multivariate approach with shared random effects could be adopted for the purpose of this analysis, as we were interested in quantifying and comparing the temporal trends of incidence and survival for those under age 50 versus those over age 50.

Materials and Methods

Iowa SEER Data

The SEER program offers publicly available cancer data from 17 registries across the United States (US). The specific registry that provides CRC data for Iowa currently includes individuals diagnosed between January 1, 1973 and December 31, 2013. The Iowa registry was selected for the long follow-up, which is not available for all SEER registries. To model the CRC incidence, a Poisson model was used, based on the county-level incidence calculated as a count of the number of individuals diagnosed in each county for a given year. The county population across the study years was also accounted for as an offset in the Poisson model. Because Iowa has 99 counties and 41 years were considered, the count outcome was a matrix with 99 rows and 41 columns. To model survival following CRC diagnosis, we combined the documented individual survival time in months together with a CRC mortality underlying cause of death indicator (International Classification of Diseases, Tenth Revision: C18, C26.0, and C19-C20). Thus, an individual was considered censored either at their last follow-up time, at the end of the study period (December 31, 2013), or at the time of their death owing to other causes. The final number of individuals diagnosed with CRC and considered for analysis was reduced from 83,664 to 82,828 based on requiring known survival time. This removed only 1% (n = 836) from the total analysis, 0.4% (n = 18) from young-onset and 1% (n = 818) from older-onset CRC.

To determine an individual's location, we utilized the SEER registry-provided Federal Information Processing Standard county code. With this information, we assigned an individual to 1 of the 99 Iowa counties. On average, there were 20.6 (range, 0-227) incidences per county per year. There were 4 counties with no CRC diagnosed ever across the study time; all these were smaller counties, as they were consistently in the twelfth or less percentile of counties for year-specific population size. When reducing the outcome to either above or below the age of 50, the average number of incidences within a county per year were 19.5 (range, 0-210) and 1.1 (range, 0-30), respectively. There were 4 county-year combinations without older onsets, and 2014 without young onsets.

The demographic and clinical covariates considered in the analysis of CRC survival were selected for their known associations with CRC mortality[15] and their availability in the SEER database. There were some missing in the covariates considered in the survival model. However, it was important to include this information so that we could maximize the number of individuals contributing within each county and year for the spatial and temporal components, particularly for the young-onset age group. Thus, we incorporated unknown categories when needed. Specifically, the covariates included were: race (White vs. other), marital status at diagnosis (single, currently married, separate/divorced/widowed/unknown), age at diagnosis, high cancer grade (yes, no/unknown), previous history of reportable malignant, in situ, benign, and borderline primary tumors (no, yes), CRC surgery (no/unknown, yes), and radiation therapy (no/unknown, yes). For marital status, the large majority (80%; n = 23,416) of the third category was widowed individuals. Cancer grade had the highest amount of missing (23%; n = 19,216) and the next highest was marital status (2%; n = 1622). We considered grade as our best available measure to account for severity of cancer among age groups. Other potential variables for this were: stage at diagnosis, anatomic location, or histologic behavior; however, stage at diagnosis was missing for a large majority of individuals (84%; n = 70,286), a laterality variable offered no information about placement of the tumor, and a comparison of histologic behavior showed no difference between the age groups.

County-level Data

We used county-level data as a means to adjust for risk factors in the Poisson model, and these data were collected from the following sources: the Area Health Resources Files,[16] the University of Wisconsin Population Health Institute's County Health Rankings and Roadmaps,[17] and the Centers for Disease Control and Prevention.[18] The risk factors included for this purpose were chosen based on their association with increased risk of CRC,[15] and specifically, these were: percent white population, percent current smokers, percent of individuals with high education (4 year of college or more), median home value (in $10,000s), percent urban population, and county-level diabetes rate. Note that these risk factors were only available for a single year, and not all were from the same year; we made an attempt to account for such discrepancies within the statistical model.

Statistical Methods

Incidence Model. A Bayesian Poisson model was utilized in the assessment of temporal trends in CRC incidence. In this model, the mean of the Poisson distribution is defined as the expected rate times the relative risk ($e_{ij}\theta_{ij}$), and both are specified for each Iowa county ($i = 1,\ldots,99$) and available year ($j = 1,\ldots 41$ corresponding to the years 1973 to 2013). The expected rate is the offset term and is calculated as the year-specific county incidence divided by the county population for the same year.[19,20] Then, the fixed and random effects of interest are related to the relative risk through a log link: $\log(\theta_{ij}) = X_i\beta_{j^*} + w_i + \gamma_j + \phi_{ij}$ where X_i represents the county-level risk factors of interest and the β_{j^*} coefficients capture fixed effects of those factors that are specific to time intervals, with $j^* = 1,\ldots,8$ corresponding to every 5 years.[19,20] Also, w_i is the correlated spatial random effect, γ_j is the temporal random effect,

and φ_{ij} is the spatio-temporal interaction term. We used the 5-year timeframe for the fixed effect parameter (β_{j^*}) to minimize the identifiability issues with the temporal random effect and to have a more interpretable number of parameters.

A multivariate specification for this Poisson model was used to separately model individuals with young and older onset CRC. The parameterization of this model is as follows for $h = 1,2$ corresponding to the young and older onset incidences respectively:

$$y_{ijh} \sim Poisson\left(\mu_{ijh}\right)$$

$$\mu_{ijh} = e_{ijh}\theta_{ijh}$$

$$\log\left(\theta_{ijh}\right) = X_i\beta_{j^*h} + w_i + \gamma_{jh} + \phi_{ij}$$

Here, the spatial random effect and spatio-temporal interaction term are shared between the multivariate outcomes as a means to improve the fit for the rarer outcome, incidence of CRC for those younger than 50. Note that the offset terms are specific to the corresponding age-based subpopulations. This allows fixed effect parameter estimates and temporal random effect estimates to be specific to the age group, but the spatial and temporal effect to be the same for these 2 age groups, which is a reasonable assumption in practice.

