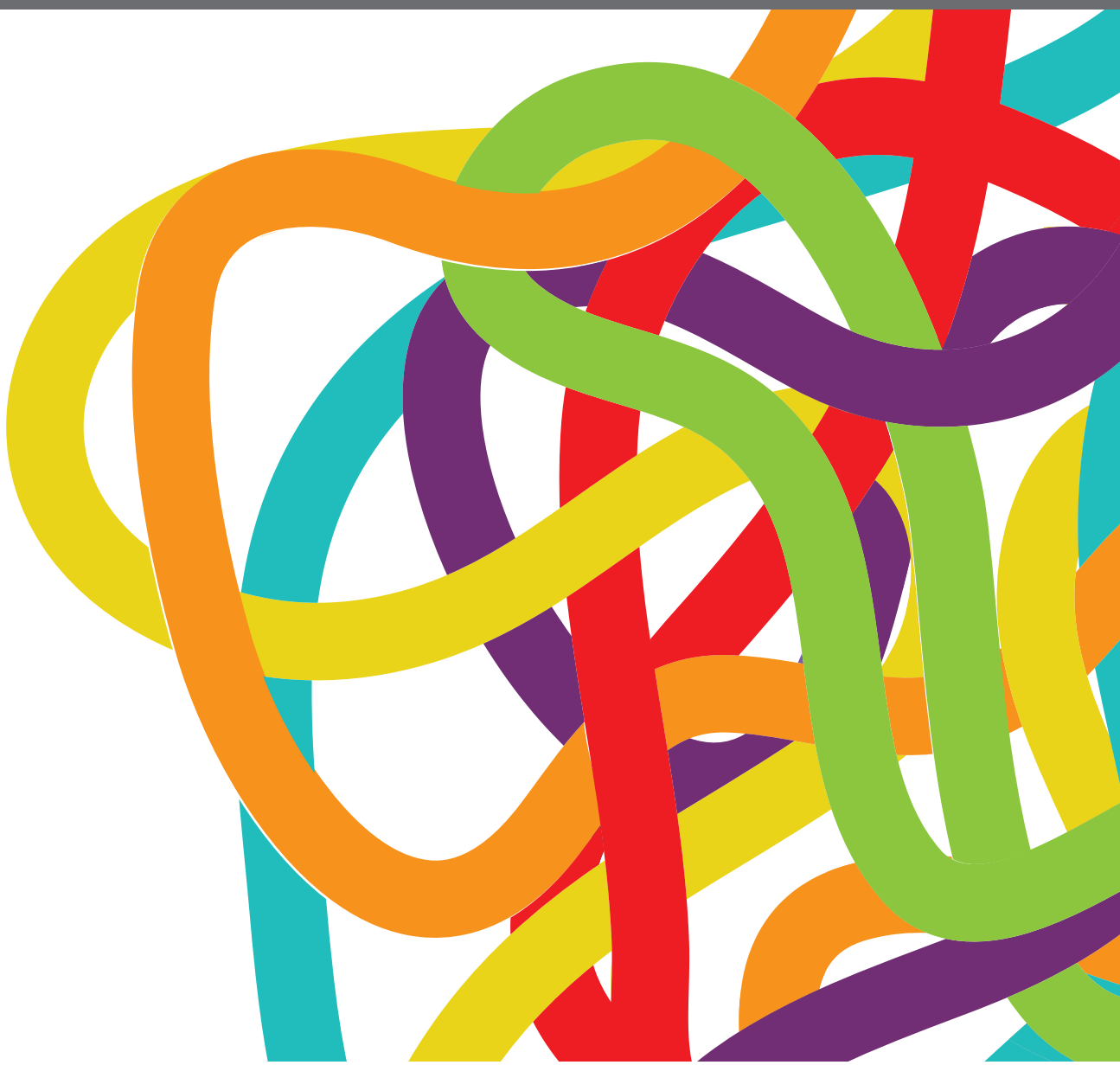
INSIGHTS IN CANCER EPIDEMIOLOGY AND PREVENTION: 2021

EDITED BY: Paolo Boffetta and Dana Kristjansson
PUBLISHED IN: Frontiers in Oncology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-792-6
DOI 10.3389/978-2-83250-792-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

INSIGHTS IN CANCER EPIDEMIOLOGY AND PREVENTION: 2021

Topic Editors:
Paolo Boffetta, Icahn School of Medicine at Mount Sinai, United States
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Citation: Boffetta, P., Kristjansson, D., eds. (2022). Insights in Cancer Epidemiology and Prevention: 2021. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-792-6

Table of Contents

frontiers in Oncology

ORIGINAL RESEARCH
published: 30 August 2021
doi: 10.3389/fonc.2021.726672

Glucose Intolerance and Cancer Risk: A Community-Based Prospective Cohort Study in Shanghai, China

*Juzhong Ke[1,2], Tao Lin[2], Xiaolin Liu[2], Kang Wu[2], Xiaonan Ruan[2], Yibo Ding[1], Wenbin Liu[1], Hua Qiu[2], Xiaojie Tan[1], Xiaonan Wang[2], Xi Chen[1], Zhitao Li[2] and Guangwen Cao[1]**

[1] Department of Epidemiology, Second Military Medical University, Shanghai, China, [2] Pudong New Area Center for Disease Control and Prevention, Pudong Institute of Preventive Medicine, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Yawei Zhang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Ni Li,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Junmei Miao Jonasson,
University of Gothenburg, Sweden

***Correspondence:**
Guangwen Cao
gcao@smmu.edu.cn

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: 17 June 2021
Accepted: 11 August 2021
Published: 30 August 2021

Citation:
Ke J, Lin T, Liu X, Wu K, Ruan X, Ding Y, Liu W, Qiu H, Tan X, Wang X, Chen X, Li Z and Cao G (2021) Glucose Intolerance and Cancer Risk: A Community-Based Prospective Cohort Study in Shanghai, China. Front. Oncol. 11:726672. doi: 10.3389/fonc.2021.726672

Background: Cancer becomes the leading cause of premature death in China. Primary objective of this study was to determine the major risk factors especially glucose intolerance for cancer prophylaxis.

Methods: A cluster sampling method was applied to enroll 10,657 community-based adults aged 15-92 years in Shanghai, China in 2013. A structured questionnaire and physical examination were applied in baseline survey. Prediabetes was diagnosed using 75-g oral glucose tolerance test. After excluding 1433 subjects including 224 diagnosed with cancer before and 1 year after baseline survey, the remaining 9,224 subjects were followed-up to December 31, 2020.

Results: A total of 502 new cancer cases were diagnosed. The cancer incidence was 10.29, 9.20, and 5.95/1,000 person-years in diabetes patients, those with prediabetes, and healthy participants, respectively ($p<0.001$). The multivariate Cox regression analysis indicated that age, prediabetes and diabetes, were associated with an increased risk of cancer in those <65 years, the hazard ratios (95% confidence interval) for prediabetes and diabetes were, 1.49(1.09-2.02) and 1.51(1.12-2.02), respectively. Glucose intolerance (prediabetes and diabetes) were associated with increased risks of stomach cancer, colorectal cancer, and kidney cancer in those <65 years. Anti-diabetic medications reduced the risk of cancer caused by diabetes. The multivariate Cox analysis showed that age, male, <9 years of education, and current smoking were associated with increased risks of cancer in those ≥65 years independently.

Conclusions: Glucose intolerance is the prominent cancer risk factor in adults <65 years. Lifestyle intervention and medications to treat glucose intolerance help prevent cancer in this population.

Keywords: type 2 diabetes mellitus, prediabetes, cancer, prospective cohort study, cancer prevention

INTRODUCTION

With the socioeconomic development, cancer has become the first leading cause of premature death (death before the mean life of a given population) in the most regions of China including Shanghai (1). The occurrence profiles of all cancer and site-specific cancers are changing, especially in younger adults. Incidence among this population is increasing for some site-specific cancers related to metabolic syndrome but decreasing for some cancers associated with infections or smoking (2–4). Update of controllable risk factor exposure is extremely important for the specific prophylaxis of cancer in population with an altered socioeconomic situation.

Type 2 diabetes mellitus and cancer are the major health problems worldwide. The age-standardized incidence of diabetes keeps increasing (5). Given that a substantial number of cancer cases are attributable to diabetes in different populations (6, 7), the increase in diabetes-related health burden and its impact on cancer risk represents an ongoing challenge. However, studies that examined cancer risk before diabetes diagnosis are relatively rare. Prediabetes is an often undiagnosed condition lasts for an average duration of 9.5 years before clinical onset of diabetes (8). Some reported indicated that prediabetes may increase the overall cancer risk (9, 10). However, many studies have failed to determine the role of prediabetes and diabetes on the risk of cancer (11–13). Thus, more reliable prospective cohort studies are needed to consolidate the etiological relationship between cancer and glucose intolerance, especially at a pre-diabetic level. Furthermore, C-reactive protein (CRP), a general marker of chronic low-grade inflammation, is associated with multiple chronic diseases including diabetes (14). CRP might have a joint effect with metabolic syndrome in carcinogenesis (15). It remains to determine if CRP contributes to carcinogenesis independently. Long-term use of metformin, an anti-diabetic, has been associated with a decreased risk of cancer, possibly because metformin works directly to cancer cells and/or the microenvironment (16–18). More recently, sulfonylureas, another groups of anti-diabetics, has been demonstrated to increase the risk of colorectal cancer in diabetes patients (19). Thus, the association of anti-diabetic medications with cancer risk remains controversial.

In this community-based prospective cohort study, we aimed to identify holistic risk factors especially glucose intolerance that can be applied for active prophylaxis of cancer in young adults and elderly adults, respectively. The study subjects aged between 15 years and 64 year were defined as young adults, while those aged 65 years or older were defined as elderly adults, according to the previous reports (20, 21). This study is of significance for cancer prophylaxis in the modern society, especially for the prevention of cancer-related premature death.

MATERIALS AND METHODS

Participants

This community-based prospective cohort study was performed in Pudong New Area, Shanghai, China. Participants are permanent residents who possess Shanghai household registration. Multistage stratified random cluster sampling was employed to sample study participants. A total of 38 urban streets and rural townships in Pudong were stratified into 3 strata according to the socioeconomic disparities from the Yearbook of Pudong government. Four streets in each stratum (6 urban streets and 6 rural townships) were randomly selected. Second, 16 urban communities and 18 rural villages were randomly selected from the 6 urban streets and 6 rural towns, respectively. Third, 11.0% families in each community/village were randomly selected. Individuals with diagnosed type I diabetes and pregnant women were excluded from this survey. A total of 12,382 eligible adults aged between 15 years and 92 years were initially recruited, among whom 10657 agreed to participate the study.

Baseline Survey

Baseline survey was carried out between January 13th and July 30st, 2013. Demographic characteristics including age, sex, marital status, years of education, lifestyle factors including smoking, alcohol consumption, tea consumption, physical activity, and preexisting medical conditions including family history of cancer, history of viral hepatitis, chronic atrophic gastritis, and use of anti-inflammatory agents were collected using a structured questionnaire (**Supplementary Table 1**). This face-to-face interview was conducted by trained investigators working in the community health centers. Current smoking was defined as smoking at least one cigarette a day in the past 6 months. Alcohol consumption and tea consumption were defined as regular drinker with at least three times per week in the past 6 months. Physical activity was defined as participating in sports activity for at least once per week in the past 5 years. Cancer family history was defined as at least one first-degree relative diagnosed with cancer.

All participants were invited to take physical examinations. Glucose, lipids, and CRP in the fasting plasma were measured using a HITACHI 7170A automatic biochemical analyzer. Glucose metabolism was determined using a 75g-oral glucose tolerance test (OGTT). Diabetes was defined as fasting plasma glucose ≥7.0 mmol/L, a 2-h plasma glucose ≥11.1 mmol/L by OGTT test, or on a glucose control medication. Participants with fasting plasma glucose between 6.1 mmol/L and 7.0 mmol/L and 2h plasma glucose <7.8 mmol/L were diagnosed as impaired fasting glucose (IFG). Participants with fasting plasma glucose <6.1mmol/L and 2h plasma glucose between 7.8 mmol/L and 11.1 mmol/L were diagnosed as impaired glucose tolerance (IGT). Both IFG and IGT are categorized as prediabetes (22). Participants with fasting plasma glucose <6.1 mmol/L and 2h plasma glucose <7.8 mmol/L were categorized as normal glucose tolerance (NGT). Body mass index (BMI) was calculated as weight (kg)/height (m^2). Hypertension was defined as blood pressure ≥140/90mm Hg or on a blood pressure-lowering medication. Dyslipidemia was defined as participants with plasma triglyceride ≥2.26mmol/L, total cholesterol ≥6.20mmol/L, low-density lipoprotein (LDL) ≥4.13mmol/L, high-density lipoprotein (HDL) <1.03mmol/L or on a cholesterol-lowering medication.

Follow-Up

The participants were excluded if confirmed not to possess Shanghai household registration (n=233), not to complete questionnaire and physical examination (n=976), and to have diagnosed cancer previously (n=170). The participants were also excluded if being diagnosed with cancer within the first year of follow-up (n=54). The remaining 9,224 eligible subjects (3,395 men and 5,829 women) were followed-up every three years. The flow diagram is shown in **Supplementary Figure 1**. Information on time-varying, physician-diagnosed incident diabetes, use of anti-diabetic medications, and covariates was obtained using a questionnaire during follow-up. The study protocol conformed to the 1975 Declaration of Helsinki and was approved by the ethics committee of the Center for Disease Control and Prevention of the Pudong New Area, Shanghai, China. A signed informed consent was obtained from each participant.

The outcomes of this cohort study are the incidences of all-cause primary cancers. Incident cancer cases were annually verified by data linkage with the cancer registration and management system in Shanghai, China. This system has covered 100% of registered population since 2002. The data in this system are reliable and their quality has been approved by the World Health Organization (23). Site-specific cancer types were identified according to the International Classification of Diseases, 10th edition (ICD-10), as previously described (1).

Statistical Analysis

For each participant, the expected number of person-years of follow-up for cancer incidence was calculated as the total years between their exact age at baseline survey and their exact age at cancer diagnosis, death, or 31st December 2020, whichever came first. Patients died of conditions unrelated to cancer were censored. One-way ANOVA test and Kruskal-Wallis test were applied to compare continuous variables. Difference in categorical variables was determined using chi-square test. Hazard ratio (HR) and 95% confidence intervals (CI) were calculated using the Cox proportional hazard model. Study participants were stratified into young adults and elderly adults. Baseline glycemic status, together with other variables including age, sex, marriage status, years of education, BMI, current smoking, alcohol consumption, tea consumption, physical activity, family history of cancer, history of hypertension, dyslipidemia, viral hepatitis, chronic atrophic gastritis, use anti-inflammatory agents, and serum CRP were introduced into the Cox proportional hazard model. The significant factors in the univariate Cox regression analysis were introduced into the multivariate Cox model to determine the factors independently associated with cancer. The Kaplan-Meier method was applied to estimate the effect of the factor proven to be significant in the Cox regression analysis on the cumulative incidence of cancer. Interaction terms were added in models to test the potential interactions of these covariates with baseline glycemic status. SPSS version 22.0 (SPSS Inc., Chicago, IL) was applied for statistical analysis. All statistical tests were two-sided. A p value of <0.05 was considered to be statistically significant.

RESULTS

Baseline Characteristics

Age and sex distribution of study subjects are shown in **Supplementary Figure 2**. In this cohort, 1454 participants (15.76%) were diagnosed with prediabetes, 1790 participants (19.41%) were diagnosed with diabetes at baseline. Baseline characteristics of the participants stratified by glycemic status are presented in **Table 1**. Compared to the NGT participants, those with prediabetes or diabetes were older and had higher frequencies of hypertension and dyslipidemia and higher levels of triglycerides, total cholesterol, LDL, CRP, and BMI and a lower level of HDL. Physical activity, history of viral hepatitis, and family history of cancer did not differ between the NGT participants and those with glucose intolerance (prediabetes + diabetes) statistically.

Association Between Glycemic Status and Cancer Incidence

Over a median of 7.48 years follow-up, cancer was found in 502 participants. The cumulative incidence of total cancer per 1,000 person-years in the participants with diabetes, those with prediabetes, and those with NGT was 10.29, 9.20, and 5.95 (log-rank test p value <0.001). In the multivariate Cox regression analysis, the interaction of age and glycemic status was significantly associated with an increased risk of cancer ($p_{\text{interaction}}$ = 0.040). The associations of all the variables with cancer risk were initially evaluated in the univariate Cox regression analysis. It was found that age, prediabetes, diabetes, BMI, hypertension, and CRP were significantly associated with an increased risk of total cancer in young adults. The multivariate Cox regression analysis demonstrated that age, prediabetes and diabetes independently associated with an increased risk of total cancer after the adjustment for the above significant variables in this population. In elderly adults, age, male, <9 years of education, and current smoking were independently associated with an increased risk of total cancer in the multivariate Cox regression analysis. Age, diabetes and current smoking were independently associated with an increased risk of all cancer in all the study population (**Table 2**).

Effect of Abnormal Glycemic Status and Anti-Diabetic Treatment on Cancer Incidence

We stratified participants with abnormal glycemic status into subgroups. Participants with prediabetes were categorized into IFG only, IGT only, and both IFG and IGT. Participants with diabetes were categorized into previously diagnosed diabetes or detected during baseline screening, use of anti-diabetic medications or not, or duration since the first diagnosis of diabetes. The multivariate Cox regression analysis demonstrated that, compared to participants with NGT at baseline, cancer incidence was significantly higher in prediabetes patients with IFG only, in diabetes patients detected during baseline screening rather than in those diagnosed previously, in diabetes patients without anti-diabetic

TABLE 1 | Baseline participant characteristics stratified by glycemic status.

