Published in final edited form as:

Stat Med. 2020 March 30; 39(7): 1011–1024. doi:10.1002/sim.8458.

Survival model methods for analyses of cancer incidence trends in young adults

Wan Yang¹, Rebecca D Kehm¹, Mary Beth Terry^{1,2}

¹Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York

²Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, New York

Abstract

Recent studies have reported increases in cancer incidence in adults under 50 years. However, there remains uncertainty about whether these are true increases or a result of incidental findings from increased medical imaging. To evaluate these trends, we propose an alternative method to age-period-cohort analyses based on survival modeling. Simulations show that our method is capable of quantifying cohort effects within various backgrounds including increasing medical imaging. We applied the method to analyze the changes in cancer incidence rates for 44 anatomic sites, stratified by sex, by birth cohort for individuals born from 1945 to 1969 in the US based on incidence data from the Surveillance, Epidemiology, and End Results (SEER) program, and tested the validity of our models using later birth cohorts (1970-1974 and 1975-1979). We found that cancer risks have increased significantly in 15 sites (9 in men and 11 in women) for 25-49 year-olds. These results were consistent with previous findings from age-period-cohort analyses. Further, based on our simulations, these increases were independent of increased medical imaging and support substantial, increased extrinsic risks in the identified cancers. Although our approach has several limitations including the restriction to the younger age range and requirement of complete data for all ages of interest, we demonstrate many advantages of our approach including the ease in implementation and interpretation of cohort effects, robustness to various period backgrounds, and ability to make predictions. Our approach should help epidemiologists evaluate cohort effects using incidence data for cancer or other diseases.

Keywords

cohort effect; survival function; cancer risk; young adults; time-trends; cancer incidence

Correspondence to: Dr. Wan Yang, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th Street, Room 520, New York, NY 10032 (Phone: 212-305-0421; wy2202@columbia.edu).

Conflict of Interest Disclosure: The authors have no conflicts of interest to disclose.

Data and code availability: The cancer incidence data are publicly available from the Surveillance, Epidemiology, and End Results program (<https://seer.cancer.gov>). Model code to implement the models and simulations is publicly available at https://github.com/wan-yang/cancer_cohort_trends.

1. Introduction

Analyses of temporal trends in cancer incidence are important for identifying changes in extrinsic risk factors that can be potentially prevented. Temporal trends can be affected by age, period, and cohort-specific effects. Age is an essential intrinsic factor determining the risk of cancer development, as genetic alternations (both genetic and epigenetic changes) tend to accumulate with age while many biological functions key to carcinogenesis including DNA repair capacity, metabolism, and immune defense tend to degrade with age.¹⁻³ A period effect is a change that occurs at a particular time and affects all ages and cohorts uniformly. It could come from both time-specific exposures or changes in detection or measurement classification. For the former, for example, changes in US Food and Drug Administration (FDA)'s regulation in 1973 on the dose of folic acid allowed in supplements and subsequent mandate in 1998 on fortification of flour and grains have increased the intake of folate since 1973 and particularly after 1998; this secular change has been thought to be influential in changing colorectal cancer incidence in the US.⁴ In contrast, changes in detection or measurement classification, e.g., due to detection method or diagnosis criteria, can lead to apparent changes in incidence not related to risk. For instance, from 1980 to 2006, the number of computed tomography (CT) scanning medicine procedures increased from 3.3 to 67 millions; this increase in medical imaging use has been cited as a reason behind the recent rise in thyroid cancer incidence.⁵ A cohort effect is a change related to a specific birth cohort or generation of population, due to exposure associated with birth, specific age groups, or generations. Therefore, quantifying the contribution from age, period, and cohort effects is essential in diagnosing population-level changes in risk factors.

Age-period-cohort models⁶⁻⁸ have been commonly applied to estimate the contributions of age, period, and cohort effects. These models decompose the three effects into additive components in order to estimate each effect based on the temporal trend in the data. However, because cohort (i.e. birth year) is the interaction between period (i.e. calendar time) and age, the three effects are fundamentally linked and cannot be identified separately in additive models. Due to this identifiability issue, age-period-cohort models instead estimate the combined effect of period and cohort and further diagnose the relative contributions of the two by graphing the age-cohort or age-period association. A recent study⁸ further proposed a measure termed C-P score (i.e. Cohort-Period score) computed from the model to measure the relative contributions of period and cohort effect. Nevertheless, the implementation and interpretation of age-period-cohort models remain challenging given the identifiability issues.

In this study, we propose an alternative method to estimate the cohort effect based on survival modeling. A survival function, in the context of cancer incidence, is a function that gives the probability that a population of a given age will not be diagnosed with cancer. Its complement thus gives the probability of the population being diagnosed with cancer at a given age, or the age-specific cancer incidence rate. As such, we propose to first model, for each cohort, the changes in risk by age using a survival function with age as the sole dependent variable. Because the survival function is highly simplified and only models changes in incidence by age in a given cohort, period effects that are related to a specific time and thus differentially affect different ages of the same cohort would likely be treated

as noise or model error. This rigidity allows the survival function to filter out the period effects and specifically estimates the impact of age for each cohort. To estimate the cohort effect, the method can then compare the risk estimates for different cohorts computed by the survival function, which reduce confounding from period effects. For instance, to detect risk changes by cohort, we can pool the cohort-specific risk estimates from the survival function and uses linear regression model against cohort to quantify the change by cohort.