Survival Model. To assess CRC survival time following diagnosis, we used the AFT model.[21] This model allows for a direct relationship of the logarithm of survival time with both the risk factors and the spatial, temporal, and spatio-temporal random effects.[22-24] This capability, along with the models' general flexibility in terms of assumptions. has led to the AFT model's increase in popularity and, for our purposes, an ideal interpretation of the spatial, temporal, and spatio-temporal frailty estimates. The AFT model for an individual k diagnosed in county i at time j can be written as: $\log(t_{ijk}) = \lambda_{ijk} + \sigma\varepsilon_{ijk}$ where t_{ijk} is the survival time, λ_{ijk} is the linear predictor of interest, ε_{ijk}'s are the independent random errors, and σ is a scale parameter. A single parameterization of λ_{ijk} is considered, and it is such that $\lambda_{ijk} = X_{ijk}\beta + w_i + \gamma_j + \phi_{ij}$. This parameterization is used in both univariate and multivariate settings where, similar to the Poisson model, a multivariate analysis is constrained to share random effects between the 2 age-specific case groups to gain age-group-specific temporal random effect estimates. The multivariate specification is as follows: $\lambda_{ijkh} = X_{ijk}\beta_h + w_i + \gamma_{jh} + \phi_{ij}$ where $h = 1,2$ for the 2 age groups considered. From here on, model results specific to the univariate, multivariate for younger onset, and the multivariate for older onset will be referred to as "all," "< 50," and "≥ 50."

Results

Incidence Assessment Results

Figure 1 displays maps of the county-level risk factors that were adjusted for in these Poisson models for assessing incidence, and Supplemental Table 1 (in the online version) shows the correlation matrix for these risk factors. Together, these illustrated that the distribution of these county-level variables was spatially structured because closer counties were more alike, and there were some similarities in the spatial distributions across variables because some correlations were fairly high. Specifically, the highest Pearson correlation is between median home value and percent high education ($\rho = 0.73$). The high correlation suggested that some of these variables could be representing similar characteristics.

Figure 2 displays the temporal random effect estimates for the "all" as well as both age subgroup incidence models. From this

Figure 1 Iowa County-level Risk Factors Adjusted for in the Poisson Models for Assessing Colorectal Cancer Incidence. Here, Lighter Shading Indicates Lower Values and Darker Shading Indicates Higher Values of the Given Variable

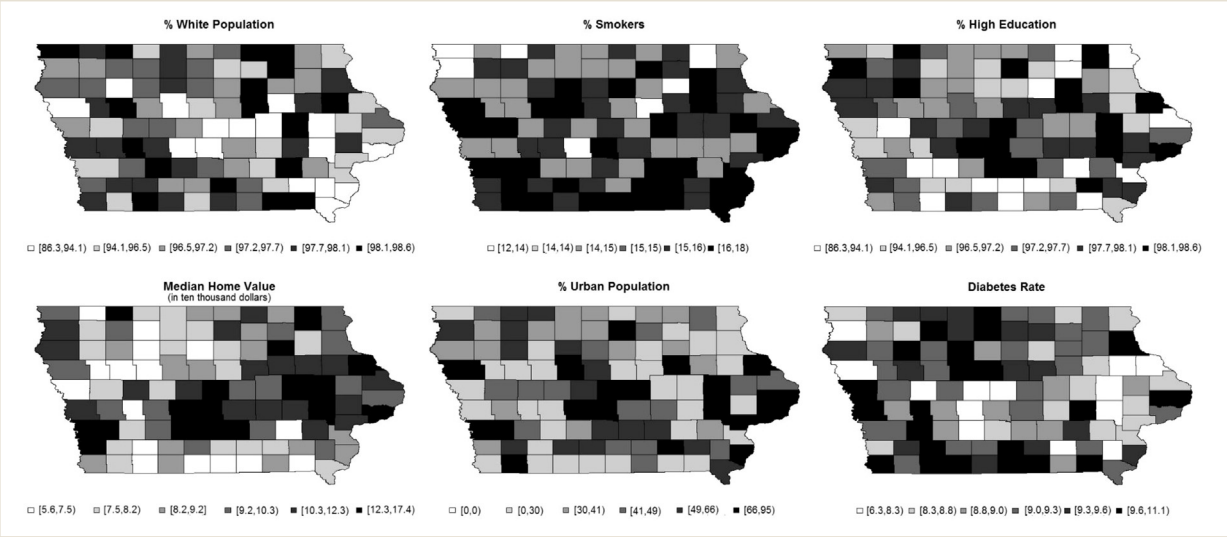
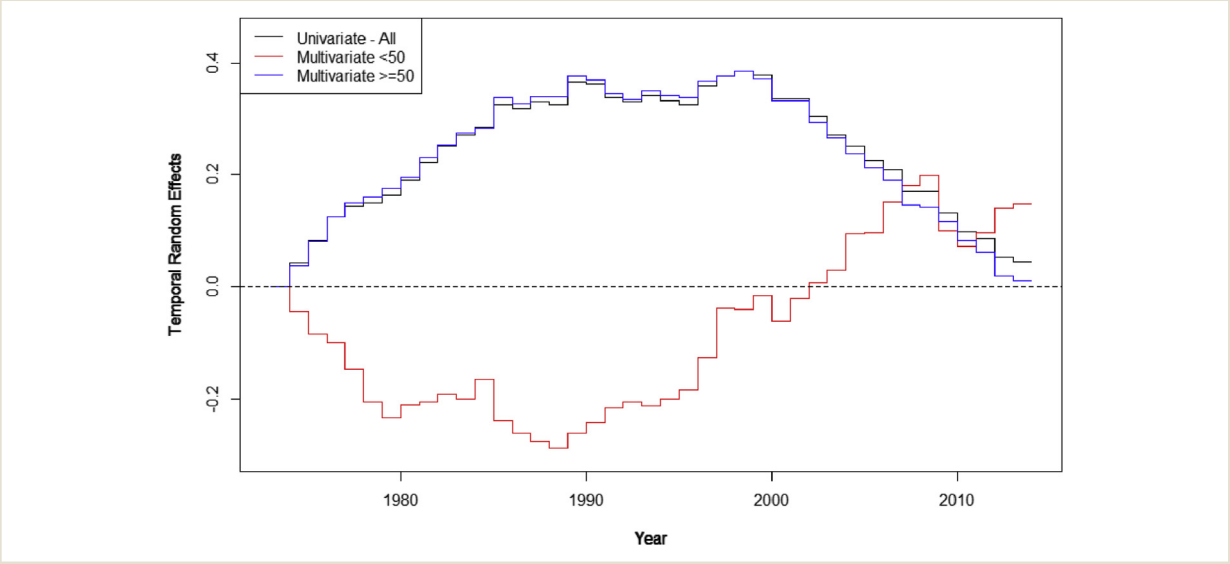