Age group (Years old)	Characteristics	Glycemic status			Total	*p*
		NGT	Prediabetes	Diabetes		
15-64	Age (years)	50.81 ± 11.18	55.29 ± 7.63	56.05 ± 6.77	52.25 ± 10.41	<0.001$ 1*** 3***
	Male (%)	1538 (32.72%)	309 (33.70%)	440 (42.19%)	2287 (34.33%)	<0.001§ 1* 2*
	Urban (%)	2818 (59.94%)	496 (54.09%)	583 (55.90%)	3897 (58.50%)	0.001§ 1* 3*
	Married (%)	4215 (89.66%)	867 (94.55%)	976 (93.58%)	6058 (90.95%)	<0.001§ 1* 3*
	>= 9 years of education (%)	4186 (89.04%)	762 (83.10%)	858 (82.26%)	5806 (87.16%)	<0.001§ 1* 3*
	Current smoking (%)	817 (17.38%)	146 (15.92%)	255 (24.45%)	1218 (18.29%)	<0.001§ 1* 2*
	Alcohol consumption (%)	524 (11.15%)	126 (13.74%)	155 (14.86%)	805 (12.09%)	0.001§ 3*
	Tea consumption (%)	1320 (28.08%)	265 (28.90%)	361 (34.61%)	1946 (29.21%)	<0.001§ 1* 2*
	Physical activity (%)	1194 (25.40%)	210 (22.90%)	258 (24.74%)	1662 (24.95%)	0.274§
	Family history of cancer (%)	286 (6.08%)	77 (8.40%)	58 (5.56%)	421 (6.32%)	0.017§ 1* 2*
	Hypertension (%)	1245 (26.48%)	443 (48.31%)	565 (54.17%)	2253 (33.82%)	<0.001§ 1* 2* 3*
	Dyslipidemia (%)	1901 (40.44%)	529 (57.69%)	650 (62.32%)	3080 (46.24%)	<0.001§ 1* 3*
	Triglycerides (mmol/L)	1.53 ± 1.29	1.98 ± 1.42	2.25 ± 2.16	1.70 ± 1.51	<0.001$ 1*** 2*** 3***
	Total cholesterol (mmol/L)	5.42 ± 1.10	5.69 ± 1.08	5.71 ± 1.22	5.50 ± 1.12	<0.001$ 1*** 3***
	LDL (mmol/L)	3.00 ± 0.98	3.33 ± 0.99	3.32 ± 1.03	3.10 ± 1.00	<0.001$ 1*** 3***
	HDL (mmol/L)	1.41 ± 0.34	1.33 ± 0.34	1.29 ± 0.32	1.38 ± 0.34	<0.001$ 1*** 2** 3***
	BMI (kg/m^2)	24.39 ± 3.84	25.94 ± 3.57	26.23 ± 3.77	24.89 ± 3.87	<0.001$ 1*** 3***
	HbA1c (%)	5.18 ± 0.64	5.59 ± 0.81	6.89 ± 1.75	5.51 ± 1.11	<0.001$ 1*** 2*** 3***
	History of viral hepatitis (%)	227 (4.83%)	46 (5.02%)	54 (5.18%)	327 (4.91%)	0.883§
	Chronic atrophic gastritis (%)	165 (3.51%)	38 (4.14%)	22 (2.11%)	225 (3.38%)	0.030§ 2*
	Use anti-inflammatory agents (%)	113 (2.40%)	40 (4.36%)	62 (5.94%)	215 (3.23%)	<0.001§ 1* 3*
	CRP (mg/L)	0.95 ± 3.56	1.59 ± 5.14	1.84 ± 4.55	1.17 ± 3.99	<0.001# 1*** 2*** 3***
≥ 65	Age (years)	71.58 ± 5.92	72.68 ± 6.20	72.50 ± 5.98	72.08 ± 6.01	<0.001$ 1** 3**
	Male (%)	569 (44.49%)	235 (43.76%)	304 (40.70%)	1108 (43.23%)	0.242§
	Urban (%)	829 (64.82%)	341 (63.50%)	478 (63.99%)	1648 (64.30%)	0.848§
	Married (%)	1034 (80.84%)	410 (76.35%)	575 (76.97%)	2019 (78.77%)	0.037§
	≥9 years of education (%)	780 (60.99%)	292 (54.38%)	393 (52.61%)	1465 (57.16%)	<0.001§ 1* 3*
	Current smoking (%)	159 (12.43%)	77 (14.34%)	90 (12.05%)	326 (12.72%)	0.434§
	Alcohol consumption (%)	165 (12.90%)	74 (13.78%)	76 (10.17%)	315 (12.29%)	0.098§
	Tea consumption (%)	313 (24.47%)	133 (24.77%)	190 (25.44%)	636 (24.81%)	0.889§
	Physical activity (%)	397 (31.04%)	135 (25.14%)	183 (24.50%)	715 (27.90%)	0.002§ 1* 3*
	Family history of cancer (%)	76 (5.94%)	23 (4.28%)	37 (4.95%)	136 (5.31%)	0.311§
	Hypertension (%)	652 (50.98%)	332 (61.82%)	540 (72.29%)	1524 (59.46%)	<0.001§ 1* 2* 3*
	Dyslipidemia (%)	603 (47.15%)	297 (55.31%)	446 (59.71%)	1346 (52.52%)	<0.001§ 1* 3*
	Triglycerides (mmol/L)	1.47 ± 0.81	1.75 ± 1.17	1.91 ± 1.35	1.66 ± 1.09	<0.001$ 1*** 2* 3***
	Total cholesterol (mmol/L)	5.59 ± 1.11	5.64 ± 1.12	5.70 ± 1.12	5.63 ± 1.11	0.112$
	LDL (mmol/L)	3.15 ± 1.02	3.20 ± 1.03	3.26 ± 1.04	3.19 ± 1.03	0.058$
	HDL (mmol/L)	1.39 ± 0.34	1.33 ± 0.32	1.31 ± 0.31	1.35 ± 0.33	<0.001$ 1** 3***
	BMI (kg/m^2)	24.87 ± 3.29	25.41 ± 3.72	26.00 ± 3.49	25.31 ± 3.47	<0.001$ 1** 2** 3***
	HbA1c (%)	5.33 ± 0.66	5.65 ± 0.80	6.75 ± 1.53	5.81 ± 1.19	<0.001$ 1*** 2*** 3***
	History of viral hepatitis infection (%)	45 (3.52%)	15 (2.79%)	25 (3.35%)	85 (3.32%)	0.732§
	Chronic atrophic gastritis (%)	68 (5.32%)	23 (4.28%)	22 (2.95%)	113 (4.41%)	0.042§ 3*
	Use anti-inflammatory agents (%)	87 (6.80%)	38 (7.08%)	67 (8.97%)	192 (7.49%)	0.186§
	CRP (mg/L)	1.31 ± 3.95	1.82 ± 5.79	2.15 ± 5.71	1.66 ± 4.94	0.001# 1*** 3***
Total	Age (years)	55.25 ± 13.35	61.71 ± 11.02	62.92 ± 10.37	57.76 ± 12.93	<0.001$ 1*** 2* 3***
	Male (%)	2107 (35.23%)	544 (37.41%)	744 (41.56%)	3395 (36.81%)	<0.001§ 2* 3*
	Urban (%)	3647 (60.99%)	837 (57.57%)	1061 (59.27%)	5545 (60.11%)	0.041§
	Married (%)	5249 (87.78%)	1277 (87.83%)	1551 (86.65%)	8077 (87.57%)	0.424§
	≥9 years of education (%)	4966 (83.04%)	1054 (72.49%)	1251 (69.89%)	7271 (78.83%)	<0.001§ 1* 3*
	Current smoking (%)	976 (16.32%)	223 (15.34%)	345 (19.27%)	1544 (16.74%)	0.004§ 2* 3*
	Alcohol consumption (%)	689 (11.52%)	200 (13.76%)	231 (12.91%)	1120 (12.14%)	0.035§
	Tea consumption (%)	1633 (27.31%)	398 (27.37%)	551 (30.78%)	2582 (27.99%)	0.014§ 3*
	Physical activity (%)	1591 (26.61%)	345 (23.73%)	441 (24.64%)	2377 (25.77%)	0.038§
	Family history of cancer (%)	362 (6.05%)	100 (6.88%)	95 (5.31%)	557 (6.04%)	0.174§
	Hypertension (%)	1897 (31.72%)	775 (53.30%)	1105 (61.73%)	3777 (40.95%)	<0.001§ 1* 2* 3*
	Dyslipidemia (%)	2504 (41.87%)	826 (56.81%)	1096 (61.23%)	4426 (47.98%)	<0.001§ 1* 2* 3*
	Triglycerides (mmol/L)	1.52 ± 1.21	1.90 ± 1.34	2.11 ± 1.87	1.69 ± 1.40	<0.001$ 1*** 2*** 3***
	Total cholesterol (mmol/L)	5.45 ± 1.10	5.67 ± 1.09	5.70 ± 1.18	5.54 ± 1.12	<0.001$ 1*** 3***
	LDL (mmol/L)	3.03 ± 0.99	3.28 ± 1.01	3.30 ± 1.04	3.12 ± 1.01	<0.001$ 1*** 3***
	HDL (mmol/L)	1.40 ± 0.34	1.33 ± 0.33	1.29 ± 0.31	1.37 ± 0.34	<0.001$ 1*** 2** 3***
	BMI (kg/m^2)	24.49 ± 3.73	25.75 ± 3.63	26.14 ± 3.66	25.01 ± 3.77	<0.001$ 1*** 2** 3***

(Continued)

TABLE 1 | Continued

Age group (Years old)	Characteristics	Glycemic status			Total	*p*
		NGT	Prediabetes	Diabetes		
	HbA1c (%)	5.22 ± 0.65	5.61 ± 0.80	6.83 ± 1.66	5.59 ± 1.14	<0.001$ 1*** 2*** 3***
	History of viral hepatitis infection (%)	272 (4.55%)	61 (4.20%)	79 (4.41%)	412 (4.47%)	0.837§
	Chronic atrophic gastritis (%)	233 (3.90%)	61 (4.20%)	44 (2.46%)	338 (3.66%)	0.009§ 2* 3*
	Use anti-inflammatory agents (%)	200 (3.34%)	78 (5.36%)	129 (7.21%)	407 (4.41%)	<0.001§ 1* 3*
	CRP (mg/L)	1.02 ± 3.65	1.67 ± 5.39	1.97 ± 5.07	1.31 ± 4.28	<0.001# 1*** 2*** 3***

 # Comparison performed using Kruskal-Wallis test. § Comparison performed using Chi-square test. $ Comparison performed using one-way ANOVA test.
Data are n (%) or mean ± SD.
*P value indicates the statistical result for the Kruskal-Wallis, chi-square or one-way ANOVA test. The results of Post hoc multiple comparisons (Bonferroni) were indicated as follows: 1, NGT versus prediabetes; 2, prediabetes versus diabetes; 3, NGT versus diabetes. *P < 0.05; **P < 0.01; ***P < 0.001.*
NGT, normal glucose tolerance; LDL, low-density lipoprotein; HDL, high-density lipoprotein; HbA1c, glycated hemoglobin A_{1c}; BMI, body mass index; CRP, C-reactive protein.

medications rather than in those receiving regular anti-diabetic medications including insulin, euglycemic agents, sulfonylureas, biguanides, thiazolidinediones, α-glycosidase inhibitors, and Chinese traditional anti-diabetic medicine, or in diabetes patients diagnosed within 5 years rather than in those diagnosed longer than 5 years in whole participants. This effect was only evident in young adults rather than in elderly adults (**Table 3**).

Association of the Incidences of Site-Specific Cancers With Baseline Glycemic Status

The association of site-specific cancers with baseline glycemic status in the whole population was first evaluated by the Cox regression analysis, adjusted for age and sex. Female breast cancer, and kidney cancer were significantly associated with glucose intolerance (prediabetes+diabetes) (**Supplementary Table 2** and **Figure 1A**). Women with glucose intolerance had higher incidences of female breast cancer and pancreatic cancer (**Figure 1B**). Men with glucose intolerance had a higher incidence of kidney cancer (**Figure 1C**). Stratification analysis indicated that in the whole population, participants with prediabetes had increased risks of stomach cancer and kidney cancer, while participants with diabetes had increased risks of female breast cancer and kidney cancer (**Supplementary Table 3**). In young adults, glucose intolerance was significantly associated with increased risks of stomach cancer, colorectal cancer, and kidney cancer in the Cox regression analysis, adjusted for age and sex (**Table 4**). Participants with prediabetes had increased risks of stomach cancer, kidney cancer and pancreatic cancer. Participants with diabetes had increased risks of stomach cancer, colorectal cancer and kidney cancer in this population (**Supplementary Table 4**).

DISCUSSION

In this community-based prospective cohort study, diabetes and prediabetes were identified to be independently associated with increased risks of total cancer and site-specific cancers such as stomach cancer, colorectal cancer, and kidney cancer in young adults (<65 years). Anti-diabetic medications reduced the risk of cancer caused by diabetes. The outcomes of this study may reflect the current risk factors of cancer in young adults. The study population was randomly recruited from urban and rural communities in Pudong New Area, the only district with urban and rural residents in Shanghai (21). Pudong New Area has about 5 million permanent residents with diverse socioeconomic status, which is highly representative for other populations. The permanent residents possessing Shanghai household registration were recruited in this study, just because this population could be eligible to be followed-up and information of cancer occurrence could be verified by data linkage with the cancer registration and management system. This does not affect the representativeness. Thus, the findings of this study can be generalized to other populations both within and outside China.

In this study, we demonstrated that glucose intolerance was significantly associated with an increased risk of total cancer especially for stomach cancer, colorectal cancer and kidney cancer in young adults. These effects were independent of other risk factors. In elderly adults, glucose intolerance was not independently associated with increased risk of total cancer. Cancer occurs more often in aged adults than in younger ones. The effect of glucose intolerance on cancer might be covered by the overwhelming effects of age and current smoking in aged adults. Our data support that the risk factors of all cancer have shifted from the pollution and chronic infections in the past decades to metabolic syndrome at the present (23). Metabolic syndrome, which is often caused by overconsumption of calories and fat and lack of physical activity, is prevalent worldwide. An important study has demonstrated that HRs for all-site and site-specific cancers are particularly elevated during the first year following diabetes diagnosis (6). Diabetes is associated with higher risk of colorectal adenomas, a precancerous lesion of colorectal cancer, in adults aged 40-49 years (24). A cross-sectional study using data from the 2001-2014 National Health and Nutrition Examination Survey has shown that individuals <65 years have higher odds of colorectal cancer when also diagnosed with diabetes (25). It has been demonstrated that diabetes patients aged ≤50 or 55 years have a greater risk of all cancers, digestive cancers, and urinary cancers (26, 27). These findings suggest that glucose

TABLE 2 | Cox regression analysis of factors significantly affected cancer incidence in cohort participants, stratified by age group.

Age at the baseline	Variable	Persons at risk	Incident casess	Person-years	Incidence (1/1000)	Univariate analysis		Multivariate Analysis*	
						HR (95% CI)	*p*	HR (95% CI)	*p*
15-64 years	Glycemic status								
	NGT	4701	169	35880	4.71	ref.		ref.	
	Prediabetes	917	57	6892	8.27	1.76 (1.30-2.37)	<0.001	1.49 (1.09-2.02)	0.012
	Diabetes	1043	67	7808	8.58	1.82 (1.37-2.42)	<0.001	1.51 (1.12-2.02)	0.006
	Age								
	15-24	155	2	1203	1.66	ref.		ref.	
	25-34	416	4	3223	1.24	0.74 (0.14-4.06)	0.733	0.73 (0.13-3.99)	0.717
	35-44	741	11	5731	1.92	1.15 (0.26-5.20)	0.853	1.09 (0.24-4.94)	0.908
	45-54	1746	66	13303	4.96	2.98 (0.73-12.15)	0.129	2.69 (0.66-11.04)	0.168
	55-64	3603	210	27120	7.74	4.65 (1.16-18.73)	0.030	4.13 (1.02-16.75)	0.047
	Sex								
	Male	2287	97	17351	5.59	ref.		–	–
	Female	4374	196	33230	5.90	1.05 (0.83-1.34)	0.672	–	–
	Area								
	Urban	3897	170	29686	5.73	ref.		–	–
	Rural	2764	123	20895	5.89	0.98 (0.78-1.23)	0.850	–	–
	Marriage status								
	Married	6058	270	45987	5.87	ref.		–	–
	Other	603	23	4594	5.01	0.85 (0.56-1.30)	0.462	–	–
	Years of education								
	≥9	5806	251	44094	5.69	ref.		–	–
	<9	855	42	6486	6.48	1.14 (0.82-1.58)	0.434	–	–
	BMI	6661	293	50581	5.79	1.03 (1.00-1.06)	0.033	1.02 (0.98-1.05)	0.310
	Current smoking								
	No	5443	232	41383	5.61	ref.		–	–
	Yes	1218	61	9198	6.63	1.18 (0.89-1.57)	0.241	–	–
	Alcohol consumption								
	No	5856	252	44482	5.67	ref.		–	–
	Yes	805	41	6098	6.72	1.19 (0.85-1.65)	0.308	–	–
	Tea consumption								
	No	4715	198	35846	5.52	ref.		–	–
	Yes	1946	95	14734	6.45	1.17 (0.91-1.49)	0.214	–	–
	Physical activity								
	No	4999	224	37936	5.90	ref.		–	–
	Yes	1662	69	12644	5.46	0.92 (0.71-1.21)	0.569	–	–
	Family history of cancer								
	No	6264	270	47574	5.68	ref.		–	–
	Yes	397	23	3006	7.65	1.34 (0.88-2.06)	0.172	–	–
	Hypertension								
	No	4408	177	33575	5.27	ref.		ref.	
	Yes	2253	116	17006	6.82	1.29 (1.02-1.63)	0.031	0.90 (0.70-1.15)	0.393
	Dyslipidemia								
	No	3581	143	27257	5.25	ref.		–	–
	Yes	3080	150	23324	6.43	1.23 (0.98-1.54)	0.081	–	–
	Viral hepatitis								
	No	6334	275	48119	5.72	ref.		–	–

(Continued)