Here, we first test our approach on five simulated datasets representing scenarios with five different combinations of cohort and period effects. We then apply it to estimate cohort trends in cancer incidence among 25-49 year-olds, for whom incidence in several cancers have been reported to increase in recent years based on age-period-cohort analyses.⁹⁻¹¹ Whether such increases are due to increased exposure or detection based on increased medical imaging remains a topic of debate.^{5,12-18} Therefore, to inform public health action, consistent findings are needed from alternative methods capable of quantifying cohort effects within the background of increasing medical imaging. We perform this analysis for 44 anatomic sites using US population-based data from the Surveillance Epidemiology and End Results (SEER) Program.¹⁹ To further validate the proposed method, we also generate predictions for later cohorts using the model and compare the predictions to independent, out-of-sample data.

The structure of this article is as follows. In Section 2, we describe the SEER dataset that is used in our main analysis. In Section 3, we describe our proposed method in detail. In Section 4, we present simulation tests and results under various cohort and period effects. In Section 5, we present results of the SEER data analyses on cancer cohort trends. We conclude with a discussion on our findings on the recent cancer trends and how our findings compared with those from previous studies using age-period-cohort models. We also briefly discuss the relative limitations and advantages of our method compared with age-period-cohort models.

2. Data sources and processing

In this study, we combined SEER 9 (1973-1991), SEER 13 (1992-1999) and SEER 18 (2000-2015) incidence data during 1973 to 2015 and referred to the combined data as SEER data. The SEER program has a case ascertainment rate of 98% and currently covers approximately 34.6% of the US population.^{19,20} We restricted this analysis to ages between 25 and 49 years, as cancer diagnosis is rare for <25 year-olds and most population-level cancer screening programs are implemented for individuals 50 years and above. In addition, trends in cancer incidence among 25-49 year olds are particularly of interest here given the recent rapid rise in several cancers in this age group including colorectal, kidney, and thyroid cancer¹¹ and recent debates as to how much these reported increases are due to changes in exposure versus increased in medical imaging.^{5,12-18,21-23}

We classified cancer anatomic site and histology based on the primary site and the third edition of the International Classification of Disease for Oncology (ICD), using SEER's "ICD-0-3/WHO_2008" variable. Only cases defined as malignant were included in this study. Note that, for rare cancers, SEER groups anatomic sites, for example, "all soft tissue

including heart” and “other digestive organs”. We further excluded sites with very low incidence rate (<0.1 per 100,000 person-year). In total, 44 sites remained in our analyses including 37 for men and 38 for women. In addition, for three cancers with known differential etiology by subsite or histology, we conducted analyses by subsite or histology. These three cancers were 1) esophagus,^{24,25} divided into esophageal adenocarcinoma (i.e., all histology coded as 8140-8141, 8143-8145, 8190, 8200-8202, 8210-8215, 8220-8221, 8230-8231, 8260-8263, 8310, 8401, 8480-8482, 8490, 8550-8551, 8570-8574, or 8576) and esophageal squamous cell carcinoma (SSC; i.e., all histology coded as 8050-8052, 8070-8076, or 8082-8084), respectively; 2) oral cavity and pharynx, divided into subsites related to human papilloma virus (HPV) infection (i.e., all primary sites coded as C019, C024, C090, C091, C98, C99, C100-C104, C108, C109, or C142) and those related to smoking (i.e., all primary sites coded as C020-C023, C028-C031, C039-C041, C048-C052, C058-C062, C068, or C069), respectively; and 3) stomach, divided into cardia (i.e., primary site coded as C160) and non-cardia (i.e., all primary sites coded as C161-C166), respectively.

For each of the 44 sites and 6 subsites, age specific incidence rates (from 25 to 49 years in 1-year intervals) were first computed for each 1-year birth cohort and then averaged to obtain the rates for each non-overlapping 5-year birth cohort since 1890. We used 5-year rather than single-year birth cohort to reduce noise in the data. This and subsequent analyses below were performed for men and women, separately. As a sensitivity analysis, we also performed analyses using shorter cohort intervals of 4-year and 3-year, respectively.

3. Methods

3.1. Survival modeling

We used survival models to estimate the age-specific hazard rate for each birth cohort from the age-specific incidence data. The basic survival model took the following form:

$$S(a) = \exp(-H) = \exp\left(-\int_{a_0}^a h(t)dt\right) \quad [1]$$

where $S(a)$ is the survival probability at age a , computed as 1 minus the incidence rate at age a ; H is the cumulative hazard; and h is the hazard rate of developing a given cancer. The variable a_0 is the initial age with a nonzero hazard rate; it is assumed the same for all birth cohorts and estimated based on data for all cohorts included.

As the age range spans 25 years and the hazard rate varies with age, we tested three different survival models, representing three plausible changes in hazard rate with age. The first model (referred to as the “constant” model) assumed a constant hazard rate (h) for the entire 25-49 age range. The second (referred to as “two-step” model) assumed constant hazard rates h_1 for the first half of the age range (i.e. 25-37 years) and h_2 for the second half (i.e. 38-49 years). The third (referred to as the “acceleration” model) assumed continuous change in hazard rate with age. That is, $h(t) = h_0 + \delta t$, with δ as the change in hazard rate (similar to acceleration). See Table 1 for detailed model equations. To assess model goodness of fit, we computed the adjusted r^2 and the Bayesian Information Criterion (BIC) for each model fit.

For each cancer site, we selected the best-fit model based on the BIC (i.e., the one with the lowest BIC) and used the site-specific best-fit model for all analyses involving the same site.

For SEER data from 1973-2015, to allow for comparison across cohorts, we only included cohorts with complete data for all ages between 25-49 years in this study. These were: 1) using 5-year cohort interval, five birth cohorts born during 1945-1949, 1950-1954, ..., and 1965-1969; 2) using 4-year cohort interval, 6 birth cohorts born during 1946-1949, 1950-1953, ..., and 1966-1969; and 3) using 3-year cohort interval, 7 birth cohorts born during 1947-1949, 1950-1952, ..., and 1965-1967.