Figure 2 Temporal Random Effects Estimates From the Univariate and Multivariate Models for All Patients and Patients in Separate Age Groups, Examining Colorectal Cancer Incidence in Iowa Surveillance, Epidemiology, and End Results Data



figure, it is clear that the "all" model estimates (black line) were dominated by the ≥ 50 age group (blue line). In the multivariate setting, we were able to separate the temporal trends for the age groups of interest. From these, it was clear that the ≥ 50 model (blue line) suggested an increase in risk until about 1990, then risk was level throughout the 90s, and risk began to decline in the early 2000s. By contrast, the < 50 model (red line) decreased until the 1990s. Following that, the risk began increasing drastically and appeared to level off somewhat towards the end of the study time. Further, for those ≥ 50 years, the risk in 2013 was almost equal to the risk in 1973, whereas the risk for those < 50 years of age was 16% higher (exp (0.15) = 1.16). With this information, the estimates suggested that the risk for those < 50 years of age was first greater than the risk for those ≥ 50 years in 2007.

Supplemental Table 2 (in the online version) shows the relative risk for the different age groups comparing different years according to the associated temporal random effect estimate. As an example to illustrate interpretation, the relative risk of CRC in 1980 was 0.81, relative to 1973, for those < 50 years old. As such, the risk was less for younger-onset CRC in 1981 as compared with the risk in 1973. Thus, we see the same trends here as in Figure 2, where the < 50 model suggested reduced risk until the 1990s and then began increasing, whereas the ≥ 50 model first had increasing risk and later decreasing risk.

Estimates for the other model parameters, fixed effect parameters and spatial terms, are included in Supplemental Figures 1 and 2 (in the online version). The fixed effect results indicated association with incidence of CRC in Iowa for 5-year periods across the study time. From these, it was clear that the estimates differed in trend and value across many of the covariates for the 2 age-specific subgroup models. The estimates that were most different were for median home value and percent urban population and showed similarities between those produced from the "all" and the "≥ 50" models. However, the trends between these 2 models and the "< 50" model were nearly opposite. An example interpretation is as follows: Positive estimates indicate higher incidence of CRC for counties with higher percent white population. This relationship is the case for most years with respect to the "all" model that considers all age groups together and the "≥ 50" model. The 5-year intervals for which this relationship was not true were 1985 to 1990 and 2000 to 2005. The estimates in the "< 50" model suggested an opposite relationship, more incidence in counties with less percent white population, for all but the first 10 years of the study time. The spatial estimates were similar between the "all" and both age-specific subgroup models, where this effect was shared between the 2 outcomes. The darker areas, which indicate more risk, appear to be nearby the state capital, Des Moines, and in general, in the western part of the state. There was an additional cluster of increased risk in the east within counties through which Interstate 380 runs from Iowa City to Cedar Rapids then to Waterloo. The spatio-temporal estimates (data not shown) did show variation across counties and years as well as similarities between the univariate and multivariate models; however, there was no clear pattern.

Survival Following CRC Results

Table 1 displays descriptive statistics of the individual covariates adjusted for in the AFT "all," "≥ 50," and "< 50" models. These indicated that there was a difference in these risk factors of interest between the 2 survival outcome groups. Further, when comparing with the tables broken down by age-specific subgroup, there does not appear to be a difference in survival outcome for young-onset CRC with respect to age and death (*P*-value, .56) nor gender (*P*-value, 0.42), whereas older-onset CRC survival status remains statistically different for all risk factors of interest.

Figure 3 displays the temporal trends in the Iowa SEER data with respect to individual-level survival time. The temporal frailty

Table 1 Descriptive Statistics for the Individual-level Risk Factors From the Iowa Surveillance, Epidemiology, and End Results Data Used in the Accelerated Failure Time Model

Risk Factor	No CRC Death			CRC Death			*P* Value
	N	Col %	Mean (SD)	N	Col %	Mean (SD)	
All age groups							
Age at diagnosis, y			71.0 (12.4)			72.0 (12.8)	**<.001**
% Urban population			57.5 (23.3)			57.5 (23.4)	.93
Marital status							**<.001**
Single	3502	6.7		2600	8.2		
Married	31,076	59.6		17,335	55.0		
Previous	17,546	33.7		11,605	36.8		
Gender							**.02**
Male	25,323	48.9		15,062	47.8		
Female	26,801	51.4		16,478	52.2		
High grade							**<.001**
No	43,087	82.7		23,380	74.1		
Yes	9037	17.3		8160	25.9		
Previous tumors							**<.001**
No	32,601	62.5		23,828	75.5		
Yes	19,523	37.5		7712	24.5		
Radiation therapy							**<.001**
No	48,452	93.0		28,358	89.9		
Yes	3672	7.0		3182	10.1		
CRC surgery							**<.001**
No	3365	6.5		7683	24.4		
Yes	48,759	93.5		23,857	75.6		
Less than 50 years old							
Age at diagnosis, y			42.3 (6.4)			42.2 (6.5)	.56
% Urban population			57.5 (28.3)			57.5 (28.3)	1.00
Marital status							**.06**
Single	442	15.1		291	17.7		
Married	2092	71.3		1130	68.9		
Previous	399	13.6		219	13.3		
Gender							.42
Male	1534	52.3		879	53.6		
Female	1399	47.7		761	46.4		
High grade							**<.001**
No	2385	81.3		1072	65.4		
Yes	548	18.7		568	34.6		
Previous tumors							**<.001**
No	2245	76.5		1432	87.3		
Yes	688	23.5		208	12.7		
Radiation therapy							**.007**
No	2444	83.3		1314	80.1		
Yes	489	16.7		326	19.9		
CRC surgery							**<.001**
No	141	4.5		259	15.8		
Yes	2802	95.5		1381	84.2		
Greater than 50 years old							
Age at diagnosis, y			72.7 (10.5)			73.6 (10.9)	**<.001**
% Urban population			57.5 (28.3)			57.5 (28.4)	1.00

Table 1 Continued

Risk Factor	No CRC Death			CRC Death			*P* Value
	N	Col %	Mean (SD)	N	Col %	Mean (SD)	
Marital status							**<.001**
Single	3060	6.2		2309	7.7		
Married	28,984	58.9		16,205	54.2		
Previous	17,147	34.9		11,386	38.1		
Gender							**.01**
Male	23,789	48.4		14,183	47.4		
Female	25,402	51.6		15,717	52.6		
High grade							**<.001**
No	40,702	82.7		22,308	74.6		
Yes	8489	17.3		7592	25.4		
Previous tumors							**<.001**
No	30,356	61.7		22,396	74.9		
Yes	18,835	38.3		7504	25.1		
Radiation therapy							**<.001**
No	46,008	93.5		27,044	90.4		
Yes	3183	6.5		2856	9.6		
CRC surgery							**<.001**
No	3234	6.6		7424	24.8		
Yes	45,957	93.4		22,476	75.2		