TABLE 2 | Continued

Age at the baseline	Variable	Persons at risk	Incident casess	Person-years	Incidence (1/1000)	Univariate analysis		Multivariate Analysis*	
						HR (95% CI)	*p*	HR (95% CI)	*p*
	Yes	327	18	2462	7.31	1.28 (0.79-2.06)	0.313	–	–
	Chronic atrophic gastritis								
	No	6436	282	48884	5.77	ref.		–	–
	Yes	225	11	1697	6.48	1.12 (0.62-2.05)	0.704	–	–
	HbA1c	6661	293	50581	5.79	1.09 (0.99-1.19)	0.068	–	–
	Use anti-inflammatory agents								
	No	6446	282	48966	5.76	ref.		–	–
	Yes	215	11	1615	6.81	1.18 (0.65-2.16)	0.584	–	–
	CRP	6661	293	50581	5.79	1.02 (1.00-1.04)	0.014	1.02 (1.00-1.03)	0.072
≥65 years	Glycemic status								
	NGT	1279	100	9303	10.75	ref.		–	–
	Prediabetes	537	42	3864	10.87	1.01 (0.70-1.45)	0.954	–	–
	Diabetes	747	67	5218	12.84	1.20 (0.88-1.63)	0.257	–	–
	Age								
	65-74	1704	114	12565	9.07	ref.		ref.	
	75-84	775	85	5332	15.94	1.76 (1.33-2.33)	<0.001	1.60 (1.18-2.16)	0.002
	≥85	84	10	488	20.48	2.28 (1.20-4.36)	0.012	1.94 (0.99-3.79)	0.054
	Sex								
	Male	1108	106	7796	13.60	ref.		ref.	
	Female	1455	103	10589	9.73	0.71 (0.54-0.94)	0.015	0.71 (0.51-1.00)	0.048
	Area								
	Urban	1648	125	11855	10.54	ref.		–	–
	Rural	915	84	6531	12.86	0.82 (0.62-1.08)	0.158	–	–
	Marriage status								
	Married	2019	155	14684	10.56	ref.		ref.	
	Other	544	44	3702	11.89	1.39 (1.02-1.89)	0.039	1.26 (0.89-1.78)	0.184
	Years of education								
	≥ 9	1465	101	10676	9.46	ref.		ref.	
	< 9	1098	108	7709	14.01	1.48 (1.13-1.95)	0.004	1.44 (1.06-1.95)	0.020
	BMI	2563	209	18386	11.37	1.03 (0.99-1.07)	0.197	–	–
	Current smoking								
	No	2237	164	16103	10.18	ref.		ref.	
	Yes	326	45	2282	19.72	1.94 (1.39-2.69)	<0.001	1.88 (1.29-2.73)	0.001
	Alcohol consumption								
	No	2248	175	16171	10.82	ref.		–	–
	Yes	315	34	2214	15.36	1.42 (0.98-2.05)	0.062	–	–
	Tea consumption								
	No	1927	164	13794	11.89	ref.		–	–
	Yes	636	45	4591	9.80	0.82 (0.59-1.14)	0.247	–	–
	Physical activity								
	No	1848	151	13179	11.46	ref.		–	–
	Yes	715	58	5206	11.14	0.97 (0.72-1.32)	0.856	–	–
	Family history of cancer								
	No	2433	200	17430	11.47	ref.		–	–
	Yes	130	9	956	9.42	0.82 (0.42-1.59)	0.552	–	–
	Hypertension								

(Continued)

TABLE 2 | Continued

Age at the baseline	Variable	Persons at risk	Incident casess	Person-years	Incidence (1/1000)	Univariate analysis		Multivariate Analysis*	
						HR (95% CI)	*p*	HR (95% CI)	*p*
	No	1039	83	7490	11.08	ref.		–	–
	Yes	1524	126	10896	11.56	1.04 (0.79-1.38)	0.763	–	–
	Dyslipidemia								
	No	1217	109	8659	12.59	ref.		–	–
	Yes	1346	100	9727	10.28	0.82 (0.62-1.07)	0.140	–	–
	Viral hepatitis								
	No	2478	199	17790	11.19	ref.		–	–
	Yes	85	10	596	16.79	1.50 (0.80-2.83)	0.209	–	–
	Chronic atrophic gastritis								
	No	2450	198	17572	11.27	ref.		–	–
	Yes	113	11	814	13.52	1.20 (0.65-2.20)	0.559	–	–
	HbA1c	2563	209	18386	11.37	1.01 (0.90-1.13)	0.882	–	–
	Use anti-inflammatory agents								
	No	2371	194	17033	11.39	ref.		–	–
	Yes	192	15	1353	11.09	0.97 (0.58-1.65)	0.920	–	–
	CRP	2563	209	18386	11.37	1.00 (0.98-1.03)	0.807	–	–
Total	Glycemic status								
	NGT	5980	269	45184	5.95	ref.		ref.	
	Prediabetes	1454	99	10756	9.20	1.55 (1.23-1.95)	<0.001	1.24 (0.98-1.58)	0.072
	Diabetes	1790	134	13027	10.29	1.73 (1.41-2.13)	<0.001	1.42 (1.10-1.82)	0.006
	Age								
	15-24	155	2	1203	1.66	ref.		ref.	
	25-34	416	4	3223	1.24	0.75 (0.14-4.07)	0.734	0.77 (0.14-4.26)	0.766
	35-44	741	11	5731	1.92	1.15 (0.26-5.20)	0.853	1.18 (0.26-5.42)	0.832
	45-54	1746	66	13303	4.96	2.98 (0.73-12.15)	0.129	2.95 (0.71-12.3)	0.137
	55-64	3603	210	27120	7.74	4.65 (1.16-18.72)	0.030	4.60 (1.12-18.92)	0.035
	65-74	1704	114	12565	9.07	5.46 (1.35-22.09)	0.017	5.18 (1.25-21.48)	0.023
	75-84	775	85	5332	15.94	9.64 (2.37-39.17)	0.002	8.78 (2.11-36.46)	0.003
	≥85	84	10	488	20.48	12.60 (2.76-57.51)	0.001	11.75 (2.53-54.57)	0.002
	Sex								
	Male	3395	203	25147	8.07	ref.		–	–
	Female	5829	299	43819	6.82	0.84 (0.71-1.01)	0.062	–	–
	Area								
	Urban	5545	295	41541	7.10	ref.		–	–
	Rural	3679	207	27426	7.55	0.94 (0.79-1.13)	0.523	–	–
	Marriage status								
	Married	8077	425	60671	7.01	ref.		ref.	
	Other	1147	77	8296	9.28	1.33 (1.04-1.69)	0.022	1.16 (0.89-1.51)	0.283
	Years of education								
	≥9	7271	352	54771	6.43	ref.		ref.	
	<9	1953	150	14196	10.57	1.65 (1.36-2.00)	<0.001	1.11 (0.90-1.38)	0.339
	BMI	9224	502	68966	7.28	1.03 (1.01-1.06)	0.004	1.03 (1.00-1.05)	0.056
	Current smoking								
	No	7680	396	57486	6.89	ref.		ref.	
	Yes	1544	106	11480	9.23	1.34 (1.08-1.66)	0.007	1.44 (1.14-1.83)	0.002
	Alcohol consumption								

(Continued)

TABLE 2 | Continued

Age at the baseline	Variable	Persons at risk	Incident casess	Person-years	Incidence (1/1000)	Univariate analysis		Multivariate Analysis*	
						HR (95% CI)	*p*	HR (95% CI)	*p*
	No	8104	427	60654	7.04	ref.		ref.	
	Yes	1120	75	8313	9.02	1.28 (1.00-1.64)	0.047	1.10 (0.84-1.43)	0.493
	Tea consumption								
	No	6642	362	49641	7.29	ref.		–	–
	Yes	2582	140	19326	7.24	0.99 (0.82-1.21)	0.944	–	–
	Physical activity								
	No	6847	375	51116	7.34	ref.		–	–
	Yes	2377	127	17850	7.11	0.97 (0.79-1.19)	0.765	–	–
	Family history of cancer								
	No	8697	470	65004	7.23	ref.		–	–
	Yes	527	32	3962	8.08	1.11 (0.78-1.59)	0.558	–	–
	Hypertension								
	No	5447	260	41065	6.33	ref.		ref.	
	Yes	3777	242	27901	8.67	1.37 (1.15-1.63)	<0.001	0.92 (0.76-1.11)	0.363
	Dyslipidemia								
	No	4798	252	35916	7.02	ref.		–	–
	Yes	4426	250	33050	7.56	1.08 (0.90-1.28)	0.404	–	–
	Viral hepatitis								
	No	8812	474	65908	7.19	ref.		–	–
	Yes	412	28	3058	9.16	1.27 (0.87-1.86)	0.216	–	–
	Chronic atrophic gastritis								
	No	8886	480	66455	7.22	ref.		–	–
	Yes	338	22	2511	8.76	1.21 (0.79-1.86)	0.377	–	–
	HbA1c	9224	502	68966	7.28	1.09 (1.01-1.16)	0.020	0.94 (0.86-1.03)	0.201
	Use anti-inflammatory agents								
	No	8817	476	65999	7.21	ref.		–	–
	Yes	407	26	2967	8.76	1.22 (0.82-1.80)	0.332	–	–
	CRP	9224	502	68966	7.28	1.02 (1.00-1.03)	0.021	1.01 (0.99-1.02)	0.243

**Only included significant covariates in univariate analysis.*

NGT, normal glucose tolerance; BMI, body mass index; HbA1c, glycated hemoglobin A_{1c}; CRP, C-reactive protein.

TABLE 3 | Effects of glucose intolerance on cancer incidence in the study participants.

Age at the baseline	Variable	Persons at risk	Incident casess	Person-years	Incidence (1/1000)	Univariate analysis		Multivariate Analysis*	
						HR (95% CI)	*p*	HR (95% CI)	*p*
15-64 years	NGT	4701	169	35880	4.71	ref.		ref.	
	Category of prediabetes								
	IFG	323	22	2433	9.04	1.92 (1.23-3.00)	0.004	1.67 (1.07-2.61)	0.023
	IGT	436	27	3265	8.27	1.76 (1.17-2.64)	0.007	1.51 (1.00-2.27)	0.048
	IFG+IGT	158	8	1194	6.70	1.42 (0.70-2.89)	0.331	1.22 (0.60-2.48)	0.581
	Category of diabetes								
	Diagnosed previously	573	31	4308	7.20	1.53 (1.04-2.24)	0.030	1.26 (0.85-1.86)	0.248
	Screen detected at the baseline	470	36	3501	10.28	2.18 (1.52-3.13)	<0.001	1.86 (1.29-2.70)	0.001
	Diabetes patients with anti-diabetic medications								
	Yes	473	29	3538	8.20	1.74 (1.18-2.58)	0.006	1.43 (0.95-2.13)	0.083
	No	570	38	4271	8.90	1.89 (1.33-2.69)	<0.001	1.61 (1.12-2.30)	0.010
	Duration since first diagnose of diabetes								
	<5 years	749	52	5603	9.28	1.97 (1.45-2.69)	<0.001	1.68 (1.22-2.31)	0.002
	≥5 years	294	15	2206	6.80	1.45 (0.85-2.45)	0.171	1.16 (0.68-1.98)	0.590
≥65 years	NGT	1279	100	9303	10.75	ref.		ref.	
	Category of prediabetes								
	IFG	141	14	1027	13.64	1.27 (0.72-2.22)	0.406	–	–
	IGT	308	18	2223	8.10	0.75 (0.46-1.24)	0.268	–	–
	IFG+IGT	88	10	614	16.28	1.51 (0.79-2.90)	0.211	–	–
	Category of diabetes								
	Diagnosed previously	459	32	3231	9.90	0.92 (0.62-1.37)	0.689	0.89 (0.60-1.33)	0.577
	Screen detected at the baseline	288	35	1987	17.62	1.64 (1.12-2.41)	0.012	1.52 (1.03-2.24)	0.033
	Diabetes patients with anti-diabetic medications								
	Yes	378	28	2635	10.63	0.99 (0.65-1.50)	0.960	–	–
	No	369	39	2583	15.10	1.41 (0.97-2.04)	0.071	–	–
	Duration since first diagnose of diabetes								
	<5 years	435	41	3079	13.32	1.24 (0.86-1.78)	0.246	–	–
	≥5 years	312	26	2139	12.15	1.13 (0.74-1.74)	0.573	–	–
Total	NGT	5980	269	45184	5.95	ref.		ref.	
	Category of prediabetes								
	IFG	464	36	3460	10.41	1.75 (1.23-2.48)	0.002	1.50 (1.06-2.13)	0.023
	IGT	744	45	5488	8.20	1.38 (1.00-1.89)	0.047	1.04 (0.75-1.43)	0.810
	IFG+IGT	246	18	1809	9.95	1.67 (1.04-2.69)	0.035	1.32 (0.82-2.14)	0.252
	Category of diabetes								
	Diagnosed previously	1032	63	7539	8.36	1.41 (1.07-1.85)	0.015	1.04 (0.79-1.38)	0.779
	Screen detected at the baseline	758	71	5488	12.94	2.18 (1.68-2.83)	<0.001	1.69 (1.29-2.21)	<0.001

(Continued)

TABLE 3 | Continued

Age at the baseline	Variable	Persons at risk	Incident cases	Person-years	Incidence (1/1000)	Univariate analysis		Multivariate Analysis*	
						HR (95% CI)	*p*	HR (95% CI)	*p*
	Diabetes patients with anti-diabetic medications								
	Yes	851	57	6173	9.23	1.55 (1.17-2.07)	0.002	1.15 (0.86-1.54)	0.344
	No	939	77	6854	11.23	1.89 (1.47-2.44)	<0.001	1.45 (1.12-1.89)	0.005
	Duration since first diagnosis of diabetes								
	<5 years	1184	93	8681	10.71	1.8 (1.42-2.28)	<0.001	1.41 (1.11-1.80)	0.005
	≥5 years	606	41	4345	9.44	1.59 (1.14-2.21)	0.006	1.12 (0.80-1.57)	0.520

Only included significant covariates shown in **Table 2 in univariate Cox regression analysis.*
NGT, normal glucose tolerance; IFG, impaired fasting glucose; IGT, impaired glucose tolerance.

intolerance may facilitate cancer development in young adults, making this population with glucose intolerance a target population for cancer screening and interventions. Since the incidence of diabetes is increasing dramatically in the younger generation (28, 29), our finding is of public health importance in monitoring all cancer in young adults who have glucose intolerance. Public health actions including encouraging physical activity and restricting energy intake to reduce the prevalent and incident glucose intolerance should be important in reducing cancer risk in young adults.

In this study, we demonstrated that anti-diabetic medications were significantly associated with a decreased risk of all cancer in young adults with diabetes. Interestingly, diabetes patients who were diagnosed previously and diagnosed 5 years or longer did not have an increased risk of all cancer, whereas diabetes patients diagnosed at the baseline survey and within 5 years had an increased risk of cancer (**Table 3**). This is possibly because long-term anti-diabetic medications have been widely applied in the study subjects who were diagnosed as diabetes 5 years ago. Anti-diabetic medications had been covered by basic medical insurance for decades in Shanghai, China. Our result is quite consistent with another cohort study carried out in Italy (30). Lifelong use of anti-diabetics is protective for all cancer in patients with diabetes. We postulate that increase in physical activity and dietary continence should be protective for all cancer in young adults with prediabetes.

The mechanism by which glucose intolerance is associated with an increased risk of all cancer remains largely unknown. Here, we demonstrated that glucose intolerance was associated with increased risks of stomach cancer, colorectal cancer, kidney cancer, and pancreatic cancer in young adults, and female breast cancer, stomach cancer, and kidney cancer in the whole population. Data from the China Kadoorie Biobank Study have shown that glucose intolerance was associated with increased risks of certain site-specific cancers including female breast cancer, liver cancer, pancreatic cancer, and colorectal cancer (6, 31). The findings in Chinese population are mostly consistent with that in Western population (6, 25, 30). The association of glucose intolerance with stomach cancer is not evident in a cohort study in the Northern Swedish population (12), possibly because of the differences in the susceptibility of gastric cancer between study populations. Although each site-specific cancer has its own risk factors, they share a common risk factor: chronic inflammation. Metformin that was proven to inhibit cancer cell growth and modulate cancer microenvironment has been demonstrated to have potent inflammation-inhibitory effects (32). In this study, CRP, a well-established marker of systemic inflammation in metabolic syndrome (33), tend to be identified as an independent risk factor of cancer in young adults. Elevated CRP has been associated with an increased risk of diabetes in middle-aged and elderly Chinese (34). Chronic inflammation related to glucose intolerance might play an essential role in carcinogenesis. Insulin is a potent growth factor that promotes cell proliferation and carcinogenesis directly and/or through insulin-like growth factor 1 (IGF-1).

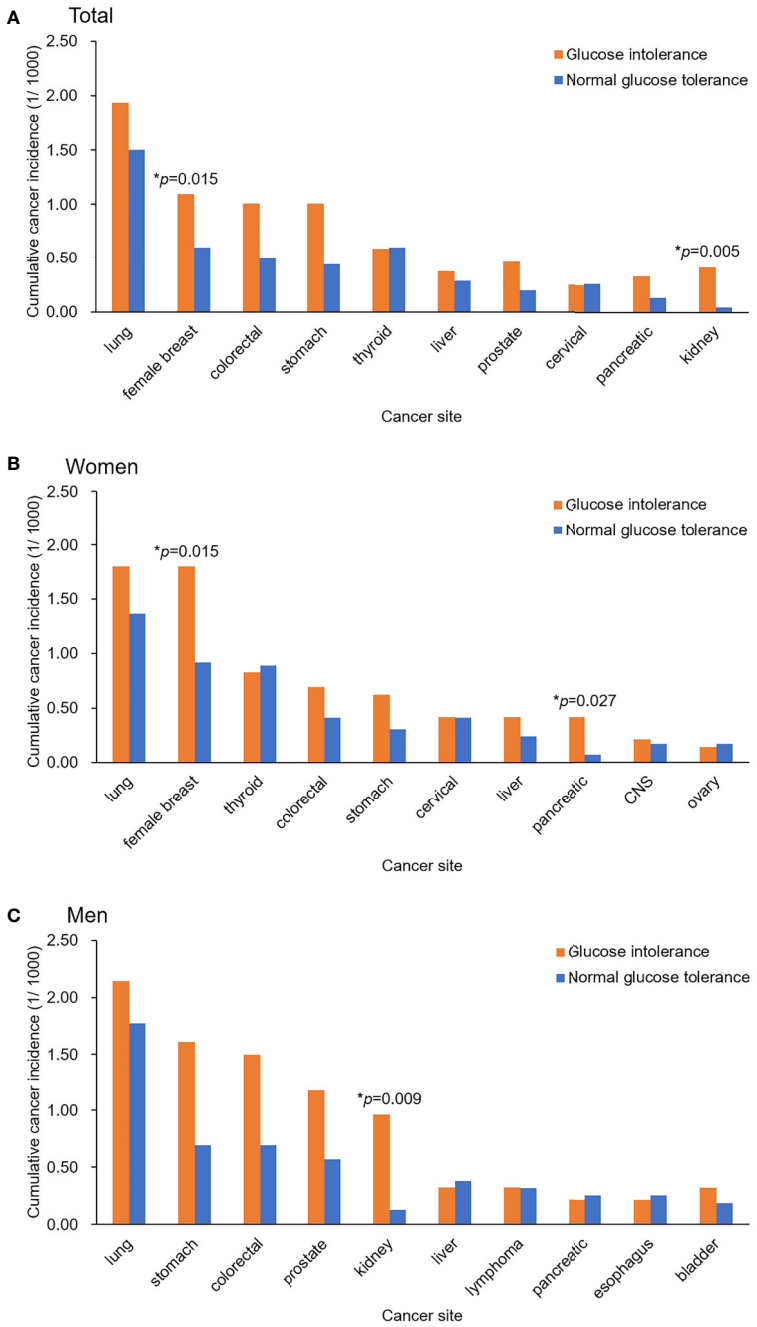


FIGURE 1 | Cumulative incidence rates of the top 10 site-specific cancers during the follow-up among the study participants with different baseline glycemic status. **(A)** Total participants, **(B)** Women, **(C)** Men. Differences in the cumulative incidence rates were tested using a Cox proportional hazards model, adjusted for age and sex.