3.2. Estimating cohort trends

To analyze changes in cancer risk across cohorts, we first computed the mean hazard rate across all ages included in the survival model described above for each cohort (see Table 1 for equations). This measure thus combines the age effects for each cohort. To detect the cohort trend, we regressed the cohort-specific mean hazard rate ($\bar{h}_c$) on birth year (c), i.e.:

$$\bar{h}_c = \beta c \quad [2]$$

The slope of the above regression (β) thus represents the changes in risk per 1-year birth cohort. As such, we considered the cohort effect statistically significant when the P value for the slope (β) is less than 0.05. Note that the linear regression here is intended to detect consistent change (either increase or decrease) across multiple birth cohorts. Changes by cohort can also be visualized by plotting the cohort-specific mean hazard rates versus cohort birth years.

To quantify the changes in risk by cohort (either increase or decrease), we further computed the relative risk change per 1-year birth cohort (ΔRR) following Equation 1 as:

$$\Delta RR = \frac{\text{sign}(\beta)(1 - \exp(-25|\beta|))}{\overline{IR}} \times 100\% \quad [3]$$

where $\text{sign}(\beta)$ gives the direction of change, i.e. '+' for increase and '-' for decrease and $25|\beta|$ gives the total absolute changes over the 25 years (i.e., age 25-49) for each 1-year birth cohort. $\overline{IR}$ is the mean incidence rate averaged over all included cohorts and ages; a different reference (e.g., mean incidence rate for the first cohort) can be used.

3.3. Predicting age-specific incidence rates for later birth cohorts

We used the combined model and parameters extrapolated from the above five cohorts to predict age-specific incidence rates for two later cohorts born during 1970-1974 and 1975-1979. For evaluation, we computed the correlation (r) between the predicted and observed age-specific incidence rates and the relative root mean squared error (RRMSE).

We conducted all analyses using the R software (version 3.3.2; <https://www.r-project.org>). Model code is publicly available at https://github.com/wan-yang/cancer_cohort_trends.

4. Simulation tests and results

We first applied our method to five test datasets representing five different scenarios: 1) Cohort effect alone, for which both the baseline risk and aging-related risk are increased by 10% per later cohort. 2) Period effect due to changes in exposure alone, for which, to mimic the aforementioned folate intake and colorectal cancer risk example, the aging-related risk is reduced by 10% for all ages from 1998 and partially reduced for all ages between 1973 and 1997 based on the prevalence of folate supplement use. 3) Period effect due to changes in detection alone, for which, to mimic the CT use and thyroid cancer diagnosis example, the diagnosis rate and thus incidence are increased based on the prevalence of CT use over time. 4) Cohort effect combining period effect due to changes in exposure, i.e., combining 1 and 2 using the same cohort/period effect thereof. and 5) Cohort effect combining period effect due to changes in detection, i.e., combining 1 and 3 using the same cohort/period effect thereof. The code we used to generate the simulation datasets and perform the analysis is publicly available at https://github.com/wan-yang/cancer_cohort_trends.

Figure 1 shows the results, with the fits to test data using a survival function (the “acceleration” model in Table 1) in the first column and estimated cohort effects in the second column. For the first scenario with cohort effect alone (Figure 1 A and B), our method observed the imposed cohort effect ($P < 0.001$ and the estimated relative risk increase per 1-year cohort $RR = 2.88\%$; 95% confidence interval (CI): 2.88%, 2.88%). Note that, as no period effect is added in this scenario, the RR estimate here is the true cohort effect. In the second and third scenarios representing two different period effects—one with decreasing incidence by period due to increased protective exposure in later years and one with increasing incidence by period due to increased detection from medical imaging—our method correctly estimated no significant cohort effects ($P = 0.08$, $RR = -0.19\%$; 95% CI: -0.42% , 0.04% and $P = 0.57$, $RR = -0.37\%$, 95% CI: -2.24% , 1.49% , respectively; Figure 1 C–F). In contrast, when the same cohort effect in Scenario 1 was present in addition to either period effect, the proposed method correctly identified significant cohort effects ($P < 0.01$ for both cases). In addition, while the uncertainty increased with the presence of period effects as shown by the larger confidence intervals, the mean estimates were close to the true imposed cohort effect ($RR = 2.78\%$, 95% CI: 2.59%, 2.98% for Scenario 4 and $RR = 2.90\%$, 95% CI: 1.15%, 4.64% for Scenario 5, compared to the true cohort effect $RR = 2.88\%$ as shown in Scenario 1; Figure 1 G–J). These results suggest that our method is able to reliably estimate the true cohort effect despite the confounding factors due to, e.g., period effects.

5. Results: Cancer cohort trends based on the SEER data

5.1. Model selection and fit

Table S1 shows the performance of the three survival models tested and the selected best-fit model for each cancer site. Figure 2 shows model fits to age-specific incidence data for three cancers with the highest, median, and lowest incidence rate for men and women, respectively (see supplementary material for model fits for all 44 sites and the 6 subsites), using the site-specific best-fit models. Despite the large range of incidence rates across ages and cancer sites, the selected model was able to accurately estimate the age-specific

incidence rates for those cancers (for men: adjusted $r^2=0.98$ for lung and bronchus, 0.97 for soft tissue including heart, and 0.79 for gallbladder cancer, respectively; for women: adjusted $r^2=1.00$ for breast, 0.96 for pancreas and 0.74 for mesothelioma, respectively).