This is for all age groups, the younger-onset age group, and the older-onset age group separately.
P value is from either a 2-sample *t* test (continuous) or a χ^2 test (categorical).
Bold *P* values indicate borderline to strong statistical significance.
Abbreviations: Col % = column percentage per category; CRC = colorectal cancer; Mean (SD) = the mean and standard deviation of the continuous variable; N = number of individuals per category.

estimates were in relation to survival time in months after CRC diagnosis. Thus, a higher number indicated longer, improved survival time. These results indicated that both age-specific subgroups have improved survival over the study time. Further, the "< 50" model indicated that the younger age group had a longer survival time following diagnosis on average. The overall rate of change for the different age groups ("all," "< 50," "≥ 50") was as follows: 0.04, 0.05, and 0.04, respectively. These rates of change estimates translated to an average of a 4% or 5% increase in survival time per year across the study time for these age groups in the state of Iowa.

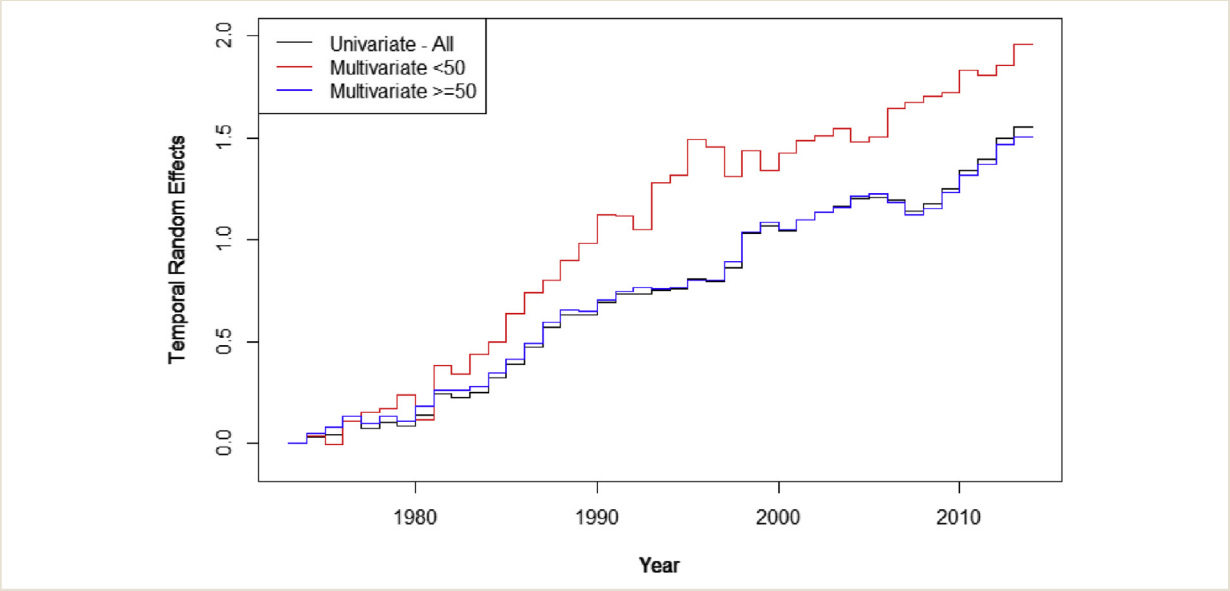
Estimates for the other model parameters, fixed effect parameters and spatial terms, are included in Supplemental Table 3 (in the online version) and Figure 3. These suggested that several of the individual covariates were associated with longer CRC survival time for the "all" model and the "≥ 50" model including: being married at the time of diagnosis, being female, not having a high grade of cancer, having a previous tumor, not having radiation therapy, having surgery, and being younger at the time of diagnosis. The estimates from the "< 50" model suggested similar relationships to those listed above except being married at the time of diagnosis; here, the estimate suggested that this lead to worse survival compared with being single at the time of diagnosis. The spatial random effect was very similar for all 3 models, and some secondary testing[25] (see Supplemental Tables 4 and 5 in the online version) suggested that this effect could be representing sociodemographic status, population health, and US Environmental Protection Agency emissions from several sources. The population health characteristics indicated longer survival for individuals in counties with higher prevalence of obesity, diabetes rates, percentage of excessive drinkers, and percentage of excess smokers. After adjusting the secondary assessment for percent urban population, many of these sociodemographic and population health associations were reduced, and the county US Environmental Protection Agency emissions found to have negative associations, indicating shorter survival times, were carbon monoxide from light-duty gas vehicles and motorcycles sources and ammonia from agriculture and forestry sources. Similarly to the incidence model estimates, the spatio-temporal estimates did show variation across counties and years as well as similarities between the univariate and multivariate models; however, there was no clear pattern (data not shown).

Discussion

Previous work has suggested that there were differences in both incidence and survival related to CRC by age group and that those younger than the age of 50 were worse off with respect to both evaluations of risk.[3] The results here agreed with respect to young-onset incidence of CRC in Iowa. However, results with respect to survival indicated that younger individuals tended to survive longer following a CRC diagnosis.

The results related to fixed effect estimates and secondary assessments of the spatial random effect in the survival models were interesting. For these, we saw improved survival in Iowa for those with previous tumors and with no radiation (not for the young-onset age group). These results could be related to more frequent

Figure 3 Temporal Variation Estimates for the Accelerated Failure Time Model Assessing Age-specific Survival Following Colorectal Cancer Diagnosis for Colorectal Cancer Underlying Cause of Death in Iowa



screenings among those with history of previous tumors, and radiation therapy could cause more harm than good for older-onset CRC. Then the secondary assessment results suggested some conflicting results. For the sociodemographic status and population health risk factors, we saw improved survival for individuals in counties with higher median home values, percent Medicaid eligible, percent with no insurance, percent persons living in poverty, percent high education, obesity prevalence, diabetes rate, percent excessive drinkers, and percent current smokers. Many of these could also be indicators of urbanicity because larger cities tend to have these sociodemographic and population health characteristics; to confirm this, we performed the secondary assessment while adjusting for percent urban population as a fixed effect estimate (see Supplemental Table 5 in the online version). After adjusting for this risk factor, many of the conflicting results became null. Further, a negative association for ammonia from agriculture and forestry sources surfaced after adjustment. Individual-level sociodemographic information could be an important measure that is unavailable with this data set.