Hyperinsulinemia leads to an increase in the bioactivity of IGF-1 by inhibiting IGF binding protein-1 (35). Apart from directly promotes cancer progression, hyperglycemia increases the levels of insulin/IGF-1 and inflammatory cytokines in circulation (36). Metabolic disorder was associated with increased risk of liver cancer (37). In this study, the association of glucose intolerance with liver cancer was not evident possibly due to few cases of liver cancer diagnosed in this cohort. Even though, glycemic control is important for cancer prevention in young adults.

TABLE 4 | The Cox regression analysis of the association of site-specific cancer with the baseline glycemic status in the adults < 65 years, adjusted for age and sex.

Glycemic status	Persons at risk	Incident cases	Person-years	Incidence (1/1000)	HR (95% CI)	*p*
Lung cancer						
NGT	4701	45	35880	1.25	ref.	
Glucose intolerance	1960	23	14700	1.56	1.04 (0.63-1.73)	0.872
Female breast cancer						
NGT	3163	23	24123	0.95	ref.	
Glucose intolerance	1211	16	9107	1.76	1.44 (0.75-2.75)	0.268
Stomach cancer						
NGT	4701	10	35880	0.28	ref.	
Glucose intolerance	1960	16	14700	1.09	3.72 (1.68-8.20)	0.001
Colorectal cancer						
NGT	4701	9	35880	0.25	ref.	
Glucose intolerance	1960	13	14700	0.88	3.51 (1.50-8.22)	0.004
Kidney cancer						
NGT	4701	2	35880	0.06	ref.	
Glucose intolerance	1960	8	14700	0.54	8.69 (1.84-40.95)	0.006
Liver cancer						
NGT	4701	5	35880	0.14	ref.	
Glucose intolerance	1960	3	14700	0.20	1.47 (0.35-6.14)	0.599
Pancreatic cancer						
NGT	4701	3	35880	0.08	ref.	
Glucose intolerance	1960	4	14700	0.27	3.27 (0.73-14.6)	0.121
Esophageal cancer						
NGT	4701	3	35880	0.08	–	–
Glucose intolerance	1960	0	14700	0.00	–	–

HR, Hazard ratio; NGT, normal glucose tolerance; Glucose intolerance, prediabetes + diabetes.

The strengths of this study include a perspective design, the high representativeness of community-based study population, holistic risk factors screening, use of standard OGTT at the baseline survey, adjustment for multiple potential confounding factors, and reliable follow-up. This study has three main implications. First, young adults with glucose intolerance are recommended to undergo appropriate cancer screenings for early diagnosis. Second, steps to prevent cancer should be taken even at pre-diabetic stage. Some forms of diabetes treatment and a reversal of obesity and prediabetes can reduce cancer risk (38). Glycemic management and lifestyle intervention are of public health significance. Third, this study provides clue to elucidate the mechanism by which glucose intolerance induces carcinogenesis.

Limitations

This study has several limitations. First, risk factors for cancer were not all included in the baseline survey, such as dietary habit, stress, and social factors, resulting loss of data. Second, the follow-up period was relatively short, resulting in small number of end-point events that weakened the statistical power. Third, information of the income was incomplete because of personal privacy. The education levels might serve as an alternative in this analysis. Fourth, small number of end-point events makes it difficult to investigate the associations of each type of anti-diabetic medicines with the risk of cancer.

CONCLUSIONS

In this community-based prospective cohort study, diabetes and prediabetes were independently associated with increased risks of total cancer and site-specific cancers such as stomach cancer, colorectal cancer, and kidney cancer in young adults. Regular monitoring of plasma glucose level could assist to identify individuals with an increased risk of cancer. Lifestyle interventions and anti-diabetic medications to prevent and treat prediabetes and diabetes are important in cancer prophylaxis in young adults.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the Center for Disease Control and Prevention of the Pudong New Area, Shanghai, China. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization: JK, TL, XR, and GC. Data curation: JK and GC. Funding acquisition: GC. Investigation: JK, TL, XL, KW, XR, WL, HQ, XT, XW, YD, and GC. Methodology: JK, TL, XL, KW, XR, HQ, XT, XW, XC, and GC. Project administration: TL, XR, and GC. Supervision: GC. Validation: XR, HQ, and XT.

Visualization: JK and GC. Writing - original draft and revising: GC.

FUNDING

This work was supported by National Natural Scientific Foundation of China grant (grant number: 81520108021, 81673250) to GC.

REFERENCES

1. Wang S, Du X, Han X, Yang F, Zhao J, Li H, et al. Influence of Socioeconomic Events on Cause-Specific Mortality in Urban Shanghai, China, From 1974 to 2015: A Population-Based Longitudinal Study. *CMAJ* (2018) 190(39):E1153–61. doi: 10.1503/cmaj.180272
2. Heer EV, Harper AS, Sung H, Jemal A, Fidler-Benaoudia MM. Emerging Cancer Incidence Trends in Canada: The Growing Burden of Young Adult Cancers. *Cancer* (2020) 126(20):4553–62. doi: 10.1002/cncr.33050
3. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, et al. Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J Natl Cancer Inst* (2017) 109(8):djw322. doi: 10.1093/jnci/djw322
4. 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):e137–47. doi: 10.1016/S2468-2667(18)30267-6
5. Yu M, Zhan X, Yang Z, Huang Y. Measuring the Global, Regional, and National Burden of Type 2 Diabetes and the Attributable Risk Factors in All 194 Countries. *J Diabetes* (2021) 13(8):613–39. doi: 10.1111/1753-0407.13159
6. Dankner R, Boffetta P, Balicer RD, Boker LK, Sadeh M, Berlin A, et al. Time-Dependent Risk of Cancer After a Diabetes Diagnosis in a Cohort of 2.3 Million Adults. *Am J Epidemiol* (2016) 183(12):1098–106. doi: 10.1093/aje/kwv290
7. Pan XF, He M, Yu C, Lv J, Guo Y, Bian Z, et al. Type 2 Diabetes and Risk of Incident Cancer in China: A Prospective Study Among 0.5 Million Chinese Adults. *Am J Epidemiol* (2018) 187(7):1380–91. doi: 10.1093/aje/kwx376
8. Reis JP, Allen NB, Bancks MP, Carr JJ, Lewis CE, Lima JA, et al. Duration of Diabetes and Prediabetes During Adulthood and Subclinical Atherosclerosis and Cardiac Dysfunction in Middle Age: The CARDIA Study. *Diabetes Care* (2018) 41(4):731–8. doi: 10.2337/dc17-2233
9. Boursi B, Finkelman B, Giantonio BJ, Haynes K, Rustgi AK, Rhim AD, et al. A Clinical Prediction Model to Assess Risk for Pancreatic Cancer Among Patients With Prediabetes. *Eur J Gastroenterol Hepatol* (2021). doi: 10.1097/MEG.0000000000002052
10. Onitilo AA, Stankowski RV, Berg RL, Engel JM, Glurich I, Williams GM, et al. Breast Cancer Incidence Before and After Diagnosis of Type 2 Diabetes Mellitus in Women: Increased Risk in the Prediabetes Phase. *Eur J Cancer Prev* (2014) 23(2):76–83. doi: 10.1097/CEJ.0b013e32836162aa
11. Goto A, Yamaji T, Sawada N, Momozawa Y, Kamatani Y, Kubo M, et al. Diabetes and Cancer Risk: A Mendelian Randomization Study. *Int J Cancer* (2020) 146(3):712–9. doi: 10.1002/ijc.32310
12. Zheng J, Rutegård M, Santoni G, Wallner B, Johansson I, Sund M, et al. Prediabetes and Diabetes in Relation to Risk of Gastric Adenocarcinoma. *Br J Cancer* (2019) 120(12):1147–52. doi: 10.1038/s41416-019-0470-1
13. Falk RS, Tretli S, Paulsen JE, Sandvik L, Erikssen J, Heir T. Response to Intravenous Glucose-Tolerance Test and Risk of Cancer: A Long-Term Prospective Cohort Study. *EBioMedicine* (2017) 21:117–22. doi: 10.1016/j.ebiom.2017.06.018
14. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-Reactive Protein, Interleukin 6, and Risk of Developing Type 2 Diabetes Mellitus. *JAMA* (2001) 286(3):327–34. doi: 10.1001/jama.286.3.327
15. Xia B, He Q, Pan Y, Gao F, Liu A, Tang Y, et al. Metabolic Syndrome and Risk of Pancreatic Cancer: A Population-Based Prospective Cohort Study. *Int J Cancer* (2020) 147(12):3384–93. doi: 10.1002/ijc.33172
16. Park YM, Bookwalter DB, O'Brien KM, Jackson CL, Weinberg CR, Sandler DP. A Prospective Study of Type 2 Diabetes, Metformin Use, and Risk of Breast Cancer. *Ann Oncol* (2021) 32(3):351–9. doi: 10.1016/j.annonc.2020.12.008
17. Di Matteo S, Nevi L, Overi D, Landolina N, Faccioli J, Giulitti F, et al. Metformin Exerts Anti-Cancerogenic Effects and Reverses Epithelial-to-Mesenchymal Transition Trait in Primary Human Intrahepatic Cholangiocarcinoma Cells. *Sci Rep* (2021) 11(1):2557. doi: 10.1038/s41598-021-81172-0
18. Kurelac I, Umesh Ganesh N, Iorio M, Porcelli AM, Gasparre G. The Multifaceted Effects of Metformin on Tumor Microenvironment. *Semin Cell Dev Biol* (2020) 98:90–7. doi: 10.1016/j.semcdb.2019.05.010
19. Shin CM, Kim N, Han K, Kim B, Jung JH, Oh TJ, et al. Anti-Diabetic Medications and the Risk for Colorectal Cancer: A Population-Based Nested Case-Control Study. *Cancer Epidemiol* (2020) 64:101658. doi: 10.1016/j.canep.2019.101658
20. Xiao L, Yin X, Di X, Nan Y, Lyu T, Wu Y, et al. Awareness and Prevalence of E-Cigarette Use Among Chinese Adults: Policy Implications. *Tob Control* (2021). doi: 10.1136/tobaccocontrol-2020-056114
21. Murphy CC, Gerber DE, Pruitt SL. Prevalence of Prior Cancer Among Persons Newly Diagnosed With Cancer: An Initial Report From the Surveillance, Epidemiology, and End Results Program. *JAMA Oncol* (2018) 4(6):832–6. doi: 10.1001/jamaoncol.2017.3605
22. Alberti KG, Zimmet PZ. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus Provisional Report of a WHO Consultation. *Diabetes Med* (1998) 15(7):539–53. doi: 10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S
23. Li X, Deng Y, Tang W, Sun Q, Chen Y, Yang C, et al. Urban-Rural Disparity in Cancer Incidence, Mortality, and Survivals in Shanghai, China, During 2002 and 2015. *Front Oncol* (2018) 8:579. doi: 10.3389/fonc.2018.00579
24. Vu HT, Ufere N, Yan Y, Wang JS, Early DS, Elwing JE. Diabetes Mellitus Increases Risk for Colorectal Adenomas in Younger Patients. *World J Gastroenterol* (2014) 20(22):6946–52. doi: 10.3748/wjg.v20.i22.6946
25. Restifo D, Williams JS, Garacci E, Walker RJ, Ozieh MN, Egede LE. Differential Relationship Between Colorectal Cancer and Diabetes in a Nationally Representative Sample of Adults. *J Diabetes Complications* (2018) 32(9):819–23. doi: 10.1016/j.jdiacomp.2018.06.007
26. Lu S, Wang A, Miao S, Zhang X, Jing S, Shan T, et al. Association Between Type 2 Diabetes and Cancer Incidence in China: Data in Hospitalized Patients From 2006 to 2013. *Ann Transl Med* (2020) 8(5):176. doi: 10.21037/atm.2020.01.101
27. Yang WS, Chen PC, Lin HJ, Su TC, Hsu HC, Chen MF, et al. Association Between Type 2 Diabetes and Cancer Incidence in Taiwan: Data From a Prospective Community-Based Cohort Study. *Acta Diabetol* (2017) 54(5):455–61. doi: 10.1007/s00592-017-0966-1
28. Wang Z, Wu Y, Wu J, Wang M, Wang X, Wang J, et al. Trends in Prevalence and Incidence of Type 2 Diabetes Among Adults in Beijing, China, From 2008 to 2017. *Diabetes Med* (2021) 38(9):e14487. doi: 10.1111/dme.14487
29. Wang L, Gao P, Zhang M, Huang Z, Zhang D, Deng Q, et al. Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. *JAMA* (2017) 317(24):2515–23. doi: 10.1001/jama.2017.7596
30. Valent F. Diabetes Mellitus and Cancer of the Digestive Organs: An Italian Population-Based Cohort Study. *J Diabetes Complications* (2015) 29(8):1056–61. doi: 10.1016/j.jdiacomp.2015.07.017

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.726672/full#supplementary-material

Supplementary Figure 1 | Diagram of this community-based prospective cohort study

Supplementary Figure 2 | Age and sex distribution of study subjects.

31. Pang Y, Kartsonaki C, Guo Y, Chen Y, Yang L, Bian Z, et al. Diabetes, Plasma Glucose and Incidence of Colorectal Cancer in Chinese Adults: A Prospective Study of 0.5 Million People. *J Epidemiol Community Health* (2018) 72 (10):919–25. doi: 10.1136/jech-2018-210651
32. Bai B, Chen H. Metformin: A Novel Weapon Against Inflammation. *Front Pharmacol* (2021) 12:622262. doi: 10.3389/fphar.2021.622262
33. Van Alsten SC, Rabkin CS, Sawada N, Shimazu T, Charvat H, Yamaji T, et al. Metabolic Syndrome, Physical Activity, and Inflammation: A Cross-Sectional Analysis of 110 Circulating Biomarkers in Japanese Adults. *Cancer Epidemiol Biomarkers Prev* (2020) 29(8):1639–46. doi: 10.1158/1055-9965.EPI-19-1513
34. Yang X, Tao S, Peng J, Zhao J, Li S, Wu N, et al. High-Sensitivity C-Reactive Protein and Risk of Type 2 Diabetes: A Nationwide Cohort Study and Updated Meta-Analysis. *Diabetes Metab Res Rev* (2021). doi: 10.1002/dmrr.3446
35. Xu CX, Zhu HH, Zhu YM. Diabetes and Cancer: Associations, Mechanisms, and Implications for Medical Practice. *World J Diabetes* (2014) 5(3):372–80. doi: 10.4239/wjd.v5.i3.372
36. Ryu TY, Park J, Scherer PE. Hyperglycemia as a Risk Factor for Cancer Progression. *Diabetes Metab J* (2014) 38(5):330–6. doi: 10.4093/dmj.2014.38.5.330
37. Benhammou JN, Lin J, Hussain SK, El-Kabany M. Emerging Risk Factors for Nonalcoholic Fatty Liver Disease Associated Hepatocellular Carcinoma. *Hepatoma Res* (2020) 6:35. doi: 10.20517/2394-5079.2020.16
38. Shlomai G, Neel B, LeRoith D, Gallagher EJ. Type 2 Diabetes Mellitus and Cancer: The Role of Pharmacotherapy. *J Clin Oncol* (2016) 34(35):4261–9. doi: 10.1200/JCO.2016.67.4044

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 26 November 2021
doi: 10.3389/fonc.2021.752948

Cohort Profile: The Socioeconomic Consequences in Adult Life After Childhood Cancer in Scandinavia (SALiCCS) Research Programme

Friederike Erdmann[1,2], Line Elmerdahl Frederiksen[1†], Hanna Mogensen[3†], Camilla Pedersen[1], Luzius Mader[1,4], Mats Talbäck[3], Andrea Bautz[1], Elli Hirvonen[5], Anniina Kyrönlahti[5,6], Liisa Maria Korhonen[5,6], Henrik Hasle[7], Nea Malila[5], Laura-Maria Madanat-Harjuoja[5,8], Maria Feychting[3] and Jeanette Falck Winther[1,9]*

[1] Childhood Cancer Research Group, Danish Cancer Society Research Center, Copenhagen, Denmark, [2] Division of Childhood Cancer Epidemiology, Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Johannes Gutenberg University Mainz, Mainz, Germany, [3] Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, [4] Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland, [5] Institute for Statistical and Epidemiological Cancer Research, Finnish Cancer Registry, Helsinki, Finland, [6] New Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland, [7] Department of Paediatrics, Aarhus University Hospital, Aarhus, Denmark, [8] Dana Farber Cancer Institute/Boston Children's Cancer and Blood Disorders Center, Harvard Medical School, Boston, MA, United States, [9] Department of Clinical Medicine, Faculty of Health, Aarhus University and Aarhus University Hospital, Aarhus, Denmark

OPEN ACCESS

Edited by:
Jianguang Ji,
Lund University, Sweden

Reviewed by:
Chi-kong Li,
The Chinese University of Hong Kong, China
Volkert Siersma,
University of Copenhagen, Denmark

***Correspondence:**
Friederike Erdmann
friederike.erdmann@uni-mainz.de

†These authors have contributed equally to this work and share second authorship

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *17 August 2021*
Accepted: *26 October 2021*
Published: *26 November 2021*

Citation:
Erdmann F, Frederiksen LE, Mogensen H, Pedersen C, Mader L, Talbäck M, Bautz A, Hirvonen E, Kyrönlahti A, Korhonen LM, Hasle H, Malila N, Madanat-Harjuoja L-M, Feychting M and Winther JF (2021) Cohort Profile: The Socioeconomic Consequences in Adult Life After Childhood Cancer in Scandinavia (SALiCCS) Research Programme. Front. Oncol. 11:752948. doi: 10.3389/fonc.2021.752948

Introduction: The growing number of survivors of childhood cancer, with many years of life ahead, demonstrates the increasing clinical and public health relevance of investigating the risks of social and socioeconomic impairment after a childhood cancer diagnosis and the life-saving treatment. To enrich understanding of the mental, social and socioeconomic difficulties that childhood cancer survivors may face during their life-course, identify particularly vulnerable survivors and overcome the limitations of previous research, we initiated the Socioeconomic Consequences in Adult Life after Childhood Cancer in Scandinavia (SALiCCS) research programme.