5.2. Cancers with *increasing incidence* by birth cohort

Cancer risk significantly increased with birth year in 9 sites in men (Table 2) and 11 sites in women (Table 2 and 2 additional cancers in Table S2). Of those, 5 cancer sites had significantly increased risk for both men and women (all $P<0.05$; Figure 3 for men and Figure 4 for women; middle columns). Specifically, for each later 1-year birth cohort, the risk of being diagnosed with colorectal cancer increased by 4.04% (95% CI: 1.88, 6.21) in men and 3.71% (95% CI: 2.84, 4.59%) in women, compared to the mean risk across cohorts. For thyroid cancer, the risk increased by 8.64% (95% CI: 5.24, 12.04%) in men and 8.08% (95% CI: 5.35, 10.80%) in women per cohort. For kidney (including renal pelvis; same below), the cancer risk increased by 8.31% (95% CI: 5.85, 10.77%) in men and 8.92% (95% CI: 6.36, 11.48%) in women per cohort. For myeloma, the risk increased by 3.14% (95% CI: 0.22, 6.06%) in men and 6.89% (95% CI: 2.39, 11.40%) in women per cohort. For small intestine, the cancer risk increased by 6.57% (95% CI: 3.22, 9.91%) in men and 7.02% (95% CI: 2.14, 11.90%) in women per cohort. Further, our model estimates predict that the incidence of these five cancers will continue to increase in later birth cohorts (Figure 3 for men and Figure 4 for women; right columns).

In men, the other 3 cancer sites with statistically significant trends by birth year were anus, anal canal and anorectum, other biliary, and penis (all $P<0.05$; Table 2). All these cancers were relatively rare in 25-49 year-old men (<1 per 100,000 person-year). In women, the other 6 sites with statistically significant trends by birth year were melanoma of the skin, leukemia, stomach, other female genital organs, and other endocrine including thymus (all $P<0.05$; Table 2 and Table S2). Of note, unlike women, men had decreasing but not significant trend in incidence for melanoma of the skin and stomach cancer (either overall or by subsite; Table S2).

5.3. Cancers with *decreasing incidence* by birth cohort

In men aged 25-49 years, five cancers had significantly decreased risk by birth year (all $P<0.05$); these were non-Hodgkin lymphoma (RR=−5.64%, 95% CI: −9.68, −1.61%; Table 3 and Figure S1), lung and bronchus (RR=−11.31%; 95% CI: −15.23, −7.38%;), urinary bladder (RR=−6.66%, 95% CI: −8.15, −5.16%), larynx (RR=−9.24%, 95% CI: −13.24, −5.24%), and esophageal squamous cell carcinoma (RR=−14.36%, 95% CI: −17.91, −10.8%). In addition, continued decreases in risk in these cancer for later birth cohorts were predicted and, as shown in Figure S1 (right column), these predicted estimates closely matched with available observations for those born during 1970-1974 and 1975-1979.

In women aged 25-49 years, although not statistically significant, similar to men, a decrease in risk by birth year was also estimated for lung and bronchus cancer (RR=−5.40%; 95% CI: −13.71, 2.91%;) and larynx cancer (RR=−3.06%, 95% CI: −12.10, 5.98%;). A significant decrease in esophageal squamous cell carcinoma was also evident in women

($RR=-13.47\%$, 95% CI: $-21.45, -5.48\%$). In addition, another three cancers had significantly decreased risk by birth year (all $P<0.05$); these were cervix uteri ($RR=-2.05\%$, 95% CI: $-3.08, -1.02\%$), ovary ($RR=-6.99\%$, 95% CI: $-11.89, -2.09\%$), and other non-epithelial skin ($RR=-6.38\%$, 95% CI: $-11.02, -1.74\%$). Further, risk decreases in these three cancers are likely to continue in later birth cohorts born during 1970-1974 and 1975-1979, as shown in Figure S2 (right column).

5.4. Sensitivity analysis

Our first sensitivity analysis stratifying three cancers (i.e. esophagus, oral cavity and pharynx, and stomach) by histology/subsite showed differential cohort trends by histology/subsite. For esophageal cancer, the overall incidence decreases were mainly driven by significant decreases in incidence rate of esophageal squamous cell carcinoma ($RR=-14.36\%$, 95% CI: $-17.91, -10.8\%$ in men and $RR=-13.47\%$, 95% CI: $-21.45, -5.48\%$ in women; Table S2) whereas the incidence rates of esophageal adenocarcinoma have increased ($RR=4.24\%$, 95% CI: $-3.88, 12.37\%$ in men and $RR=6.38\%$, 95% CI: $-8.22, 20.99\%$ in women). For oral cavity and pharynx cancer, there were differential trends by subsite and sex: subsites related to smoking showed a significant decrease in men ($RR=-6.77\%$, 95% CI: $-4.83\%, -2.89\%$) but not in women ($RR=0.26\%$, 95% CI: $-3.18, 3.70\%$); in addition, subsites related to HPV infection appeared to increase by cohort in both men ($RR=4.95\%$, 95% CI: $-4.33, 14.23\%$) and women ($RR=2.71\%$, 95% CI: $-2.65, 8.07\%$; Table S2). For stomach cancer, in women, we found a significant incidence increase in the non-cardia subsites ($RR=3.88\%$, 95% CI: $1.57, 6.20\%$) and nonsignificant increase in cardia subsites ($RR=3.41\%$, 95% CI: $-4.13, 10.95\%$); in contrast, in men incidence rates decreased in both cardia and non-cardia subsites, although neither was significant (Table S2).