Another important aspect of trend over time is change points owing to significant events. We attempted to detect change points in both incidence and survival as in Carroll et al.[11,12] However, the addition of this parameter did not indicate an improvement in model fit over fitting a yearly trend as in the results presented here. Previous work suggested that the decrease in incidence for older-onset CRC and the rise in incidence for young-onset CRC both began in the mid-1980s.[3] However, our results suggested that both changes were later, with the change for decrease in older-onset at around diagnosis year 2000 and the change for young-onset at around diagnosis year 1989. Our finding related to the change in incidence of older-onset CRC at 2000 could be related to the implemented screening recommendation in 1996, assuming an approximately 4-year lag. All these statements are made with the caveat that this is specific to the Iowa SEER data.

For temporal trend in survival, previous studies have suggested that observed data on young-onset CRC also indicated worse survival for this age group,[26] but they did not suggest any changes in trend. This data from Iowa indicated that survival was better and the same as older onset for those diagnosed prior to their 50th birthday until 1998. Another study suggested that survival for young-onset cases could be significantly worse for non-Hispanic blacks,[27] and our population was largely white (99%; n = 82,434); thus, the results here might not illustrate the full decline of young-onset survival. Race was not adjusted for in our models as there were very few non-whites, and the percentage of missing failure time (1% for each race category) and CRC-specific death (36% of non-whites and 38% for whites) for whites and non-whites was nearly identical.

The most recent change in CRC screening recommendations occurred in May 2018 as a result of recent research.[6] This recommendation suggested that regular screenings should begin at age 45 rather than at age 50. This will likely help the risk associated with incidence of young-onset CRC as 47% of individuals in the young-onset group used in this study were diagnosed between ages 45 and 50. However, 53% of individuals classified as having young-onset CRC would remain uncaptured by the newly recommended screenings. Alternatively, a decrease to screenings beginning at 40 years old would encompass another 27% of the individuals considered here. The cost and burden of further lowering the recommended screening age should be considered.

This study was not without limitations. First, the data involved here was exclusively from the Iowa SEER registry; thus, interpretations are with respect to that specific population. Moreover, this population is largely white, and there are differences in

CRC among race groups. In fact, there were so few non-white individuals in this registry that adjustment for race at the individual level was not feasible. Another data limitation involved the anatomic and histologic information that was not available for the entirety of the SEER data. The only variable associated with severity of disease that was also available across all years of interest was grade, and although this does indicate severity, other variables could be more appropriate. The final limitation of the data involved lack of demographic information. From these results, we believe that disparities play a role in CRC incidence and survival; however, SEER data only offers age, race, and gender.

Conclusion

The results of this project suggested that, based on the population in Iowa, there are age—group-specific differences in CRC incidence and survival. We found that timing for changes in incidence of young- and older-onset CRC were slightly later than what had previously been suggested and that the revised estimated inflection year appeared to be more consistent with the 1996 change in recommended screenings for older individuals. For CRC survival, we found that all age group definitions tested indicated positive trend (eg, toward ever-improving survival times), with the highest rate of change in the younger-onset group. These results suggest that improved screening for people over 50 decreased the incidence of older-onset CRC, but did not benefit the younger population. This analysis underscores the importance of the revised recommendation that screening should instead begin at 45. The finding that survival following a CRC diagnosis continues to improve across the study time further underscores the importance of early detection for CRC.

Clinical Practice Points

- CRC is common worldwide with 140,250 diagnoses and 50,630 deaths estimated for the United States in 2018. Guidelines current to the most recent individuals in our analysis suggested regular screenings beginning at age 50 have reduced the incidence of CRC. However, incidence continues to rise among those under 50. Less is known about survival following CRC diagnosis, but research has suggested that younger cases may also have worse survival.
- Our results indicated that both overall and older onset CRC incidence began to decline in the early 2000s, while younger onset incidences decreased until the late 1980s but then increased steeply through the 2000s. The risk for those younger than 50 years of age first exceeded the risk for those 50 years or older in 2007. Survival times did increase for overall CRC, older onset CRC, and young onset CRC throughout the study period, with young onset CRC increasing at a higher rate. The spatial variation assessment indicated that the survival was positively associated with several variables of interest, most notably disparities including better access to healthcare and higher sociodemographic status.
- This analysis suggested that improved screening for people over 50 decreased the incidence of older onset CRC but did not benefit the younger population. Further, this work underscores the importance of the revised recommendation that screening should instead begin at 45. The finding that survival following a CRC diagnosis continues to improve across the study time further underscores the importance of early detection for CRC.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences.

Disclosure

The authors have stated that they have no conflicts of interest.

Supplemental Data

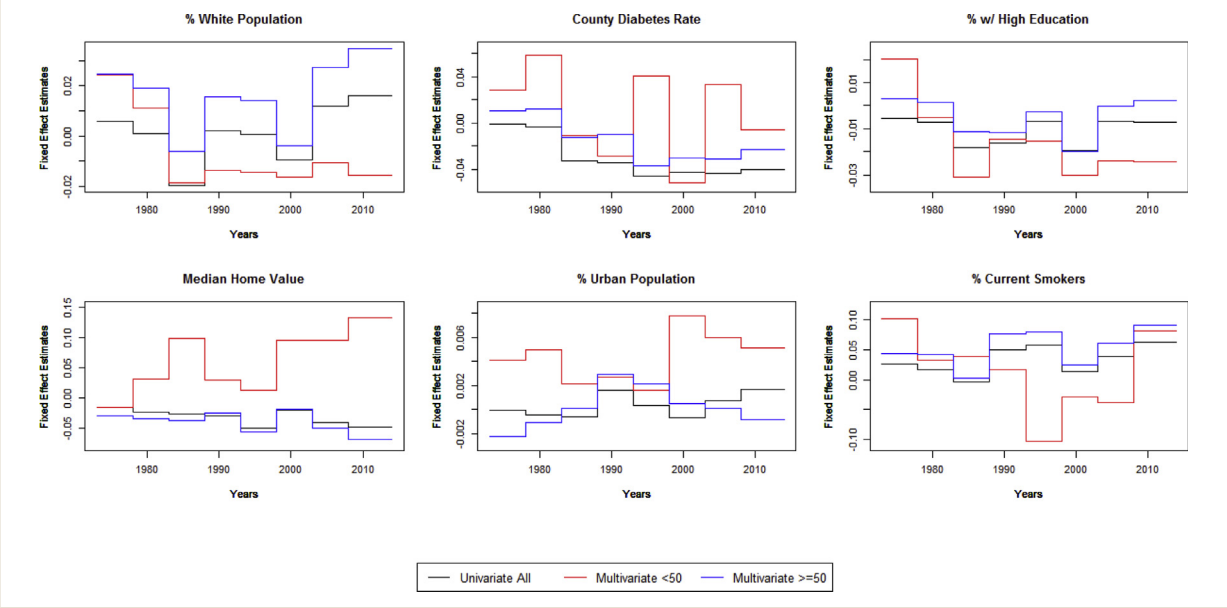
Supplemental tables and figures accompanying this article can be found in the online version at https://doi.org/10.1016/j.clcc.2018.12.001.