Methods: This Nordic cross-border research programme is a collaboration between the Danish Cancer Society, the Finnish Cancer Registry and Karolinska Institutet to investigate a broad range of mental, social and socioeconomic conditions in long-term childhood cancer survivors in Denmark, Finland and Sweden. SALiCCS is based on a registry-based matched cohort design, comprising five-year survivors of cancer diagnosed at ages 0–19 years (1971–2008 in Denmark, 1971–2009 in Finland, 1971–2011 in Sweden), age-, sex- and country-matched population comparisons and sibling comparisons who were followed over time. Outcomes of interest included mental disorders, educational achievements, employment and profession, family life and the need of social security benefits. Individual-level data linkage among various national registries provided the data for the research programme.

Results: The SALiCCS core population comprises 21,292 five-year survivors, 103,303 population comparisons and 29,644 siblings as a second comparison group. The most

common diagnoses in survivors were central nervous system tumours, leukaemias and lymphomas.

Discussion: SALiCCS is the largest, most comprehensive population-based research initiative in this field, based on high-quality registry data with minimal risk of bias. The findings will be informative for evidence-based survivorship care targeting not only somatic late effects but also psychosocial impairments.

Keywords: childhood cancer survivors, survivorship, social and socioeconomic outcomes, family life, register-based research, Denmark, Finland, Sweden

INTRODUCTION

Childhood cancer is of increasing public health concern, as approximately 35,000 new cases are diagnosed yearly in children and adolescents in Europe, and about 500,000 European Union citizens are childhood cancer survivors, with complex needs for medical and psycho-social care (1). Although a growing body of research has addressed a broad range of potential risk factors, the aetiology of most childhood cancers is still largely unknown (2, 3). With remarkable advances in diagnostics and treatment (4, 5), the five-year survival after childhood cancer has improved from 30% in the 1960s to more than 80% nowadays in most of Europe (6–8). As a result of the increasing survival and lack of primary preventive measures (2, 3, 9), the number of childhood cancer survivors in society is growing steadily. This growing population is at risk of long-term health consequences (i.e. late effects) induced by the cancer or the intensive treatment at a young age (8, 10–12). Although many survivors are well after therapy, a wide spectrum of long-term adverse health consequences in childhood cancer survivors has been described (8, 12–18), indicating higher risks of a broad range of somatic and mental late effects, including second cancers (8, 12, 17, 18), higher overall mortality rates (12, 16), severe chronic health conditions (8, 12–15), mental disorders (19, 20) and use of antidepressants (21).

The experience of cancer during childhood and adverse somatic or mental health conditions may also have consequences for social and family life and for socioeconomic achievement later in life. Previous research has shown that childhood cancer survivors are at increased risk of several adverse socioeconomic and social outcomes, including scholastic difficulties, such as requiring special education or attending learning disability programmes, lower levels of attained education and lower income than their peers (22). There are only few studies on the uptake of social security benefits in survivors, and almost all are from the Nordic countries. Increased uptake of various social security benefits by survivors was reported consistently (22). Empirical observations on employment and occupation are less conclusive (22, 23), with heterogeneous findings from Europe (22, 23) and a consistently higher risk of unemployment among survivors in studies from Canada and the USA (23–25).

The current evidence is limited by methodological shortcomings. Most previous research is based on self-reported information from surveys and are thereby susceptible to non-participation, which might have affected the outcomes. Further, many studies included survivors of only one or a few specific childhood cancer types, did not involve repeated measurements of social and socioeconomic outcomes throughout the life-course or suffered from substantial loss to follow-up. Further limitations of previous studies include insufficient sample size of survivors and short follow-up. The mechanisms that lead to adverse social and socioeconomic conditions, especially in vulnerable subgroups of survivors, are still poorly understood (22, 26), and better knowledge would be of significant importance for developing interventions and supportive strategies for these vulnerable groups.

The Socioeconomic Consequences in Adult Life after Childhood Cancer in Scandinavia (SALiCCS) research programme was initiated to address these gaps and enrich understanding of the impairments and social and socioeconomic difficulties that survivors of childhood cancer may face during their life-course. This Nordic research programme is a collaboration among the Danish Cancer Society Research Center, the Finnish Cancer Registry and Karolinska Institutet in Sweden.

The Welfare Systems of Denmark, Finland, and Sweden

In 2020, Denmark, Finland and Sweden had populations of 5.8, 5.5 and 10.3 million, respectively (27). The Nordic countries are well known for their generous welfare systems (28), with the core principles of solidarity and universalism and the overall aim of providing equal access for all citizens to welfare services, including health care, education and social security benefits. Decommodification in the three countries ensures that such services are provided independently of an individual's affiliation to the labour force (28). The welfare services are tax-funded and, as a result, Denmark, Finland and Sweden have some of the highest tax revenues in the world of above 40% of the gross domestic product (GDP) (29). In general, the welfare

Abbreviations: ALiCCS, Adult Life after Childhood Cancer in Scandinavia; CNS, Central Nervous System; NOPHO, Nordic Society of Paediatric Haematology and Oncology; SALiCCS, Socioeconomic Consequences in Adult Life after Childhood Cancer in Scandinavia.

services provided in these countries result in a high standard of living, demonstrated for instance by high life expectancy (30).

The welfare systems of the three countries are largely comparable. Citizens are entitled to an education free of charge, from primary schooling, which is compulsory, to advanced tertiary educational levels (31). Additionally, students enrolled in tertiary education, such as college or university, may receive direct financial support and loans from the government. The unemployment rates in Denmark, Finland and Sweden are low (32) (from 5.7% to 8.8% in 2017), and few citizens rely on income support from the government (i.e. social security benefits) (32). Increasing numbers of women in the Nordic countries have entered the workforce over the past six decades (33), and the active labour force has almost equal gender distribution (32). This is to some extent enabled by supportive arrangements such as state-subsidised childcare provision, generous parental leave schemes and, often, flexible working hours (34, 35).

In general, all citizens of the Nordic countries have equal access to government-subsidized primary health care services provided by general practitioners and specialised health care in hospitals or provided by specialist physicians. Out-of-pocket expenses and reimbursement schemes vary, however, in the three countries (36–38). The health care services in all three countries also cover all costs directly related to the diagnosis and treatment of childhood cancer and for the vast majority of health care for any late effects. In Denmark and Sweden, more than 80% of all health care expenditure is covered by the public tax-financed system (36, 37, 39). In Finland, total public coverage of health care expenditures is also about 80%, funded primarily through taxation (around 60%) but also through health insurance contributions paid to the tax administration (38).

Childhood Cancer Treatment in the Nordic Countries

As childhood cancers are a heterogeneous group consisting of very different diseases, survival and developments in survival over time differ widely by cancer type (6). Since the 1970s, as a result of advances in molecular tumour biology, imaging, pharmacology, risk grouping and treatment combinations, overall five-year survival rose from 30% in the pre-chemotherapeutic era to more than 80% nowadays (6). Early collaborative clinical research to identify effective therapy for children with cancer dates back to the 1950s, when children with acute lymphoblastic leukaemia were some of the earliest participants in clinical trials of new drugs for cancer treatment (5). Participation in clinical trials is today considered the standard of paediatric cancer care, and a large number of children in Europe and North America are enrolled in protocols (4, 40–42) developed by collaborative study groups, such as the Nordic Society of Paediatric Haematology and Oncology (NOPHO) (4, 40, 41).

Current therapy for some malignancies is highly intensive, and, while survival has gradually increased, the risk of treatment-related acute toxicity and late effects may also be increasing. The aim of many current protocols is to identify subgroups of patients for whom the intensity of therapy can be reduced to decrease toxicity.

Collaborative clinical trials to standardise childhood leukaemia treatment protocols in all Nordic countries began in 1981 (4, 40, 43), and since 1992 almost all children with leukaemia have been treated with the standardised NOPHO protocols. Treatment for lymphomas and solid tumours has been similarly standardised and is based mainly on the protocols of international collaborative study groups with NOPHO participation.

Aims and Objectives of the SALiCCS Research Programme

The SALiCCS research programme is a Nordic register-based cohort study of social and socioeconomic consequences in long-term survivors of childhood cancer in Denmark, Finland and Sweden. SALiCCS is based on the Nordic research programme Adult Life after Childhood Cancer in Scandinavia (ALiCCS), which investigates late effects of cancer therapy in children to better understand the risk and mechanisms of treatment-induced somatic disease (44). Data from ALiCCS are enriched by longitudinal information on mental disorders and social and socioeconomic outcomes after a childhood cancer diagnosis. With this strategy, the overarching goal of SALiCCS is to identify groups of childhood cancer survivors for whom early intervention would minimise later mental disorders and adverse social and socioeconomic consequences of the cancer and its treatment.

The main objectives of the SALiCCS research programme are:

i. to ascertain hospital contacts for mental disorders in childhood cancer survivors;
ii. to examine how survivors of childhood cancer transition from childhood to adulthood by determining the following social and socioeconomic conditions and attainments: scholastic achievements, attained educational level and educational delays, income, employment, occupational position and professional attainment, leaving the parental household to live independently, cohabitation with a partner, getting married and founding a family; and
iii. to assess the socioeconomic burden of childhood cancer and treatment on survivors by determining the uptake of social security benefits, such as unemployment benefits, social assistance, sickness allowance, disability pension and rehabilitation benefits.

MATERIAL AND METHODS

Design and Research Setting

The SALiCCS research programme is based on a registry-based matched cohort design. Denmark, Finland and Sweden have civil registration systems with numerous national administrative registries (45–47) that contain individual-level data in various

fields, including cancer diagnoses (48), hospitalisation for somatic and mental disorders (49–52), vital status (46, 53), emigration and immigration (45, 46), perinatal and birth characteristics (54–56) and socio-demographic and socioeconomic characteristics (57–64). Nordic citizens are assigned a unique personal identification number (Denmark since 1968, Finland since 1964 and Sweden since 1947) that is used in all national registries, enabling accurate linkage of information among registries (45–47, 65). National legislations permit and supports registry-based research. Data linkage among the registries is the basis of the SALiCCS research programme. Data from all three countries are collected and harmonized to enable pooled analyses (see section 2.5 for information on data access and data protection). **Figure 1** gives an overview of the SALiCCS study design and the registers used.

The infrastructure of the population-based registers in the Nordic countries, with longstanding, high-quality, comprehensive health, socio-demographic and socioeconomic data, are an ideal, unique basis for large-scale epidemiological studies of childhood cancer survivorship. None of the Nordic countries, however, has a sufficiently large population to provide adequate statistical power for a detailed assessment of social and socioeconomic outcomes in childhood cancer survivors in a life-course perspective, particularly not for determining the underlying mechanisms of adverse social and socioeconomic outcomes and identifying particularly vulnerable groups of survivors. Combination of data from several Nordic countries is required for such purposes. As the Nordic countries have longstanding, largely standardised diagnostic and treatment procedures and similar welfare systems, it was considered reasonable to combine data on childhood cancer survivors across Nordic countries.

Study Population

The SALiCCS core population comprises all five-year survivors of a first childhood cancer (including non-malignant central nervous system tumours) diagnosed at ages 0–19 years in Denmark (1971–2008), Finland (1971–2009) and Sweden (1971–2011) (**Table 1** and **Figure S1**). For individual SALiCCS studies however, also more broadly defined criteria for the study population may be applied (e.g. 1-year survival as eligibility criteria or including all childhood cancer cases). We identified childhood cancer cases from the respective nationwide cancer registries, which are of excellent quality and high completeness (48).

The cancer diagnoses were classified according to the International Classification of Childhood Cancer, in which tumours are classified into 12 major diagnostic groups and detailed subgroups according to the nomenclature of the International Classification of Diseases – Oncology (66–69).

We established two independent comparison groups. We randomly sampled five population-based comparisons per survivor from the populations of Denmark, Finland and Sweden from the national population registries. Survivors and population comparisons were individually matched by year of birth, sex and country of residence (and municipality of residence in Sweden) (**Figure 1** and **Figure S1**). The second comparison group constitutes all biological full and half siblings as well as adopted (except in Finland) siblings of the childhood cancer survivors, defined as having either the same (biological or adoptive) mother or father. One sibling could serve as a sibling comparison for several survivors if more than one child in the same family was a childhood cancer survivor. We included only siblings with a maximum age difference of 10 years in order to allow meaningful comparisons of outcomes between siblings and corresponding childhood cancer survivors, leaving the possibility to reduce the age difference even more for some outcomes when relevant. Sibling and population comparisons had to be alive and not lost to follow-up five years after the reference date, and cancer-free until the age of 20 years to be eligible as comparisons. The reference date was defined as the date of diagnosis of cancer

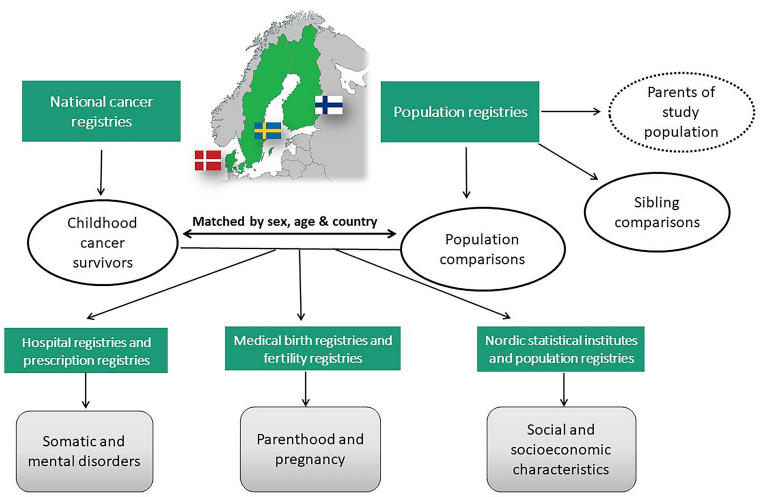


FIGURE 1 | Overview of the SALiCCS study design and data sources.

in the corresponding matched survivor for population comparisons and, for siblings, as the date on which the sibling was of the same age as the corresponding survivor at cancer diagnosis.

As a cancer predisposition syndrome may confound associations with mental, social and socioeconomic outcomes, we excluded individuals with Down syndrome, neurofibromatosis or tuberous sclerosis, resulting in a final SALiCCS core population of 21,292 five-year childhood cancer survivors, 103,303 population comparisons and 29,644 siblings (**Figure S1**).

Follow-Up, Social and Socioeconomic Outcomes and Information on Potential Mediators and Confounders

The childhood cancer survivors and their comparisons were followed from five years after the cancer diagnosis or reference date until death, first emigration, loss to follow-up or end of follow-up, whichever came first. Information on end of follow-up varied somewhat by outcome data, depending on the registry used as the data source (**Table 1**).

For childhood cancer survivors and their comparisons, we obtained comprehensive information on highest attained education, scholastic achievements, educational delays and other educational information, individual and household disposable income, labour market affiliation, occupation position and the uptake of various social security benefits, including annual unemployment, sickness and disability benefits, social assistance and rehabilitation allowances from the social registries administered by the Nordic statistical institutes. "Disposable income" refers to annual individual and household income after taxes, including social security benefits.

Information on family structure, including cohabitation and marriage, place of residence, parenthood and other socio-demographic information, was obtained from the population registries, while data on birth characteristics were obtained from the medical birth registers, which contain mandatory, regularly updated reports on all births in the respective countries. From the national patient registries, we received comprehensive histories of hospital admissions (including outpatient contacts) for somatic and mental disorders and the respective discharge diagnoses. We obtained information on prescribed drugs from the nationwide prescribed drug registries, whereby prescription data from Finland was limited to contraceptive medications, antidepressants, pain killers, and psychiatric drugs.

Apart from diagnostic data from the national cancer registries, clinical and treatment information was overall sparse. The Finnish Cancer Registry included limited and non-validated treatment information on surgery, radiotherapy and chemotherapy (with incomplete coverage). The treatment information is given on binary level for curative, palliative or unknown intention (48). For Denmark and Sweden, some clinical and treatment data was available from the Swedish Childhood Cancer Registry and from the NOPHO database, however only for leukaemia survivors and for a limited follow-up period.

Parental Information

We collected basic socio-demographic and socioeconomic information for the biological and adoptive parents (for

TABLE 1 | Study period and data availability of parental information and information about the index subjects in the SALiCCS research programme by country.

	Denmark	Finland	Sweden
Study period of cancer diagnosis/ reference	1971-2008	1971-2009	1971-2011
End of follow-up	11 Aug 2017	31 Dec 2014	31 Dec 2016
Time period with available sociodemographic and socioeconomic information			
Region of residence	1980-2017	1936-2016	1968-2016
Educational achievements	1970-2016	1970, 1975, 1980, 1985, 1987-2016	1960, 1963-2016[a]
Employment	1980-2017	1970, 1975, 1980, 1985, 1987-2016	1960, 1965, 1970, 1975, 1980, 1985, 1990-2015
Occupational position	1993-2017	1990, 1993, 1995, 2000, 2004-2015	1960, 1970, 1975, 1980, 1985, 2001-2015
Disposable income	1980-2018	1995-2016	1970[b], 1975, 1985, 1990-2015
Social security benefits	1980-2016 (differs between benefit types)	1985-2016 (differs between benefit types)	1990-2015 (differs between benefit types)
Marital status	1968-2017	1970-2016	1960, 1965, 1968-2016
Cohabitation status	1980-2017	1987-2016	1970, 1975, 1980, 1985, 1990-2015[d]
Parenthood	1968-2017	1971-2016	1961-2016
Time period with available information on hospital contacts for somatic and psychiatric care and prescribed drugs			
Somatic diseases	1977-2017	1969-2014	1964[c]-2016
Mental disorders	1969-2017	1975-2014	1973[c] -2016
Prescribed drugs	1995-2017	1995-2016	2005-2017

[a] *Differences in the availability of educational information for the early time period.*
[b] *Net income is used in 1970 as disposable income was not available.*
[c] *Nationwide since 1987.*
[d] *1990-2015: Restricted to cohabitating couples with common children.*

Finland only for biological parents) of the childhood cancer survivors, the population comparisons, and the sibling comparisons. Furthermore, to account for the increasing number of reconstituted families in the Nordic countries, we also defined the "social parents" for the Danish and Swedish SALiCCS populations, for whom we collected the same socioeconomic and socio-demographic information as for biological parents. In Denmark, social parents were defined as individuals living at the same address with the index person the year before the reference year, at least 16 years older and not a full or half sibling of the index person. In Sweden, social parents were defined as individuals registered by Statistics Sweden as a parent or guardian in the same household in which the index person was defined, the year before the reference date. Social parents could be identified from 1980 onwards in Denmark and from 1990 onwards in Sweden. Corresponding information was not collected for the Finnish SALiCCS population.