Sensitivity analyses on different cohort intervals suggest highly consistent results using either 5, 4, or 3 years (see site-specific comparison in Tables S3 for men and S4 for women). In particular, the five sites shown to experience significant incidence increases in both men and women in the main analysis (i.e. using 5-year cohort interval) remained in analyses using 4 or 3-year cohort interval (see Figs S3–4 for 4-year and Figs S5–6 for 3-year). Similarly, all the sites shown to experience significant incidence decreases remained so using shorter cohort intervals (see Figs S7–8 for 4-year and Figs S9–10 for 3-year). However, with shorter cohort intervals, two cancers showing marginally significant increasing cohort trends in both men and women in the main analysis—i.e., cancer of the anus, anal canal and anorectum and cancer of other digestive organs (Table 2 vs. Figs S3–6)—and three cancers showing marginally significant decreasing cohort trends in the main analysis—i.e., stomach cancer in men (Table S2 vs. Fig S7), testicular cancer (Table S2 vs. Fig S9), and non-Hodgkin Lymphoma in women (Table 3 vs. Fig S8 and S10)—became statistically significant. Of note, with shorter cohort intervals, more cohorts were included in the analysis and the cohort trend tended to be less linear (see, e.g., middle panels in Figs S3–6).

6. Discussion

Motivated by the recent debates regarding temporal trends in cancer incidence,^{5,12–18,21–23,26} here we proposed an alternative method to estimate changes in cancer incidence by cohort (i.e. cohort effect) in the presence of age and period effects. In particular, to inform public health action, we aimed to test whether the increasing trends in cancer incidence in adults under 50 years are due to changes in exposures or increasing medical imaging. Simulation tests suggest that our method is able to accurately estimate the cohort effect in the presence of period effects including those due to increased detection (Figure 1). Further, for cancers that have population-level screening for ages under 50 years, like breast, we did not find any cohort effects due to screening (Table S2). Using this alternative approach, we were thus able to rule out that increased medical imaging is the sole explanation for the increase in incidence. The accuracy of our forecasted estimates also supports that the increases in identified cancers under age 50 years are real.

We found that risks of developing cancer have increased significantly in 9 sites in 25–49 year-old men and 11 sites in 25–49 year-old women since cohorts born in 1945–1949. In particular, 3 major solid cancers with incidence rates >1 per 100,000 person-year (i.e., colorectal, thyroid, and kidney) and 2 blood cancers (leukemia and myeloma) have had consistent risk increases with birth year in both 25–49 year-old men and women. These results are consistent with a recent analysis¹¹ using age-period-cohort models, which found that incidence in the same five cancers have increased significantly in adults aged 25–49 years in later birth cohorts. Given that the intrinsic cancer risk due to hereditary^{27–30} and random genetic alterations^{23,31,32} would be the same across birth cohort, our findings point to a substantial increased extrinsic risk to development of these cancers for younger adults in the recent decades.

The specific extrinsic risk factors behind these recent increases in younger cancers, however, remain largely unknown. While several lifestyle factors including obesity and physical inactivity have been cited as common risk factors for many cancers,^{33–35} most of these findings have come from epidemiologic and etiologic studies in older adults. Whether the same risk factors are responsible for the recent cancer risk increases in adults under age 50 remains unclear. As shown in Figure 3 (for men) and Figure 4 (for women), our models predicted that, without effective public health intervention, incidence rates in 25–49 year-olds would be even higher for later cohorts born in 1970–1974 and 1975–1979.

Our study also identified several cancers with decreased risk across birth cohorts; these included non-Hodgkin lymphoma, cancers of the lung and bronchus, larynx, smoking-related oral and pharynx subsites, esophagus (SCC), and urinary bladder, for which the decreases were significant in men but often not in women. These results are in general consistent with findings from Sung et al.¹¹ based on age-period-cohort models (note that Sung et al. did not distinguish men and women in their analysis). The decrease in non-Hodgkin lymphoma was likely due to the control of HIV/AIDS epidemic whereas the decreases in the other five aforementioned cancers reflect and consolidate the success in smoking reduction through regulations (e.g., smoking laws) and individual behavior changes (e.g., smoking cessation).^{36,37} However, our results support sex-specific differences in

smoking-related cancer decreases and suggest that such regulations may have different impact and that specific programs to young women are needed.

In addition, for 25-49 year-old women, significant decreases were estimated for both cervix uteri and ovary cancer. For cervical cancer, the decrease was likely due to earlier detection through screening,³⁸ as cohorts included in our analysis were born in 1945-1949 to 1964-1969. The impact of HPV vaccination³⁹⁻⁴¹ may further reduce incidence in HPV-related cancers in later birth cohorts. Nevertheless, contrary to the general decreases in HPV-related cancers in women, we found that, in 25-49 year-old men, penile and anal cancers (i.e., penis, anus, anal canal and anorectum) as well as HPV-related oral and pharynx cancer have increased by birth cohort. As HPV is a known cause of these cancers⁴²⁻⁴⁴ and the uptake of the HPV vaccine has been lower in males than females,⁴⁵ the diverging trends suggest that enhanced education and HPV vaccination in boys and young men may be needed to reduce the risk of these cancers in these subpopulations.

Comparing with the age-period-cohort models, our proposed method has several limitations. Unlike the age-period-cohort methods that include data for all ages and cohorts, to ensure fair comparison across cohorts, our method only includes cohorts with complete data for all ages of interest (e.g., 25-49 years here). In this study, we also restricted the analysis to younger ages to reduce impacts not due to biological aging (e.g., population-level cancer screening); however, extension to the full age range is possible with the use of more flexible survival functions. Finally, our method does not estimate the period effect.