References

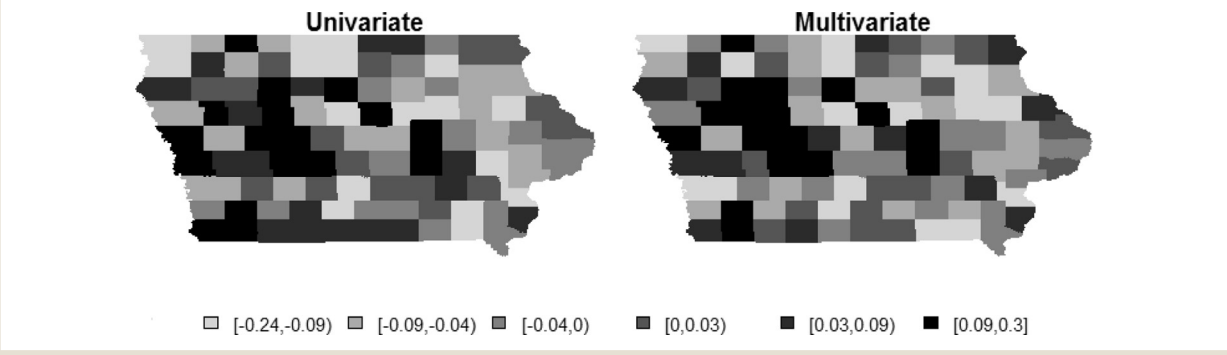
1. American Cancer Society. *Cancer Facts and Figures 2018*. Atlanta: American Cancer Society; 2018. Available at: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf. Accessed: May 2, 2018.
2. You YN, Xing Y, Feig BW, Chang GJ, Cormier JN. Young-onset colorectal cancer: is it time to pay attention? *Arch Intern Med* 2012; 172:287-9.
3. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J Natl Cancer Inst* 2017; 109.
4. Kasi PM, Shahjehan F, Cochuyt J, Li Z, Colibaseanu D, Merchea A. Rising proportion of young individuals with rectal and colon cancer [Epub ahead of print]. *Clin Colorectal Cancer* 2018. https://doi.org/10.1016/j.clcc.2018.10.002.
5. Rodriguez L, Brennan K, Karim S, Nanji S, Patel SV, Booth CM. Disease characteristics, clinical management, and outcomes of young patients with colon cancer: a population-based study. *Clin Colorectal Cancer* 2018; 17:e651-61.
6. Siegel RL, Miller KD, Jemal A. Colorectal cancer mortality rates in adults aged 20 to 54 years in the United States, 1970-2014. *JAMA* 2017; 318:572-4.
7. Kopetz S, Chang GJ, Overman MJ, et al. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. *J Clin Oncol* 2009; 27:3677-83.
8. Knorr-Held L. Bayesian modeling of inseperable space-time variation in disease risk. *Stat Med* 2000; 19:2555-67.
9. Knorr-Held L, Besag J. Modelling risk from a disease in time and space. *Stat Med* 1998; 17:2045-60.
10. Carroll R, Lawson AB, Jackson CL, Zhao S. Assessment of spatial variation in breast cancer-specific mortality using Louisiana SEER data. *Soc Sci Med* 2017; 193:1-7.
11. Carroll R, Lawson AB, Zhao S. Temporally dependent accelerated failure time model for capturing the impact of events that alter survival in disease mapping [Epub ahead of print]. *Biostatistics* 2018. https://doi.org/10.1093/biostatistics/kxy023.
12. Carroll R, Lawson AB, Zhao S. *A data-driven approach for estimating the change-points and impact of major events on disease risk* 2018, *Spat Spatiotemporal Epidemiol*.
13. Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Extensions to multivariate space time mixture modeling of small area cancer data. *Int J Environ Res Public Health* 2017; 14.
14. Lawson AB, Carroll R, Faes C, Kirby RS, Aregay M, Watjou K. Spatiotemporal multivariate mixture models for Bayesian model selection in disease mapping. *Environmetrics* 2017; 28.
15. Centers for Disease Control and Prevention. *What are the risk factors of colorectal cancer?*. Atlanta: Centers for Disease Control and Prevention; 2016. Available at: https://www.cdc.gov/cancer/colorectal/basic_info/risk_factors.htm. Accessed: May 15, 2018.
16. Bureau of Health Workforce. *Area Health Resource Files (AHRF) Rockville, MD*. US Department of Health and Human Services, Health Resources and Services Administration; 2015. Available at: http://ahrf.hrsa.gov/. Accessed: September 2016.
17. University of Wisconsin Population Heath Institute. *County Health Rankings - Adult Smoking, Madison, WI* 2018. Available at: http://www.countyhealthrankings.org/app/iowa/2018/measure/factors/9/data, Accessed: April 2018.
18. Centers for Disease Control and Prevention. *Diabetes County Data Indicators*. Atlanta: Centers for Disease Control and Prevention; 2016. Available at: https://www.cdc.gov/diabetes/data/countydata/countydataindicators.html. Accessed: April 2018.
19. Lawson AB. Disease map reconstruction. *Stat Med* 2001; 20:2183-204.
20. Lawson AB, Banerjee S, Haining R, Ugarte MD. *Handbook of Spatial Epidemiology*. Boca Raton, FL: CRC Press; 2016.
21. Christensen R, Johnson W. Modelling accelerated failure time with a Dirichlet process. *Biometrika* 1988; 75:693-704.