Data Access, Data Protection and Other Ethical Considerations

The SALiCCS research programme has been approved by Statistics Denmark, the Regional Ethical Review Board in Stockholm, Sweden (dnr 2016/25-31/5, 2016/1561-32, 2017/1656-32, 2017/1990-32, 2017/2340-32, 2018/1165-32), Findata (Dnro THL/5543/14.06.00/2020) prolonging the former approvals by the National Institute for Health and Welfare and Social Insurance (KELA) and Statistics Finland (TK-53-394-17) in Finland. For the European Union General Data Protection Regulation (GDPR), the SALiCCS project is listed in a local archive (2018-DCRC-0044) at the Danish Cancer Society Research Center, which provides an accurate, updated overview of ongoing projects and of ongoing research projects involving personal data under the GDPR. The 2018-DCRC-0044 replaces the former notification from the Danish Data Protection Agency.

The SALiCCS research programme is conducted in compliance with the requirements of the GDPR and other applicable laws in the respective countries, as well as the respective procedures at Statistics Denmark, the Danish Cancer Society Research Center, Karolinska Institutet and the Finnish Cancer Registry. All data have been stored, linked and pooled and are analysed at a secure remote platform at Statistics Denmark, with controlled remote access only for individually approved SALiCCS project members. Personal identification numbers were replaced by pseudonymised ID numbers, and the key code is kept only by the original register holders or at the respective statistical institutes. All the results of the statistical analyses will only be presented as aggregated data.

SPECIFIC CHARACTERISTICS OF THE STUDY POPULATION

The core SALiCCS population comprises 21,292 five-year survivors, 103,303 population comparisons and 29,644 siblings as the second comparison group. The distribution of diagnostic characteristics differed only slightly in Denmark, Finland and Sweden (**Table 2**). The most common diagnoses were tumours of the central nervous system (23%), leukaemias (23%) and lymphomas (15%). Slightly over half of the childhood cancer survivors were male (53%), 31% were diagnosed before the age of 5 years, and about 60% were diagnosed between 1990 and 2009. The length of follow-up after the date of cancer diagnosis ranged from 5.0 to 46.6 years, with a median follow-up time of 20.7 years in Denmark, 19.1 years in Finland and 20.6 years in Sweden (**Table 2**).

The SALiCCS core population is defined by individuals who survived the five-year survival point. To describe this population, the survival probabilities of all children diagnosed with cancer in Denmark, Finland and Sweden, their population comparisons and siblings are shown in **Figure 2** (including children who died or emigrated within the first five years of diagnosis/reference date) and a comparison of the distribution of diagnostic groups and specific cancer types between all incident childhood cancer cases in Denmark, Finland and Sweden and the SALiCCS 5-year childhood cancer survivors is given in **Table S1**. Overall survival figures for this extended population are also given by diagnostic decade, age at diagnosis, diagnostic group and country (**Figure 2** and **Supplementary Material S2**–**S5**). We intentionally excluded any adverse social and socioeconomic conditions that may have arisen while being hospitalised and receiving cancer treatment by starting follow-up at five years after the childhood cancer diagnosis, since these acute and direct outcomes would not reflect the social and socioeconomic consequences in long-term survivors.

Tables 3 and **S2** present the distribution of socio-demographic characteristics in the SALiCCS core population, consisting of all individuals who survived five years after diagnosis or the respective reference date for comparisons. The distribution of place of residence, parental age and parental socioeconomic characteristics differed only slightly for childhood cancer survivors and population and sibling comparisons (**Table 3**).

This SALiCCS core population will serve as the basis for individual studies of the objectives of the SALiCCS research programme, as outlined in section 1.3. Additional inclusion and exclusion criteria may be applied to the core population in individual SALiCCS studies, which may be restricted to specific time periods, depending on the respective research objectives and register coverage.

DISCUSSION

The growing number of survivors of childhood cancers diagnosed at a young age, with many years of life ahead, indicates the increasing clinical and public health relevance of investigating the long-term social and socioeconomic consequences of the cancer diagnosis and the life-saving treatment. Some previous evidence points to higher risks of impaired social functioning and adverse socioeconomic outcomes in adult life. Nevertheless, additional research is urgently required to fully understand the long-term social consequences of a diagnosis of cancer in

TABLE 2 | Characteristics of 5-year childhood cancer survivors diagnosed in 1971-2008 (DEN), 2009 (FIN), 2011 (SWE) by country and for the three countries combined.

	Denmark		Finland		Sweden		Three-Country Wide pooled data	
	N	%	*N*	%	*N*	%	*N*	%
Total	5343		5672		10277		21292	
Sex								
Boys	3000	56.1	2939	51.8	5405	52.6	11344	53.3
Girls	2343	43.9	2733	48.2	4872	47.4	9948	46.7
Age at diagnosis (years)								
<1	306	5.7	377	6.7	607	5.9	1290	6.1
1 – 4	1294	24.2	1434	25.3	2475	24.1	5203	24.4
5 – 9	964	18.0	948	16.7	1844	17.9	3756	17.6
10 – 14	1046	19.6	1106	19.5	2005	19.5	4157	19.5
15 – 19	1733	32.4	1807	31.9	3346	32.6	6886	32.3
Decade of diagnosis								
1971 – 1979	824	15.4	797	14.1	1578	15.4	3199	15.0
1980 – 1989	1259	23.6	1316	23.2	2408	23.4	4983	23.4
1990 – 1999	1592	29.8	1765	31.1	2841	27.6	6198	29.1
2000 – 2009	1668	31.2	1794	31.6	2803	27.3	6265	29.4
2010 - 2011	–	–	–	–	647	6.3	647	3.0
Decade of birth								
1951 – 1959	182	3.4	235	4.1	440	4.3	857	4.0
1960 – 1969	722	13.5	645	11.4	1285	12.5	2652	12.5
1970 – 1979	1342	25.1	1310	23.1	2377	23.1	5029	23.6
1980 – 1989	1508	28.2	1728	30.5	2659	25.9	5895	27.7
1990 – 1999	1178	22.1	1270	22.4	2427	23.6	4875	22.9
2000 – 2009	411	7.7	484	8.5	1048	10.2	1943	9.1
2010 – 2011	–	–	–	–	41	0.4	41	0.2
Cancer type[a]								
Leukaemias	1189	22.3	1382	24.4	2324	22.6	4895	23.0
Lymphoid leukaemia[b]	1000	18.7	1170	20.6	1895	18.4	4065	19.1
Acute myeloid leukaemia[c]	134	2.5	130	2.3	235	2.3	499	2.3
Other leukaemia	55	1.0	82	1.5	194	1.9	331	1.6
Lymphomas	759	14.2	952	16.8	1465	14.3	3176	14.9
Hodgkin lymphoma[d]	440	8.2	569	10.0	813	7.9	1822	8.6
Non-Hodgkin lymphoma[e]	196	3.7	364	6.4	213	2.1	773	3.6
Other lymphoma	123	2.3	19	0.3	439	4.3	581	2.7
CNS tumours[f]	1277	23.9	1110	19.6	2524	24.6	4911	23.1
Ependymoma	103	1.9	110	1.9	230	2.2	443	2.1
Astrocytoma and other gliomas	510	10.0	697	12.3	1088	10.6	2295	10.8
Embryonal CNS tumours	130	2.4	108	1.9	384	3.7	622	2.9
Other specified or unspecified CNS tumour	534	10.0	195	3.4	822	8.0	1551	7.3
Sympathetic nervous system tumours	169	3.2	209	3.7	244	2.4	622	2.9
Retinoblastomas	148	2.8	124	2.2	226	2.2	498	2.3
Renal tumours	232	4.3	273	4.8	487	4.7	992	4.7
Hepatic tumours	35	0.7	35	0.6	80	0.8	150	0.7
Malignant bone tumours	180	3.4	202	3.6	396	3.9	778	3.7
Soft tissue sarcomas	276	5.2	315	5.6	488	4.8	1079	5.1
Germ cell tumours	466	8.7	322	5.7	642	6.3	1430	6.7
Malignant epithelial neoplasms	533	10.0	676	11.9	1110	10.8	2319	10.9
Other & unspecified malignant neoplasms	79	1.5	72	1.3	291	2.8	442	2.1
Secondary malignancy before age of 20 years								
Yes	66	1.2	65	1.2	101	1.0	232	1.1
No	5277	98.8	5607	98.8	10176	99.0	21060	98.9
End event								
Death	569	10.7	542	9.6	831	8.1	1942	9.1
Emigration (1st emigration)	391	7.3	59	1.0	433	4.2	883	4.2
Lost to follow-up	5	0.1	0	0	0	0	5	<0.1
End of study	4378	81.9	5071	89.4	9013	87.7	18462	86.7
DEN: 11 Aug 2017								

(Continued)

TABLE 2 | Continued

	Denmark		Finland		Sweden		Three-Country Wide pooled data	
	N	%	*N*	%	*N*	%	*N*	%
SWE: 31 Dec 2016 FIN: 31 Dec 2014 **Median time of follow-up in years (range)**	20.7 (5.0-46.6)		19.1 (5.0-44.0)		20.6 (5.0-46.0)		20.1 (5.0-46.6)	

No missing information for any of the characteristics given in this table.
[a]*Classified by the International Classification of Childhood Cancer.*
[b]*Lymphoid leukaemia defined as ICCC1 group I a-b and ICCC3 group Ia.*
[c]*Acute myeloid leukaemia defined as ICCC1 group Ic and ICCC3 group Ib.*
[d]*Hodgkin lymphoma defined as ICCC1 and ICCC3 group IIa.*
[e]*Non-Hodgkin lymphoma defined as ICCC1 and ICCC3 group IIb.*
[f]*CNS tumor subtypes were grouped as follows: Ependymoma (defined by ICCC 1 and ICCC3 group 3a), astrocytoma and other gliomas (ICCC 1 and ICCC 3 groups 3b and 3d combined), and embryonal CNS tumours (defined by ICCC 1 and ICCC3 group 3c).*

childhood, to identify vulnerable survivors at particular risk for adverse social and socioeconomic impairments in adulthood and ultimately to provide scientific knowledge for evidenced-based survivorship care that addresses not only somatic late effects but also psychosocial impairments.

Data linkage among various population-based registries in Denmark, Finland and Sweden gave us the unique possibility of setting up the largest, most comprehensive population-based research initiative on the social and socioeconomic consequences of childhood cancer so far, comprising more than 21,000 five-year childhood cancer survivors and two independent comparison groups. The large number of survivors will enable detailed analyses, allowing for identification of vulnerable subgroups in terms of e.g. diagnostic or socio-demographic characteristics. Use of high-quality population-based register data with high coverage, virtually no loss to follow-up and no self-reporting or non-participation allows reliable estimation of the social and socioeconomic outcomes with minimal risk of bias in countries with similar welfare systems and generally equal access to education and other services. Particularly valuable strengths of the SALiCCS research programme include annual information on social and socioeconomic outcomes, facilitating the study of trajectories and enabling a comprehensive long-term follow-up, as well as the availability of detailed information on hospital contacts, allowing stratified analyses by somatic or mental late effects. A limitation of the SALiCCS research programme is the lack of comprehensive and three-countrywide information on cancer treatment and other relevant clinical characteristics such as subtype of disease and tumour stage or grade. Such information would have been of considerable value for assessing the underlying mechanisms leading to adverse social and socioeconomic conditions, and to identify survivors of childhood cancer that are particularly vulnerable to adverse outcomes. For instance, especially cranial radiation therapy has been associated with various somatic late effects, including long-term neurocognitive impairment, as well as with adverse educational attainments, higher risk of unemployment and low income (22). Understanding the causal

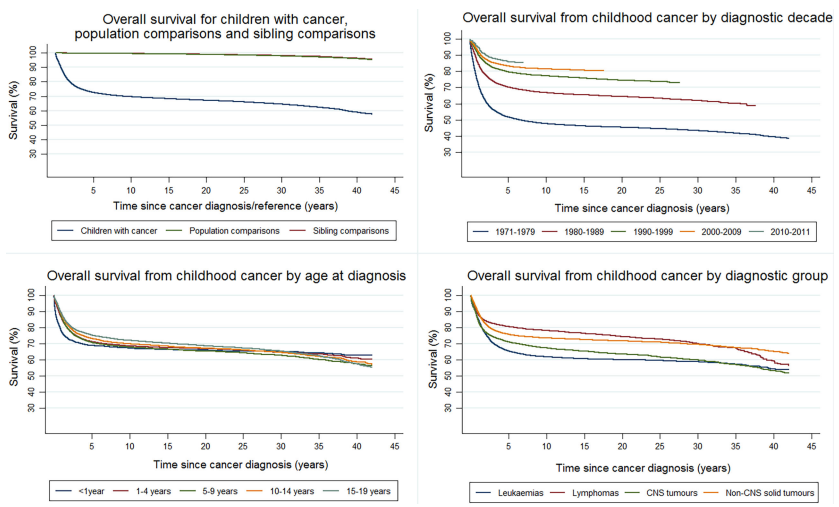


FIGURE 2 | Overall survival in the SALiCCS population: by children with cancer, population comparisons and siblings, by diagnostic decade, by age at diagnosis and by diagnostic group.

TABLE 3 | Sociodemographic and socioeconomic characteristics of 5-year childhood cancer survivors, population comparisons and siblings, pooled data from the three Nordic countries.

	Three-country wide pooled data					
	Childhood cancer survivors		Population comparisons		Siblings	
	N	%	N	%	N	%
Total	21292		103303		29644	
Sex						
Boys	11334	53.3	55134	53.4	15137	51.1
Girls	9948	46.7	48169	46.6	14507	48.9
Decade of birth						
1951 – 1959	857	4.0	4129	4.0	1065	3.6
1960 – 1969	2652	12.5	12884	12.5	4166	14.1
1970 – 1979	5029	23.6	24381	23.6	6908	23.3
1980 – 1989	5895	27.7	28516	27.6	8650	29.2
1990 – 1999	4875	22.9	23738	23.0	6667	22.5
2000 – 2009	1943	9.1	9454	9.2	2149	7.3
2010 – 2011	41	0.2	201	0.2	39	0.1
Region of residence[a,b]						
Major city	6363	32.4	30699	32.1	8301	30.2
Town & suburb	7354	37.4	35606	37.3	10085	36.7
Rural areas	5743	29.2	28086	29.4	8652	31.5
Missing[a]	211	1.1	1177	1.2	453	1.7
Parents						
Biological mothers	21073	99.0	101873	98.6	29447	99.3
Biological fathers	20763	97.5	100179	97.0	29159	98.4
Maternal age[c,d]						
≤25	7684	36.1	38420	37.2	11853	40.0
26 – 30	7297	34.3	34561	33.5	10086	34.0
31 – 35	4299	20.2	20484	19.8	5585	18.8
36 – 40	1535	7.2	7175	7.0	1705	5.8
41 – 45	250	1.2	1198	1.2	214	0.7
≥46	11	0.1	54	0.1	4	<0.1
Missing	216	1.0	1411	1.4	197	0.7
Paternal age[c,d]						
≤25	4338	20.4	21644	21.0	6599	22.3
26 – 30	7034	33.0	33657	32.6	9960	33.6
31 – 35	5411	25.4	26278	25.4	7633	25.8
36 – 40	2647	12.4	12394	12.0	3422	11.5
41 – 45	944	4.4	4404	4.3	1124	3.8
≥46	409	1.9	1981	1.9	419	1.4
Missing	509	2.4	2945	2.9	487	1.6
Maternal education[a,c,d,e]						
Short	5729	28.1	29109	29.4	8821	31.2
Medium	7979	39.1	37964	38.4	10563	37.3
Higher	5249	25.7	24251	24.5	6702	23.7
Missing[a]	1437	7.1	7614	7.7	2231	7.9
Paternal education[a,c,d,e]						
Short	5643	27.7	28015	28.3	8690	30.7
Medium	8514	41.8	40679	41.1	11269	39.8
Higher	4712	23.1	21868	22.1	6213	21.9
Missing[a]	1525	7.5	8376	8.5	2145	7.6
Maternal employment status[a,c,d]						
Employed	14447	74.2	69027	73.0	18293	67.4
Unemployed	4760	24.5	23841	25.2	7976	29.4
Missing[a]	257	1.3	1703	1.8	860	3.2
Paternal employment status[a,c,d]						
Employed	16673	85.7	80048	84.6	22666	85.6
Unemployed	2269	11.7	11296	12.0	3339	12.3
Missing[a]	522	2.7	3227	3.4	1124	4.1
Maternal disposable income[a,c,d,f]						
1st quartile	3791	22.0	18680	22.4	5207	22.1
2nd quartile	4044	23.5	20138	24.2	5269	22.4
3rd quartile	4508	26.2	21446	25.7	5839	24.8
4th quartile	4591	26.7	21401	25.7	6582	27.9

(Continued)

TABLE 3 | Continued

	Three-country wide pooled data					
	Childhood cancer survivors		Population comparisons		Siblings	
	N	%	N	%	N	%
Missing[a]	266	1.6	1686	2.0	678	2.9
Paternal disposable income[a,c,d,f]						
1st quartile	3732	21.7	18892	22.7	5353	22.7
2nd quartile	4080	23.7	20485	24.6	5651	24.0
3rd quartile	4419	25.7	20776	24.9	5824	24.7
4th quartile	4458	25.9	20149	24.2	5861	24.9
Missing[a]	511	3.0	3049	3.7	886	3.8
Average days/year with hospital visit[g]						
	Average days/year		Average days/year		Average days/year	
0-4 years after reference date	18.2		0.8		0.9	
5-9 years after reference date	6.2		1.0		1.1	
10-14 years after reference date	3.8		1.3		1.4	
15-19 years after reference date	3.4		1.4		1.6	
20-24 years after reference date	3.4		1.6		1.8	
25-29 years after reference date	3.9		1.8		1.9	
30-34 years after reference date	4.4		1.7		1.9	
35-39 years after reference date	5.5		1.8		2.4	
End event						
Death	1942	9.1	1416	1.4	389	1.3
Emigration (1st emigration)	883	4.2	5606	5.4	1526	5.2
Lost to follow-up	5	<0.1	26	<0.1	3	<0.1
End of study DEN: 11 Aug 2017 SWE: 31 Dec 2016 FIN: 31 Dec 2014	18462	86.7	96255	93.2	27726	93.5
Median time of follow-up (years)	20.1 (5.0-46.6)		21.4 (5.0-46.6)		22.1 (5.0-46.6)	
Type of sibling[h]						
Full sibling	–	–	–	–	24710	83.4
Half sibling	–	–	–	–	4934	16.6
Age difference between survivor and sibling						
Less than 5 years older	–	–	–	–	9827	33.2
5-10 years older	–	–	–	–	6009	20.3
Less than 5 years younger	–	–	–	–	8807	29.7
5-10 years younger	–	–	–	–	5001	16.9

Reference date corresponds to the date of diagnosis for the population comparisons. For siblings the reference date corresponds to the date when a sibling was of the same age as the respective survivor at diagnosis.