However, the strengths of our method likely outweigh these limitations in a number of important ways. First, the statistical model results are straightforward. Both the survival modeling used to estimate the mean hazard rate (or risk) for each cohort and the linear regression to estimate the cohort trend are commonly applied and intuitive. We note, however, for the second step, nonlinear methods such as joinpoint regression⁴⁶ (vs. linear regression here) can be used to more precisely describe the cohort trend; and we recommend doing so for better diagnosis of changing trends that reflect nonlinear patterns, e.g., myeloma in men (Fig. 3) and cancer of the small intestine in women (Fig. 4) as well as when shorter cohort intervals are used and more cohorts are included (e.g., 3 or 4-year cohort intervals in our sensitivity analyses). Secondly, the method provides direct estimates of cohort effect without needing to rely on any combined or relative measures. Thirdly, as the survival model estimates the risk for each age, predictions for each age in later cohorts can be made easily using the same model (e.g., the predictions shown in Figures 3 and 4). These predictions not only can be used to further validate model estimates but also to guide public health action.

In summary, we have proposed and tested a new method for estimating cohort effects. While the method is introduced in the context of cancer research, extension to other research is possible as survival modeling is commonly applied in epidemiology. Using the proposed method, we have estimated the changes in cancer risk for 44 anatomic sites among adults under 50 years and identified significant increases in selected cancers among this age group. These findings are consistent with those based on age-period-cohort models. Altogether,

they support increased extrinsic risks in several cancers and call for improved etiologic research and public health intervention to prevent these younger cancers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

WY was partly supported by NIH/NIEHS grant ES009089; RDK was supported by NIH/NCI grant T32-CA094061; MBT was supported by NIH/NCI grants 2U01CA164920 and P30CA013696.

Abbreviations:

SEER Surveillance, Epidemiology, and End Results program

References:

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–674. [PubMed: 21376230]
2. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep*. 2014;15(12):1243–1253. [PubMed: 25381661]
3. Vander Heiden MG, DeBerardinis RJ. Understanding the Intersections between Metabolism and Cancer Biology. *Cell*. 2017;168(4).
4. Keum N, Giovannucci EL. Folic acid fortification and colorectal cancer risk. *Am J Prev Med*. 2014;46(3 Suppl 1):S65–72. [PubMed: 24512932]
5. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA*. 2006;295(18):2164–2167. [PubMed: 16684987]
6. Holford TR. Understanding the Effects of Age, Period, and Cohort on Incidence and Mortality-Rates. *Annu Rev Publ Health*. 1991;12:425–457.
7. Rosenberg PS, Check DP, Anderson WF. A Web Tool for Age-Period-Cohort Analysis of Cancer Incidence and Mortality Rates. *Cancer Epidem Biomar*. 2014;23(11):2296–2302.
8. Holford TR, Chen HS, Annett D, Krapcho M, Dorogaeva A, Feuer EJ. CP*Trends: An Online Tool to Compare Cohort and Period Trends Across Cancer Sites. *Am J Epidemiol*. 2019.
9. Zhu CR, Zheng TZ, Kilfoy BA, et al. A Birth Cohort Analysis of the Incidence of Papillary Thyroid Cancer in the United States, 1973–2004. *Thyroid*. 2009;19(10):1061–1066. [PubMed: 19732011]
10. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the United States, 1974–2013. *J Natl Cancer Inst*. 2017;109(8).
11. Sung H, Siegel RL, Rosenberg PS, Jemal A. Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. *Lancet Public Health*. 2019;4(3):E137–E147. [PubMed: 30733056]
12. Chen AY, Jemal A, Ward EM. Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. *Cancer*. 2009;115(16):3801–3807. [PubMed: 19598221]
13. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. *Nat Rev Endocrinol*. 2016;12(11):646–653. [PubMed: 27418023]
14. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974–2013. *JAMA*. 2017;317(13):1338–1348. [PubMed: 28362912]
15. Vigneri R, Malandrino P, Vigneri P. The changing epidemiology of thyroid cancer: why is incidence increasing? *Curr Opin Oncol*. 2015;27(1):1–7. [PubMed: 25310641]
16. Chow WH, Devesa SS. Contemporary epidemiology of renal cell cancer. *Cancer J*. 2008;14(5):288–301. [PubMed: 18836333]