22. Onicescu G, Lawson A, Zhang J, Gebregziabher M, Wallace K, Eberth JM. Bayesian accelerated failure time model for space-time dependency in a geographically augmented survival model. *Stat Methods Med Res* 2017; 26: 2244-56.
23. Orbe J, Ferreira E, Nunez-Anton V. Comparing proportional hazards and accelerated failure time models for survival analysis. *Stat Med* 2002; 21:3493-510.
24. Zhang J, Lawson AB. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *J Appl Stat* 2011; 38:591-603.
25. Carroll R, Zhao S. Spatial heterogeneity: gaining relevance from the random. *Spat Spatiotemporal Epidemiol* 2018; 25:11-7.
26. Khan SA, Morris M, Idrees K, et al. Colorectal cancer in the very young: a comparative study of tumor markers, pathology and survival in early onset and adult onset patients. *J Pediatr Surg* 2016; 51:1812-7.
27. Holowatyj AN, Ruterbusch JJ, Rozek LS, Cote ML, Stoffel EM. Racial/ethnic disparities in survival among patients with young-onset colorectal cancer. *J Clin Oncol* 2016; 34:2148-56.

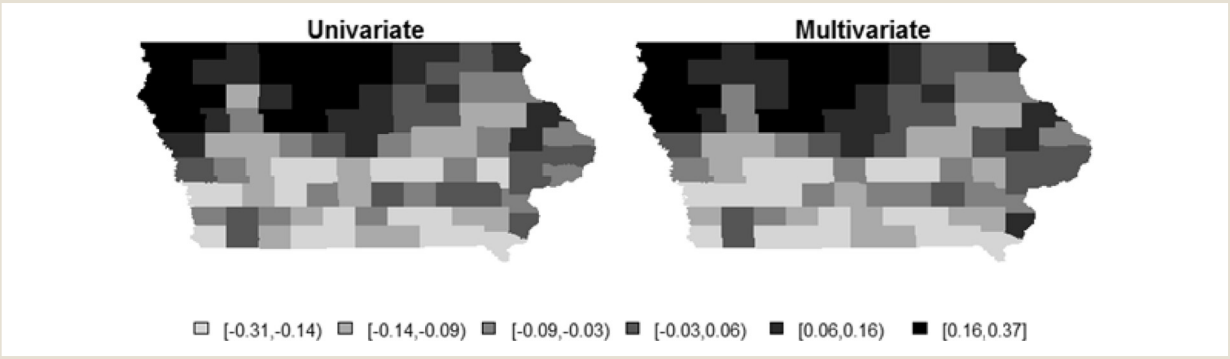
Supplemental Figure 1 Fixed Effect Estimates for the Univariate and Multivariate Specifications Across Both Age Groups From the Poisson Incidence Model Using Iowa Surveillance, Epidemiology, and End Results Data



Supplemental Figure 2 Spatial Estimates Produced for the Univariate and Multivariate Specifications Across Both Age Groups From the Poisson Incidence Models With Iowa Surveillance, Epidemiology, and End Results Data



Supplemental Figure 3 Spatial Random Effect Estimates for the Accelerated Failure Time Survival Models With Iowa Surveillance, Epidemiology, and End Results Data



Supplemental Table 1 Correlation Matrix for the Risk Factors From Various Sources Adjusted for in the Poisson Models

	% White Population	% Smokers	% High Education	Median Home Value (in $10,000)	% Urban Population	Diabetes Rate
% White population	1.00	−0.25	−0.63	−0.33	−0.65	0.24
% Smokers	−0.25	1.00	−0.21	−0.35	0.13	0.14
% High education	−0.63	−0.21	1.00	0.73	0.52	−0.53
Median home value	−0.33	−0.35	0.73	1.00	0.47	−0.58
% Urban population	−0.65	0.13	0.52	0.47	1.00	−0.29
Diabetes rate	0.24	0.14	−0.53	−0.58	−0.29	1.00

Supplemental Table 2 Table of Change in Risk Calculations From the Poisson Incidence Model Using Iowa Surveillance, Epidemiology, and End Results Data

yr_1	yr_2 1973	1980	1990	2000	2010
Age at diagnosis <50 y					
1973	1				
1980	0.81	1			
1990	0.78	0.94	1		
2000	1.08	0.97	1.16	1	
2010	1.33	1.20	1.37	1.14	1
Age at diagnosis ≥50 y					
1973	1				
1980	1.21	1			
1990	1.45	1.39	1		
2000	1.09	1.19	1.15	1	
2010	0.89	0.96	0.75	0.78	1

This calculation was done as follows: $\Delta_{yr_1 yr_2} = \exp(yr_2 - yr_1)$. The result from this calculation is interpreted as: the relative risk of CRC in yr_1 was $\Delta_{yr_1 yr_2}$, relative to y_2, for a specific age group.

Supplemental Table 3 Individual-level Fixed-effect Estimates for the Accelerated Failure Time Survival Model With Iowa Surveillance, Epidemiology, and End Results Data

Covariate	Univariate, β (95% CI)	Multivariate, β (95% CI) <50 y	≥50 y
Marital status			
Single (Ref)	0.00	0.00	0.00
Married	0.30 (0.25, 0.32)[a]	0.53 (0.33, 0.73)[a]	0.63 (0.46, 0.77)[a]
Previously married	−0.08 (−0.52, 0.08)	0.36 (−0.67, 1.43)	0.48 (0.42, 0.54)[a]
Gender			
Male (Ref)	0.00	0.00	0.00
Female	0.28 (0.23, 0.30)[a]	0.11 (−0.08, 0.33)	0.17 (−0.36, 0.75)
High grade			
No (Ref)	0.00	0.00	0.00
Yes	−1.31 (−1.37, −1.29)[a]	−1.61 (−1.81, −1.41)[a]	−1.41 (−1.48, −1.35)[a]
Previous tumors			
No (Ref)	0.00	0.00	0.00
Yes	1.02 (0.97, 1.04)[a]	1.64 (1.34, 1.94)[a]	1.00 (0.94, 1.06)[a]
Radiation therapy			
No (Ref)	0.00	0.00	0.00
Yes	−0.18 (−0.27, −0.15)[a]	−0.32 (−0.56, −0.07)[a]	−0.14 (−0.25, −0.05)[a]
CRC surgery			
No (Ref)	0.00	0.00	0.00
Yes	3.96 (3.89, 3.98)[a]	2.91 (2.55, 3.24)[a]	3.79 (3.71, 3.86)[a]
Age at diagnosis	−0.03 (−0.03, −0.03)[a]	−0.01 (−0.03, 0.00)	−0.03 (−0.04,−0.03)[a]

These are fixed across space and time. Other components in these models include: w_i, γ_j for the univariate or γ_{j1} and γ_{j2} for the multivariate, and φ_{ij}.
Abbreviations: CI = credible interval; CRC = colorectal cancer; Ref = reference.
[a]Indicates a significant association.