[a]*Information tied to the years of register coverage in the respective country (see* ***Table 1****). Parental socioeconomic information from Finland were tied to the years from 1980 onwards.*

[b]*Corresponds to reference year.*

[c]*Characteristics correspond to the biological parents.*

[d]*Corresponds to the year before reference year; if not available, then to the year closest to the year before reference year.*

[e]*Parental highest education was grouped into short [early childhood education, primary and lower secondary education, ISCED levels 0-2], medium [upper secondary including vocational upper secondary education, ISCED level 3] and higher [ISCED level 4-8] education, following the International Standard Classification of Education (ISCED). Parents with missing education in Finland have been allocated to lowest education category, as only education from secondary level and above is registered in Finland.*

[f]*Annual disposable income was categorised into four groups based on the sex- and calendar-year specific income distribution (quartiles) of the population comparisons in the respective country.*

[g]*Average number of days/year with inpatient and outpatient hospital contact for any somatic and mental disorders during 5-year periods after reference date.*

[h]*Adoptive siblings from Sweden were assigned to the group of full siblings, whereas adoptive siblings from Denmark cannot be specifically identified and therefore may be found in both groups.*

pathway of those adverse outcomes would be of substantial clinical and public health relevance. We do, however, have information on hospital contacts, including the full history, with dates of in- and outpatient hospital contacts including corresponding diagnoses, which may be used in some individual SALiCCS studies as indicators of severity of disease, length of treatment and disease burden.

The novel findings resulting from this research programme may serve as a basis for recommendations on interventions for vulnerable subgroups of survivors. Such recommendations may not be limited to the Nordic countries, with their extensive welfare systems, but may also be applicable to other countries, especially within Europe.

DATA AVAILABILITY STATEMENT

The data that support the information of this manuscript were accessed remotely on a secure platform at Statistics Denmark. Pseudonymised individual-level data were obtained from national registry holders after ethical approval (where applicable) and secrecy assessment. According to Danish,

Finnish and Swedish laws and regulations, individual-level sensitive data can only be made available for researchers who fulfil legal requirements for access to personal sensitive data. Please contact Jeanette Falck Winther (jeanette@cancer.dk), the Principal Investigator of the SALiCCS research programme, for further questions about data access.

ETHICS STATEMENT

The SALiCCS research programme has been approved by Statistics Denmark, the Regional Ethical Review Board in Stockholm, Sweden (dnr 2016/25-31/5, 2016/1561-32, 2017/1656-32, 2017/1990-32, 2017/2340-32, 2018/1165-32), Findata (Dnro THL/5543/14.06.00/2020) prolonging the former approvals by the National Institute for Health and Welfare and Social Insurance (KELA) and Statistics Finland (TK-53-394-17) in Finland. For the European Union General Data Protection Regulation (GDPR), the SALiCCS project is listed in a local archive (2018-DCRC-0044) at the Danish Cancer Society Research Center, which provides an accurate, updated overview of ongoing projects and of ongoing research projects involving personal data under the GDPR. The 2018-DCRC-0044 replaces the former notification from the Danish Data Protection Agency.

AUTHOR CONTRIBUTIONS

FE and JW developed the concept and outline of the manuscript. All authors contributed to the acquisition and preparation of data. FE, LF, HM, NM, L-MM-H, MF, and JW developed the strategy for the descriptive analysis and presentation of the study population. FE, LF, HM, and JW drafted the manuscript. All authors provided critical feedback, critically reviewed the manuscript for important intellectual content, and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by NordForsk under grant 76111, the Danish Childhood Cancer Foundation under grant 2016-0293, Aarhus University under fellowship 43239402, the Swiss National Science Foundation under grant P2LUP3_175288 (personal fellowship to LM), the Tømrermester Jørgen Holm og Hustru Elisa F. Hansens Mindelegat under grant 20088 and the Swedish Childhood Cancer Foundation under grant PR2020-0130. The funding sources had no involvement in the content or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors are grateful to Anja Krøyer (Danish Cancer Society Research Center, Childhood Cancer Research Group) for her technical support with data acquisition and preparation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.752948/full#supplementary-material

Supplementary Figure 1 | Flow diagram of the sampling of the SALiCCS core population.

Supplementary Figure 2 | Overall survival for children with cancer, population comparisons and sibling, for the entire SALiCCS population and by country (reference date in 1971-2008 (DEN), 2009 (FIN), 2011 (SWE)).

Supplementary Figure 3 | Overall survival from childhood cancer by diagnostic decade, for the entire SALiCCS population and by country (children diagnosed with cancer in 1971-2008 (DEN), 2009 (FIN), 2011 (SWE)).

Supplementary Figure 4 | Overall survival from childhood cancer by age at diagnosis, for the entire SALiCCS population and by country (children diagnosed with cancer in 1971-2008 (DEN), 2009 (FIN), 2011 (SWE)).

Supplementary Figure 5 | Overall survival from childhood cancer by diagnostic group, for the entire SALiCCS population and by country (children diagnosed with cancer in 1971-2008 (DEN), 2009 (FIN), 2011 (SWE)).

Supplementary Figure 6 | Overall survival from childhood non-CNS solid tumours by tumour type, for the entire SALiCCS population and by country (children diagnosed with cancer in 1971-2008 (DEN), 2009 (FIN), 2011 (SWE)).

Supplementary Table 1 | Comparison of the distribution of diagnostic groups and specific cancer types between all incident childhood cancer cases and the SALiCCS 5-year childhood cancer survivors.

Supplementary Table 2 | Sociodemographic and socioeconomic characteristics of 5-year childhood cancer survivors, population comparisons and siblings by country.

REFERENCES

1. The European Society for Paediatric Oncology (SIOPE). *The SIOPE Strategic Plan - a European Cancer Plan for Children and Adolescents*. Brussels, Belgium (2015).
2. Schuz J, Erdmann F. Environmental Exposure and Risk of Childhood Leukemia: An Overview. *Arch Med Res* (2016) 47:607–14. doi: 10.1016/j.arcmed.2016.11.017
3. Spector LG, Pankratz N, Marcotte EL. Genetic and Nongenetic Risk Factors for Childhood Cancer. *Pediatr Clinics North Am* (2015) 62:11–25. doi: 10.1016/j.pcl.2014.09.013
4. Schmiegelow K, Forestier E, Hellebostad M, Heyman M, Kristinsson J, Soderhall S, et al. Long-Term Results of NOPHO ALL-92 and ALL-2000 Studies of Childhood Acute Lymphoblastic Leukemia. *Leukemia* (2010) 24:345–54. doi: 10.1038/leu.2009.251
5. Pritchard-Jones K, Pieters R, Reaman GH, Hjorth L, Downie P, Calaminus G, et al. Sustaining Innovation and Improvement in the Treatment of Childhood Cancer: Lessons From High-Income Countries. *Lancet Oncol* (2013) 14:e95–103. doi: 10.1016/S1470-2045(13)70010-X
6. Gatta G, Botta L, Rossi S, Aareleid T, Bielska-Lasota M, Clavel J, et al. Childhood Cancer Survival in Europe 1999–2007: Results of EUROCARE-5—A Population-Based Study. *Lancet Oncol* (2014) 15:35–47. doi: 10.1016/S1470-2045(13)70548-5
7. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Niksic M, et al. Global Surveillance of Trends in Cancer Survival 2000-14 (CONCORD-3):

Analysis of Individual Records for 37 513 025 Patients Diagnosed With One of 18 Cancers From 322 Population-Based Registries in 71 Countries. *Lancet* (2018) 391:1023–75. doi: 10.1016/S0140-6736(17)33326-3
8. Erdmann F, Frederiksen LE, Bonaventure A, Mader L, Hasle H, Robison LL, et al. Childhood Cancer: Survival, Treatment Modalities, Late Effects and Improvements Over Time. *Cancer Epidemiol* (2021) 71(Pt B):101733. doi: 10.1016/j.canep.2020.101733
9. Erdmann F, Ghantous A, Schüz J. Environmental Agents and Childhood Cancer. In: Nriagu J, editor. *Encyclopedia of Environmental Health*. 2. Burlington: Elsevier B.V. (2019). p. 336–47.
10. Bhatia S, Armenian SH, Armstrong GT, van Dulmen-den Broeder E, Hawkins MM, Kremer LC, et al. Collaborative Research in Childhood Cancer Survivorship: The Current Landscape. *J Clin Oncol* (2015) 33:3055–64. doi: 10.1200/JCO.2014.59.8052
11. Hudson MM, Armenian SH, Armstrong GT, Chow EJ, Henderson TO. Optimization of Health and Extension of Lifespan Through Childhood Cancer Survivorship Research. *J Clin Oncol* (2018) 36(21):2133–4. doi: 10.1200/JCO.2018.79.0477
12. Norsker FN, Pedersen C, Armstrong GT, Robison LL, McBride ML, Hawkins M, et al. Late Effects in Childhood Cancer Survivors: Early Studies, Survivor Cohorts, and Significant Contributions to the Field of Late Effects. *Pediatr Clinics North America* (2020) 67:1033–49. doi: 10.1016/j.pcl.2020.07.002
13. Oeffinger K, Mertens AC, Sklar CA, Kawashima T, Hudson MM, Meadows AT, et al. Chronic Health Conditions in Adult Survivors of Childhood Cancer. *N Engl J Med* (2006) 355(15):1572–82. doi: 10.1056/NEJMsa060185
14. Geenen MM, Cardous-Ubbink MC, Kremer LC, van den Bos C, van der Pal HJ, Heinen RC, et al. Medical Assessment of Adverse Health Outcomes in Long-Term Survivors of Childhood Cancer. *JAMA* (2007) 297:2705–15. doi: 10.1001/jama.297.24.2705
15. de Fine Licht S, Rugbjerg K, Gudmundsdottir T, Bonnesen TG, Asdahl PH, Holmqvist AS, et al. Long-Term Inpatient Disease Burden in the Adult Life After Childhood Cancer in Scandinavia (ALiCCS) Study: A Cohort Study of 21,297 Childhood Cancer Survivors. *PloS Med* (2017) 14:e1002296. doi: 10.1371/journal.pmed.1002296
16. Fidler MM, Reulen RC, Winter DL, Kelly J, Jenkinson HC, Skinner R, et al. Long Term Cause Specific Mortality Among 34 489 Five Year Survivors of Childhood Cancer in Great Britain: Population Based Cohort Study. *Bmj* (2016) 354:i4351. doi: 10.1136/bmj.i4351
17. Olsen JH, Moller T, Anderson H, Langmark F, Sankila R, Tryggvadottir L, et al. Lifelong Cancer Incidence in 47,697 Patients Treated for Childhood Cancer in the Nordic Countries. *J Natl Cancer Inst* (2009) 101:806–13. doi: 10.1093/jnci/djp104
18. Robison LL, Hudson MM. Survivors of Childhood and Adolescent Cancer: Life-Long Risks and Responsibilities. *Nat Rev Cancer* (2014) 14:61–70. doi: 10.1038/nrc3634
19. Lund LW, Winther JF, Dalton SO, Cederkvist L, Jeppesen P, Deltour I, et al. Hospital Contact for Mental Disorders in Survivors of Childhood Cancer and Their Siblings in Denmark: A Population-Based Cohort Study. *Lancet Oncol* (2013) 14:971–80. doi: 10.1016/S1470-2045(13)70351-6
20. Nathan PC, Nachman A, Sutradhar R, Kurdyak P, Pole JD, Lau C, et al. Adverse Mental Health Outcomes in a Population-Based Cohort of Survivors of Childhood Cancer. *Cancer* (2018) 124:2045–57. doi: 10.1002/cncr.31279
21. Lund LW, Winther JF, Cederkvist L, Andersen KK, Dalton SO, Appel CW, et al. Increased Risk of Antidepressant Use in Childhood Cancer Survivors: A Danish Population-Based Cohort Study. *Eur J Cancer* (2015) 51:675–84. doi: 10.1016/j.ejca.2015.01.001
22. Frederiksen LE, Mader L, Feychting M, Mogensen H, Madanat-Harjuoja L, Malila N, et al. Surviving Childhood Cancer: A Systematic Review of Studies on Risk and Determinants of Adverse Socioeconomic Outcomes. *Int J Cancer* (2019) 144(8):1796–823. doi: 10.1002/ijc.31789
23. Mader L, Michel G, Roser K. Unemployment Following Childhood Cancer. *Dtsch Arztebl Int* (2017) 114:805–12. doi: 10.3238/arztebl.2017.0805
24. Kirchhoff AC, Leisenring W, Krull KR, Ness KK, Friedman DL, Armstrong GT, et al. Unemployment Among Adult Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *Med Care* (2010) 48:1015–25. doi: 10.1097/MLR.0b013e3181eaf880
25. Pang JW, Friedman DL, Whitton JA, Stovall M, Mertens AC, Robison LL, et al. Employment Status Among Adult Survivors in the Childhood Cancer Survivor Study. *Pediatr Blood Cancer* (2008) 50:104–10. doi: 10.1002/pbc.21226
26. Erdmann F, Feychting M, Mogensen H, Schmiegelow K, Zeeb H. Social Inequalities Along the Childhood Cancer Continuum: An Overview of Evidence and a Conceptual Framework to Identify Underlying Mechanisms and Pathways. *Front Public Health* (2019) 7:84. doi: 10.3389/fpubh.2019.00084
27. Nordic Co-operation. *Nordic Statistics Database*. (2021).
28. Esping-Andersen G. *The Three Worlds of Welfare Capitalism*. Princeton, New Jersey: Princeton University Press (1990).
29. OECD. *Revenue Statistics 2018 - Tax Revenue Trends in the OECD* (2018). Available at: http://www.oecd.org/tax/revenue-statistics-2522770x.htm.
30. OECD. *Life Expectancy at Birth (Indicator), Both Genders 1960-2017*. (2019). doi: 10.1787/27e0fc9d-en
31. OECD. *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing (2018). doi: 10.1787/eag-2018-en.
32. Agerskov U, Mertsola SN.C.o. *Ministers, Nordic Statistics 2018*. Copenhagen, Denmark: Nordic Council of Ministers (2018). Available at: http://norden.diva-portal.org/smash/get/diva2:1257993/FULLTEXT01.pdf.
33. OECD. *Labour Force (Indicator) - Labour Force Women, 1959-2017*. (2019). doi: 10.1787/ef2e7159-en
34. OECD. *Family Benefits Public Spending (Indicator), 1980-2017*. (2019). doi: 10.1787/8e8b3273-en
35. Lammi-Taskula J, Brandth B, Duvander A-Z, V. Gíslason I, Björk Eydal G, Rostgaard T. *Parental Leave, Childcare and Gender Equality in the Nordic Countries*. Copenhagen: Nordic Council of Ministers (2012). doi: 10.6027/TN2011-562
36. Olejaz M, Nielsen AJ, Rudkjøbing A, Birk HO, Krasnik A, Hernandez-Quevedo C. *Denmark - Health System Review, Health Systems in Transition*. Copenhagen: World Health Organization (2012). 14(2):1–192.
37. Annel A, Glenngaard AH, Merkur S. *Sweden - Health System Review, Health Systems in Transition*. Vol. 14, no. 5. Copenhagen: World Health Organization (2012). p. 1–159.
38. Vuorenkoski L. *Finland - Health System Review, Health Systems in Transition*. Vol. 10, no. 4. Copenhagen: World Health Organization (2008). p. 1–168.
39. OECD. *Health a Glance 2017 - OECD Indicators*. Paris: OECD Publishing (2017). doi: 10.1787/health_glance-2017-en
40. Lie SO, Abrahamsson J, Clausen N, Forestier E, Hasle H, Hovi L, et al. Long-Term Results in Children With AML: NOPHO-AML Study Group–Report of Three Consecutive Trials. *Leukemia* (2005) 19:2090–100. doi: 10.1038/sj.leu.2403962
41. Márky I, Björk O, Forestier E, Jónsson ÓG, Perkkiö M, Schmiegelow K, et al. Intensive Chemotherapy Without Radiotherapy Gives More Than 85% Event-Free Survival for Non-Hodgkin Lymphoma Without Central Nervous Involvement: A 6-Year Population-Based Study From the Nordic Society of Pediatric Hematology and Oncology. *J Pediatr Hematol/ Oncol* (2004) 26:555–60. doi: 10.1097/01.mph.0000139772.98685.d2
42. Pui CH, Carroll WL, Meshinchi S, Arceci RJ. Biology, Risk Stratification, and Therapy of Pediatric Acute Leukemias: An Update. *J Clin Oncol* (2011) 29:551–65. doi: 10.1200/JCO.2010.30.7405
43. Gustafson G, Kreuger A, Clausen N, Garwicz S, Kristinsson J, Lie S, et al. Nordic Society of Paediatric Haematology and Oncology (NOPHO), Intensified Treatment of Acute Childhood Lymphoblastic Leukaemia has Improved Prognosis, Especially in Non-High-Risk Patients: The Nordic Experience of 2648 Patients Diagnosed Between 1981 and 1996. Nordic Society of Paediatric Haematology and Oncology (NOPHO). *Acta Paediatrica* (1998) 87:1151–61. doi: 10.1080/080352598750031149
44. Asdahl PH, Winther JF, Bonnesen TG, De Fine Licht S, Gudmundsdottir T, Anderson H, et al. The Adult Life After Childhood Cancer in Scandinavia (ALiCCS) Study: Design and Characteristics. *Pediatr Blood Cancer* (2015) 62:2204–10. doi: 10.1002/pbc.25661
45. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health* (2011) 39:22–5. doi: 10.1177/1403494810387965
46. Ludvigsson JF, Almqvist C, Bonamy AK, Ljung R, Michaelsson K, Neovius M, et al. Registers of the Swedish Total Population and Their Use in Medical Research. *Eur J Epidemiol* (2016) 31:125–36. doi: 10.1007/s10654-016-0117-y
47. Digital and Population Data Services Agency. *Population Information System*. Finland: Digital and Population Data Services Agency (2021).