17. King SC, Pollack LA, Li J, King JB, Master VA. Continued Increase in Incidence of Renal Cell Carcinoma, Especially in Young Patients and High Grade Disease: United States 2001 to 2010. *J Urology*. 2014;191(6):1665–1670.
18. Nepple KG, Yang L, Grubb RL, Strobe SA. Population Based Analysis of the Increasing Incidence of Kidney Cancer in the United States: Evaluation of Age Specific Trends From 1975 to 2006. *J Urology*. 2012;187(1):32–38.
19. National Cancer Institute. The Surveillance, Epidemiology, and End Results (SEER) Program. <https://seer.cancer.gov>.
20. National Cancer Institute. Overview of the SEER Program. 2019; <https://seer.cancer.gov/about/overview.html> Accessed 2/13, 2019.
21. Ashford NA, Bauman P, Brown HS, et al. Cancer risk: role of environment. *Science*. 2015;347(6223):727.
22. Song MY, Giovannucci EL. Cancer risk: Many factors contribute. *Science*. 2015;347(6223):728–729.
23. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*. 2015;347(6217):78–81. [PubMed: 25554788]
24. Zhang Y Epidemiology of esophageal cancer. *World J Gastroenterol*. 2013;19(34):5598–5606. [PubMed: 24039351]
25. Xie SH, Rabbani S, Petrick JL, Cook MB, Lagergren J. Racial and Ethnic Disparities in the Incidence of Esophageal Cancer in the United States, 1992–2013. *Am J Epidemiol*. 2017;186(12):1341–1351. [PubMed: 28641390]
26. Kehm RD, Yang W, Tehranifar P, Terry MB. 40 years of change in age- and stage-specific cancer incidence rates in US women and men. *JNCI Cancer Spectrum*. 2019:Accepted May 2019.
27. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *Jama-J Am Med Assoc*. 2016;315(1):68–76.
28. Graff RE, Moller S, Passarelli MN, et al. Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. *Clin Gastroenterol Hepatol*. 2017;15(8):1256–1264. [PubMed: 28130150]
29. Moller S, Mucci LA, Harris JR, et al. The Heritability of Breast Cancer among Women in the Nordic Twin Study of Cancer. *Cancer Epidem Biomar*. 2016;25(1):145–150.
30. Hjelmborg JB, Scheike T, Holst K, et al. The Heritability of Prostate Cancer in the Nordic Twin Study of Cancer. *Cancer Epidem Biomar*. 2014;23(11):2303–2310.
31. Tomasetti C, Li L, Vogelstein B. CANCER ETIOLOGY Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*. 2017;355(6331):1330–1334. [PubMed: 28336671]
32. Wu S, Powers S, Zhu W, Hannun YA. Substantial contribution of extrinsic risk factors to cancer development. *Nature*. 2016;529(7584):43–47. [PubMed: 26675728]
33. Cronin KA, Lake AJ, Scott S, et al. Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. *Cancer*. 2018;124(13):2785–2800. [PubMed: 29786848]
34. Yang L, Drake BF, Colditz GA. Obesity and Other Cancers. *J Clin Oncol*. 2016;34(35):4231–4237. [PubMed: 27903157]
35. Ehemann C, Henley SJ, Ballard-Barbash R, et al. Annual Report to the Nation on the status of cancer, 1975–2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer*. 2012;118(9):2338–2366. [PubMed: 22460733]
36. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975–2000. *J Natl Cancer Inst*. 2012;104(7):541–548. [PubMed: 22423009]
37. Ordóñez-Mena JM, Schottker B, Mons U, et al. Quantification of the smoking-associated cancer risk with rate advancement periods: meta-analysis of individual participant data from cohorts of the CHANCES consortium. *BMC Med*. 2016;14:62. [PubMed: 27044418]
38. Kim JJ, Burger EA, Regan C, Sy S. Screening for Cervical Cancer in Primary Care A Decision Analysis for the US Preventive Services Task Force. *Jama-J Am Med Assoc*. 2018;320(7):706–714.

39. Mesher D, Panwar K, Thomas SL, et al. The Impact of the National HPV Vaccination Program in England Using the Bivalent HPV Vaccine: Surveillance of Type-Specific HPV in Young Females, 2010–2016. *J Infect Dis.* 2018;218(6):911–921. [PubMed: 29917082]
40. Barnabas RV, Laukkanen P, Koskela P, Kontula O, Lehtinen M, Garnett GP. Epidemiology of HPV 16 and cervical cancer in Finland and the potential impact of vaccination: mathematical modelling analyses. *PLoS Med.* 2006;3(5):e138. [PubMed: 16573364]
41. Kim JJ, Goldie SJ. Health and economic implications of HPV vaccination in the United States. *N Engl J Med.* 2008;359(8):821–832. [PubMed: 18716299]
42. Beckmann AM, Daling JR, Sherman KJ, et al. Human Papillomavirus Infection and Anal Cancer. *International Journal of Cancer.* 1989;43(6):1042–1049. [PubMed: 2543642]
43. Lin CQ, Franceschi S, Clifford GM. Human papillomavirus types from infection to cancer in the anus, according to sex and HIV status: a systematic review and meta-analysis. *Lancet Infect Dis.* 2018;18(2):198–206. [PubMed: 29158102]
44. Stratton KL, Culkin DJ. A Contemporary Review of HPV and Penile Cancer. *Oncology-Ny.* 2016;30(3):245–+.
45. Lewis RM, Markowitz LE. Human papillomavirus vaccination coverage among females and males, National Health and Nutrition Examination Survey, United States, 2007–2016. *Vaccine.* 2018;36(19):2567–2573. [PubMed: 29650386]
46. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 2000;19(3):335–351. [PubMed: 10649300]

Figure 1.

Model validation. To test the proposed method, test datasets are generated using a survival model, with changes added to model parameters to mimic cohort effect alone (A and B), period effect due to changes in exposure (C and D), period effect due to changes in detection (E and F), cohort effect in addition to period effect due to changing exposure (G and H), and cohort effect in addition to period effect due to changing detection (I and J). In the left panel, points show the test data for 5 cohorts born in 1945-1949 to 1965-1969 and lines show model fits using the survival model (lighter colors show earlier cohorts and darker colors for

later cohorts). In the right panel, dots show estimated mean hazard rates for each cohort and lines show the linear regression fits. Numbers in the bottom show estimated cohort effect and significance level (those with $P < 0.05$ are highlighted in red). The true cohort effect for Scenarios 1 (A and B), 4 (G and H), and 5 (I and J) is $RR = 2.88\%$ as shown in B. The true cohort effect for Scenarios 3 (C and D) and 4 (E and F) is $RR = 0$.

Figure 2.

Example model fits to the age-specific incidence rate by cohort using the site-specific best-fit model (specified in the plot). Points show observations based on the SEER data and lines show model fitted estimates.

Figure 3.