Supplemental Table 4 Secondary Assessment Results Utilizing the Spatial Random Effect From the Accelerated Failure Time Survival Model With Iowa Surveillance, Epidemiology, and End Results Data

Risk Factor	Univariate, β (95% CI)	Multivariate, β (95% CI)
Access to and quality of healthcare		
Total hospitals	0.35 (−0.28, 0.98)[a]	0.39 (−0.24, 1.02)[a]
Total hospitals w/Medicare certification	0.29 (−0.26, 0.85)[a]	0.32 (−0.23, 0.88)[a]
Total hospitals w/colorectal screening	0.08 (−0.41, 0.58)	0.11 (−0.38, 0.61)
Sociodemographic status		
Median home value	2.10 (0.86, 3.34)[b]	2.10 (0.87, 3.34)[b]
% Urban population	0.53 (−1.43, 2.48)	0.53 (−1.43, 2.48)
% Medicaid eligible	2.98 (1.49, 4.47)[b]	3.01 (1.53, 4.49)[b]
% With no insurance	1.07 (−0.08, 2.24)[a]	1.07 (−0.10, 2.26)[a]
% Persons living in poverty	1.00 (−0.04, 2.05)[a]	0.91 (−0.13, 1.97)[a]
% Persons with high education	1.85 (−0.04, 3.72)[a]	1.83 (−0.05, 3.70)[a]
Population health		
Obesity prevalence	1.31 (−0.62, 3.24)[a]	1.31 (−0.62, 3.25)[a]
Diabetes rate	0.43 (−0.15, 1.02)[a]	0.38 (−0.20, 0.97)[a]
% Excessive drinkers	2.23 (1.34, 3.18)[b]	2.28 (1.39, 3.21)[b]
% Current smokers	0.89 (0.05, 1.79)[b]	0.87 (0.03, 1.77)[b]
EPA emissions		
Carbon monoxide		
Light-duty gas vehicles and motorcycles	55.41 (55.39, 55.44)[b]	14.41 (14.41, 14.42)[b]
Light-duty gas trucks	70.78 (70.76, 70.81)[b]	28.86 (28.85, 28.87)[b]
Non-road gasoline	0.04 (−1.92, 2.00)[b]	−283.27 (−283.45, −283.09)[b]
Filterable PM10		
Other fugitive dust	27.28 (27.10, 27.46)[b]	10.24 (10.09, 10.40)[b]
Primary PM10		
Other fugitive dust	27.28 (27.10, 27.46)[b]	10.24 (10.09, 10.40)[b]
Ammonia		
Agriculture and forestry	0.01 (−1.95, 1.97)	487.68 (487.39, 487.97)[b]

The emissions variables presented were limited to only those with a significant association in at least 1 model.
Abbreviations: CI = credible interval; EPA = United States Environmental Protection Agency; PM = particulate matter.
[a]Indicates borderline association.
[b]Indicates a significant association.

Supplemental Table 5 Urbanicity-adjusted Secondary Assessment Results Utilizing the Spatial Random Effect From the Accelerated Failure Time Survival Model With Iowa Surveillance, Epidemiology, and End Results Data

Risk Factor	Univariate, Adjusted β (95% CI)	Multivariate, Adjusted β (95% CI)
Access to and quality of healthcare		
Total hospitals	0.11 (−0.42, 0.64)	0.19 (−0.33, 0.70)
Total hospitals w/Medicare certification	0.15 (−0.32, 0.61)	0.21 (−0.24, 0.67)
Total hospitals w/colorectal screening	−0.30 (−0.74, 0.15)[a]	−0.22 (−0.65, 0.22)[a]
Sociodemographic status		
Median home value	0.49 (−0.59, 1.57)	0.49 (−0.57, 1.56)
% Urban population	0.00 (−0.18, 0.18)	0.00 (−0.17, 0.17)
% Medicaid eligible	0.88 (−0.93, 2.67)	0.95 (−0.86, 2.74)[a]
% With no insurance	0.82 (−0.25, 1.92)[a]	0.80 (−0.28, 1.89)[a]
% Persons living in poverty	0.78 (−0.14, 1.70)[a]	0.72 (−0.19, 1.63)[a]
% Persons with high education	0.71 (−1.00, 2.40)	0.60 (−1.10, 2.30)
Population health		
Obesity prevalence	0.86 (−0.98, 2.70)	0.86 (−0.98, 2.70)
Diabetes rate	0.50 (0.00, 1.01)[a]	0.50 (0.00, 1.01)[a]
% Excessive drinkers	1.30 (0.46, 2.19)[b]	1.30 (0.46, 2.19)[b]
% Current smokers	0.76 (0.01, 1.54)[b]	0.76 (0.01, 1.54)[b]
EPA emissions		
Carbon monoxide		
Light-duty gas vehicles and motorcycles	105.21 (105.18, 105.24)[b]	−0.29 (−0.29, −0.29)[b]
Light-duty gas trucks	188.14 (188.11, 188.17)[b]	303.78 (303.76, 303.81)[b]
Non-road gasoline	193.25 (193.12, 193.39)[b]	221.83 (221.63, 222.03)[b]
Filterable PM10		
Other fugitive dust	341.23 (341.06, 341.39)[b]	459.65 (459.50, 459.81)[b]
Primary PM10		
Other fugitive dust	341.23 (341.06, 341.39)[b]	459.65 (459.50, 459.81)[b]
Ammonia		
Agriculture and forestry	−409.97 (−410.26, −409.67)[b]	−339.55 (−339.70, −339.40)[b]

The emissions variables presented were limited to only those with a significant association in at least 1 model.
Abbreviations: CI = credible interval; EPA = United States Environmental Protection Agency; PM = particulate matter.
[a]Indicates borderline association.
[b]Indicates a significant association.