Cancer risk trend in 25-49 year-old *men* for the six sites with increasing incidence by birth cohort in both men and women. Left column shows model fit to the age- and cohort-specific incidence rates for each cancer; middle column shows estimated cancer risk trend across cohorts born during 1945-1969; and right column shows predicted incidence rates for cohorts born during 1970-1974 and 1975-1979, compared to available observations. In the left and right columns, points show observations based on the SEER data and lines show model fits (left column) or predicted estimates (right column). Prediction accuracy measures

including the correlation (r) and the relative root mean squared error (RRMSE) are shown in the plot.

Figure 4.

Cancer risk trend in 25-49 year-old *women* for the six sites with increasing incidence by birth cohort in both men and women. Left column shows model fit to the age- and cohort-specific incidence rates for each cancer; middle column shows estimated cancer risk trend across cohorts born during 1945-1969; and right column shows predicted incidence rates for cohorts born during 1970-1974 and 1975-1979, compared to available observations. In the left and right columns, points show observations based on the SEER data and lines show model fits (left column) or predicted estimates (right column). Prediction accuracy measures

including the correlation (r) and the relative root mean squared error (RRMSE) are shown in the plot.

Table 1.

Equations for the three survival models.

Survival model	Cumulative hazard (H)	Mean hazard rate ($\bar{h}_c$)
constant	$h(a - a_0)$	h
two-step	$\begin{cases} h_1(a - a_0), & \text{for } a_0 \leq a \leq 37 \\ h_1(37 - a_0) + h_2(a - 37), & \text{for } 37 < a \leq 49 \end{cases}$	$(h_1 + h_2)/2$
acceleration	$h_0(a - a_0) + \frac{1}{2}\delta(a - a_0)^2$	$h_0 + 12.5\delta$

Table 2.

Cancers with increasing incidence trends by cohort for both men and women. Cancer sites are ordered by the average incidence rate across sex (1st column) and cohort trend (2nd column). Relative risk changes were computed per Eqn 3; changes significant at an $\alpha=0.05$ level are bolded and designated with an thick arrow for the cohort trend; otherwise in regular font and thin arrow.

Site	Rate per 100,000	Cohort trend*	Relative Changes (%)	
			Men	Women
Colon and Rectum	10.88	↑ ↑	4.04 (1.88, 6.21); P=0.0096	3.71 (2.84, 4.59); P=0.00088
Thyroid	9.88	↑ ↑	8.64 (5.24, 12.04); P=0.0039	8.08 (5.35, 10.8); P=0.0025
Kidney and Renal Pelvis	4.53	↑ ↑	8.31 (5.85, 10.77); P=0.0017	8.92 (6.36, 11.48); P=0.0016
Leukemia	4.35	↑ ↑	0.74 (-0.78, 2.25); P=0.22	2.06 (0.58, 3.54); P=0.021
Soft Tissue including Heart	1.90	↑ ↑	1.17 (0.2, 2.14); P=0.028	1.32 (-2.04, 4.68); P=0.3
Myeloma	1.23	↑ ↑	3.14 (0.22, 6.06); P=0.042	6.89 (2.39, 11.4); P=0.017
Anus, Anal Canal and Anorectum	0.81	↑ ↑	9.8 (3.58, 16.02); P=0.015	9.81 (-1.09, 20.71); P=0.064
Small Intestine	0.65	↑ ↑	6.57 (3.22, 9.91); P=0.0083	7.02 (2.14, 11.9); P=0.02
Other Female Genital Organs	0.51	- ↑	N/A	6.92 (0.7, 13.14); P=0.038
Other Endocrine including Thymus	0.46	↑ ↑	0.9 (-3.18, 4.98); P=0.53	3.39 (0.25, 6.53); P=0.041
Other Biliary	0.28	↑ ↑	4.38 (1.35, 7.4); P=0.019	3.42 (-1.33, 8.18); P=0.11
Penis	0.24	↑ -	4.36 (0.12, 8.6); P=0.047	N/A
Other Digestive Organs	0.10	↑ ↑	13.97 (-0.15, 28.09); P=0.051	14.37 (1.28, 27.47); P=0.04

* The first arrow is for men and the second for women.

Table 3.

Cancers with decreasing incidence trends by cohort for both men and women. Cancer sites are ordered by the average incidence rate across sex (1st column) and cohort trend (2nd column). Relative risk changes were computed per Eqn 3; changes significant at an $\alpha=0.05$ level are bolded and designated with an thick arrow for the cohort trend; otherwise in regular font and thin arrow.

Site	Rate per 100,000	Cohort trend*	Relative Changes (%)	
			Men	Women
Cervix Uteri	13.56	- ↓	N/A	-2.05 (-3.08, -1.02); P=0.0079
Ovary	9.11	- ↓	N/A	-6.99 (-11.89, -2.09); P=0.02
Non-Hodgkin Lymphoma	8.85	↓ ↓	-5.64 (-9.68, -1.61); P=0.021	-1.66 (-3.34, 0.01); P=0.051
Lung and Bronchus	8.53	↓ ↓	-11.31 (-15.23, -7.38); P=0.0028	-5.4 (-13.71, 2.91); P=0.13
Urinary Bladder	2.96	↓ ↓	-6.66 (-8.15, -5.16); P=0.00076	-2.19 (-7.98, 3.61); P=0.32
Larynx	0.95	↓ ↓	-9.24 (-13.24, -5.24); P=0.0052	-3.06 (-12.1, 5.98); P=0.36
Other Non-Epithelial Skin	0.95	↓ ↓	-0.53 (-1.67, 0.61); P=0.27	-6.38 (-11.02, -1.74); P=0.022
Esophagus (SCC)	0.3	↓ ↓	-14.36 (-17.91, -10.8); P=0.001	-13.47 (-21.45, -5.48); P=0.013

* The first arrow is for men and the second for women.