48. Pukkala E, Engholm G, Hojsgaard Schmidt LK, Storm H, Khan S, Lambe M, et al. Nordic Cancer Registries - An Overview of Their Procedures and Data Comparability. *Acta Oncol* (2018) 57:440–55. doi: 10.1080/0284186X.2017.1407039
49. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health* (2011) 39:30–3. doi: 10.1177/1403494811401482
50. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. *Scand J Public Health* (2011) 39:54–7. doi: 10.1177/1403494810395825
51. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim J-L, Reuterwall C, et al. External Review and Validation of the Swedish National Inpatient Register. *BMC Public Health* (2011) 11:450. doi: 10.1186/1471-2458-11-450
52. Sund R. Quality of the Finnish Hospital Discharge Register: A Systematic Review. *Scand J Public Health* (2012) 40:505–15. doi: 10.1177/1403494812456637
53. Brooke HL, Talback M, Hornblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish Cause of Death Register. *Eur J Epidemiol* (2017) 32:765–73. doi: 10.1007/s10654-017-0316-1
54. Bliddal M, Broe A, Pottegard A, Olsen J, Langhoff-Roos J. The Danish Medical Birth Register. *Eur J Epidemiol* (2018) 33:27–36. doi: 10.1007/s10654-018-0356-1
55. Cnattingius S, Ericson A, Gunnarskog J, Källen B. A Quality Study of a Medical Birth Registry. *Scand J Soc Med* (1990) 18:143–8. doi: 10.1177/140349489001800209
56. Gissler M, Louhiala P, Hemminki K. Nordic Medical Birth Registers in Epidemiological Research. *Eur J Epidemiol* (1997) 13:169–75. doi: 10.1023/A:1007379029182
57. Thygesen LC, Daasnes C, Thaulow I, Bronnum-Hansen H. Introduction to Danish (Nationwide) Registers on Health and Social Issues: Structure, Access, Legislation, and Archiving. *Scand J Public Health* (2011) 39:12–6. doi: 10.1177/1403494811399956
58. Jensen VM, Rasmussen AW. Danish Education Registers. *Scand J Public Health* (2011) 39:91–4. doi: 10.1177/1403494810394715
59. Baadsgaard M, Quitzau J. Danish Registers on Personal Income and Transfer Payments. *Scand J Public Health* (2011) 39:103–5. doi: 10.1177/1403494811405098
60. Statistics Sweden. *Background Facts – Labour and Education Statistics 2016*. Stockholm, Sweden (2019).
61. Statistics Sweden. *Population and Housing Census 1960-1990 (TPR)*. Stockholm, Sweden (2000).
62. Ludvigsson JF, Svedberg P, Olen O, Bruze G, Neovius M. The Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA) and Its Use in Medical Research. *Eur J Epidemiol* (2019) 34:423–37. doi: 10.1007/s10654-019-00511-8
63. Official Statistics of Finland (OSF). *Population Structure [E-Publication]*. Helsinki: Statistics Finland (2021).
64. Official Statistics of Finland (OSF). *Educational Structure of Population [E-Publication]*. Helsinki: Statistics Finland (2021).
65. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish Personal Identity Number: Possibilities and Pitfalls in Healthcare and Medical Research. *Eur J Epidemiol* (2009) 24:659–67. doi: 10.1007/s10654-009-9350-y
66. Birch JM, Mardsen HB. Classification Scheme for Childhood Cancer. *Int J Cancer* (1987) 40:620–4. doi: 10.1002/ijc.2910400508
67. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, Third Edition. *Cancer* (2005) 103:1457–67. doi: 10.1002/cncr.20910
68. Kramarova E, Stiller CA. The International Classification of Childhood Cancer. *Int J Cancer* (1996) 68:759–65. doi: 10.1002/(SICI)1097-0215(19961211)68:6<759::AID-IJC12>3.0.CO;2-W
69. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al. *International Classification of Diseases for Oncology. 3rd ed*. Geneva: World Health Organisation (2000).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 01 March 2022
doi: 10.3389/fonc.2022.817045

Effect of Time Since Smoking Cessation on Lung Cancer Incidence: An Occupational Cohort With 27 Follow-Up Years

Zheng Su[1], Xin-Hua Jia[1,2], Fang-Hui Zhao[1], Qing-Hua Zhou[3,4], Ya-Guang Fan[4]* and You-Lin Qiao[5]*

[1] Department of Epidemiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] The State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, National Institute of Diagnostics and Vaccine Development in Infectious Diseases, School of Public Health, Xiamen University, Xiamen, China, [3] Sichuan Lung Cancer Institute, Sichuan Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [4] Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, China, [5] Center for Global Health, School of Population Medicine and Public Health Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Marilyn Klug, University of North Dakota, United States
Sungshim Lani Park, University of Hawaii at Manoa, United States

***Correspondence:**
Ya-Guang Fan
fanyaguang75@163.com
Fang-Hui Zhao
zhaofangh@cicams.ac.cn

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: 17 November 2021
Accepted: 28 January 2022
Published: 01 March 2022

Citation:
Su Z, Jia X-H, Zhao F-H, Zhou Q-H, Fan Y-G and Qiao Y-L (2022) Effect of Time Since Smoking Cessation on Lung Cancer Incidence: An Occupational Cohort With 27 Follow-Up Years. Front. Oncol. 12:817045. doi: 10.3389/fonc.2022.817045

Background: This special cohort reveals the effect of smoking cessation in occupational miners exposed to radon and arsenic.

Methods: A total of 9,134 tin miners with at least 10 years of underground radon and arsenic exposure were enrolled beginning in 1992 and followed for up to 27 years. Detailed smoking information was collected at baseline, and information on smoking status was consecutively collected from 1992 to 1996. The Cox proportional hazards model was used to explore the relationship between time since smoking cessation and lung cancer.

Results: A total of 1,324 lung cancer cases occurred in this cohort over 167,776 person-years of follow-up. Among populations exposed to radon and arsenic, miners after quitting smoking for 10 years or more had almost halved their lung cancer risk [adjusted hazard ratio (HR) = 0.55, 95% CI: 0.38–0.79], compared with current smokers. Among miners after quitting smoking for 5 years or more, lung cancer incidence approximately halved (HR = 0.52, 95% CI: 0.30–0.92) for squamous cell lung carcinoma, while it showed no significant decline for adenocarcinoma (HR = 0.79, 95% CI: 0.34–1.85).

Conclusion: Smoking cessation for 10 years or more halved lung cancer incidence among miners exposed to radon and arsenic, and the benefit was more pronounced among squamous cell lung carcinoma.

Keywords: smoking cessation, lung cancer, cohort, radon, arsenic

WHAT THIS PAPER ADDS

What is already known about this subject?

Preliminary studies suggest that there exist joint effects between radon, arsenic, and smoking. Quitting smoking reduces not only smoking-related lung cancer, but also smoking-radon- and smoking-arsenic-related lung cancer. However, there is no prospective cohort to report the effect of years of smoking cessation on lung cancer incidence among miners exposed to radon and arsenic.

What are the new findings?

Among miners exposed to radon and arsenic, smoking cessation of at least 10 years would halve lung cancer incidence, and the benefit was more related to squamous cell lung carcinoma.

How might it impact policy in the foreseeable future?

To reduce the burden of lung cancer, smoking cessation is urgently needed among radon and arsenic miners. The longer years of smoking cessation should be emphasized among occupational miners than the general population.

BACKGROUND

Lung cancer remains the most common cancer with respect to both incidence and mortality, both in China and throughout the world (1, 2). Tobacco smoking is the leading risk factor for lung cancer, but other factors related to lung cancer include environmental tobacco smoke, air pollution, occupational exposures, marijuana, and other recreational drugs (3–6). Either radon or arsenic exposure is evident to be the major occupational carcinogens of lung cancer concluded by the International Agency for Research on Cancer (7, 8).

While it is clear that smoking cessation reduces lung cancer risk in general populations, this topic has been rarely investigated in occupational epidemiological studies. Population-based studies suggested a sharp decrease in lung cancer risk for over 50% in the first 5 years after smoking cessation (9, 10). However, findings from general population-based studies may not be directly generalized to occupational studies, because occupational workers additionally exposed to lung carcinogens generally have a higher lung cancer risk, and there exists a joint effect between cigarette use and occupational agents such as radon and arsenic. Consequently, the effect of smoking cessation on lung cancer tends to need a much longer time in occupational groups. A historical cohort of Chinese silicotics revealed that smoking cessation for 10 years halved lung cancer mortality among silicotics (11). Similarly, an asbestos-exposed cohort showed that lung cancer mortality rate ratio dropped steeply (over 50%) during the first 10 years after quitting smoking (12).

Globally, no studies to date have reported the effect of smoking cessation on lung cancer in occupational populations exposed to radon and arsenic. Among occupational radon cohorts, although Colorado Plateau cohort and German uranium miners collected individual smoking data, they have not revealed the effect of smoking cessation, and other cohorts generally lacked complete smoking information (13–15). Similarly, for occupational arsenic cohorts, most epidemiological studies of copper smelters in Utah, Sweden, Montana, and the United States have not reported the role of smoking cessation on lung cancer (16–19).

Among Chinese Tin miner studies, the results from case–control and cohort studies for several decades have identified that radon, arsenic, and smoking are the main risk factors for lung cancer (20–22). In addition, individual exposure information about radon, arsenic, and smoking was collected in our cohort. Therefore, it provided us a unique opportunity to investigate the effect of smoking cessation on lung cancer in workers exposed to radon and arsenic.

METHODS

Study Design and Participants

The design and inclusion criteria of the Yunnan Tin Corporation (YTC) cohort was described previously (22, 23). Briefly, a total of 9,295 tin miners ≥ 40 years old who had 10 or more years of underground radon and/or arsenic exposure have been dynamically included into this cohort since 1992. All participants were followed by December 31, 2018. A total of 161 former smokers who restarted smoking from 1992 to 1996 were excluded. Then, 9,134 miners were included into the final analysis to estimate the risk of lung cancer incidence according to years since smoking cessation. In addition, as 599 women in this study were almost never smokers, all women were excluded. Finally, 8,535 male miners were included into a sensitivity analysis that was used to assess the robustness of analysis (**Figure 1**).

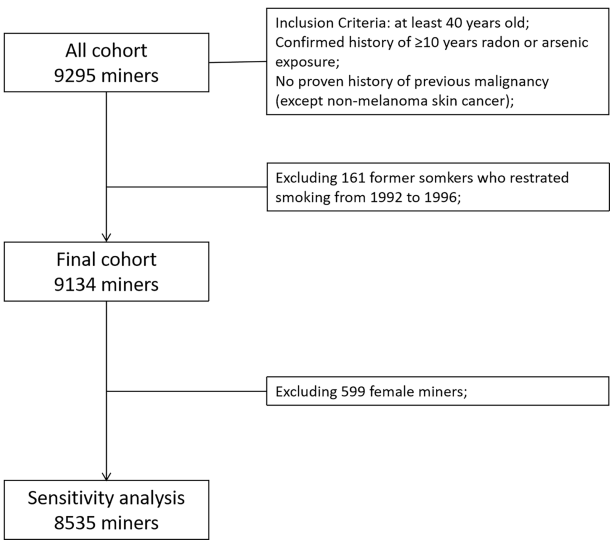


FIGURE 1 | Flow chart of selections of the participants included in the statistical analysis.

Exposure Assessment

Tobacco Use

All miners were enrolled at baseline (in 1992) and information on age of start/stop smoking, type of tobacco (cigarette, waterpipe, and long-stem pipe), and smoking status was collected. According to the smoking status at baseline, we divided miners into smokers, former smokers, and never smokers. At baseline, individuals who had smoked regularly for 6 months or longer at any time in their lives were classified as smokers, and those who have a smoking duration of less than 6 months were considered never smokers and smokers who ceased smoking at enrollment were former smokers. In addition, to test the stability of smoking status, for each miner, information on smoking status and type of tobacco was collected for five consecutive years from 1992 to 1996. The change of smoking status for at least two consecutive years was identified as "the real behavior change" from 1992 to 1996. A total of 161 former smokers who restarted smoking were excluded.

However, the impact of type of cigarette was not considered in this study, because most of the participants were mixed smokers: at baseline, 62.4% (4,306/6,899) used cigarette + waterpipe, 29.3% (2,022/6,899) used cigarette only, and 8.3% (571/6,899) used waterpipe only among current smokers; 53.6% (1,084/2,022) of cigarette-only and 53.6% (306/571) of waterpipe-only smokers became mixed smokers in the next 4 years. Thus, the impact of type of cigarette was not considered in the final analysis.

Therefore, based on previous findings, we calculated a cigarette-equivalent variable adjusting conservatively: 1 g water pipe = 1 cigarette. Smoking intensity was measured by the number of cigarettes smoked per day and pack-years were calculated as the average number of cigarettes per day (divided by 20) times the number of years of smoking. Finally, smoking pack-years as a continuous variable was included into the multiple Cox proportional hazards model.

Occupational Carcinogens

Detailed definitions of occupational radon and arsenic exposures were given elsewhere (22, 23). In this study, according to total radon exposure, participants were classified into three groups: low group: <100 cumulative working level month (WLM), medium group: ≥100 and <400 WLM, and high group: ≥400 WLM. On the other hand, according to total arsenic exposure, participants were also classified into three groups: low group: <40 mg/m^3, medium group: ≥40 and <100 mg/m^3, and high group: ≥100 mg/m^3. Finally, we combined these groups into three new groups named occupational exposure groups: lowly exposed group: low radon group and low arsenic groups; moderately exposed group: low radon–medium arsenic, medium radon–low arsenic, and medium radon–medium arsenic; highly exposed group: either high radon group or high arsenic group (a total of 5 subgroups: low radon–high arsenic, medium radon–high arsenic, high radon–high arsenic, high radon–low arsenic, and high radon–medium arsenic). In addition, other information including age, sex, education level, and prior lung disease was also collected for each participant.

Follow-Up and Case Ascertainment

From 1992 to 1999, annual follow-up was conducted combined with screening by chest radiography and sputum cytology. In 2005 and 2006, the first post-screening follow-up was performed and participants were followed until December 31, 2001. In 2019, an extended follow-up was conducted, and the survival status, total cause of death, cancer diagnosis, and death information were collected.

The primary outcome was lung cancer incidence, which was from the local cancer registration agency, medical record system, and funeral parlor, and face-to-face interviews with relatives and workmates of the participants. In the process of information extraction, participants' name, age, work units, and home address were taken into consideration. By the end of December 31, 2018, 187 participants (2.1%) were lost to follow-up, with a follow-up rate of 97.9%.

Lung cancer cases were confirmed *via* the following ways: (1) histology (from surgical resection tissue or biopsy); (2) cytology (from sputum sample or endoscopy brushing); (3) x-ray; and (4) others (e.g., death certificate listing only without other information).

Statistical Analysis

Person-years of follow-up were calculated from the date of enrollment to the date of lung cancer incidence or date of death or censoring as of December 31, 2018. Descriptive statistics was used to show various characteristics among never, former, and current smokers at the baseline. The association between time since smoking cessation and risk of lung cancer incidence was analyzed by the Cox proportional hazards model. Schoenfeld residuals were used to check the proportional hazards assumption. To control the effect of occupational exposures (radon and arsenic), Cox proportional hazards model was performed to calculate the hazard ratio (HR) and 95% confidence interval (95% CI) in three different occupational exposure subgroups (low, medium, and high groups). Furthermore, to reduce potential residual confounding by occupational exposure, we also modeled occupational exposure as continuous variables into the statistical analysis within each exposed stratum. In addition, in the multiple Cox proportional hazards model, other variables including age at entry, sex, education level, family history of lung cancer in first-degree relatives, and prior lung disease (silicosis, tuberculosis, asthma, and chronic bronchitis) were adjusted to eliminate confounding effects. Finally, since the loss of follow-up rates and missing data were very low, those with missing values were not included in the analysis. SAS and R software were used for statistical analysis.

RESULTS

The characteristics of the YTC miners in subgroups of smoking status at baseline are shown in **Table 1**. A total of 6,899 (75.5%) of them were current smokers, 772 (8.5%) were former smokers, and 1,463 (16.0%) were never smokers. The number of lung

cancer cases and person-years were 1,075 and 124,797.3 in current smokers, 121 and 12,087 in former smokers, and 128 and 30,891.0 in never smokers, respectively. Almost all women (99.7%) were never smokers, and never smokers had the youngest age (mean = 49, IQR = 41–55) at enrollment. Smoking (number of pack-years in life time) in current smokers [mean = 26.4, interquartile range (IQR) = 14.2–34.5] was significantly higher than that in former smokers (mean = 18.7, IQR = 5.9–25.5). Compared with the non-smokers, current smokers had lower educational level, more prone to having prior lung disease (silicosis, tuberculosis, asthma, and chronic bronchitis), and higher occupational exposure.

The evaluation of the stability of smoking status for 5 consecutive years in the YTC cohort is shown in **Table 2**. From current smokers to former smokers, there were only 1.6% (110/6,899); from never smokers to current smokers, there were only 2.1% (31/1,463). Among former smokers, the stability of smoking status varied significantly with increasing years since smoking cessation: the rates of quitting successfully were 57.6% (68/118) in those who quit smoking at enrollment, 74.0% (91/123) in 1 year after quitting, 77.2% (129/167) in 2–5 years after quitting, and 92.2% (484/525) in more than 5 years after cessation.

Table 3 shows the effects of years of smoking cessation on lung cancer incidence among miners exposed to radon and arsenic. Generally, a significantly negative gradient ($p < 0.001$ for trend test) of lung cancer incidence was observed with increasing years of smoking cessation for all former smokers, despite the fact that significant risk reduction did not manifest within the first 1 year (HR = 1.03, 95% CI: 0.70–1.51), 2–5 years (HR = 0.85, 95% CI: 0.56–1.30), and 6–10 years (HR = 0.66, 95% CI: 0.43–1.03) of cessation. Furthermore, the risk of lung cancer incidence was nearly halved for 10+ years (HR = 0.55, 95% CI: 0.38–0.79). Furthermore, we observed the effect of smoking cessation stratified by radon and arsenic exposure. Among miners from the highly exposed group, they showed similar patterns to all miners, which is shown in **Table 3**. Given the low sample size in the lowly and moderately exposed group, data are shown in **Table S1** and the risk of lung cancer incidence among the lowly exposed group decreased by 50% (HR = 0.50, 95% CI: 0.24–1.16) within 5 years since cessation, and it further decreased by 65% (HR = 0.45, 95% CI: 0.11–1.89) if the smokers continued

TABLE 1 | Characteristics of the YTC miners in subgroups of smoking status at baseline.

Characteristics	All subjects	Smoking status at baseline			*p*-value
		Current smokers	Former smokers	Never smokers	
No. of subjects	9,134	6,899 (75.5%)	772 (8.5%)	1,463 (16.0%)	
Age (years)					<0.01
Mean (IQR)	53 (43–61)	53 (43–61)	59 (53–66)	49 (41–55)	
Gender					<0.01
Male	8,535	6,897 (80.8%)	772 (9.1%)	866 (10.2%)	
Female	599	2 (0.3%)	0 (0.0%)	597 (99.7%)	
Education					<0.01
No	2,155	1,771 (82.2%)	223 (10.4%)	161 (7.4%)	
≤6 years	4,384	3,429 (78.2%)	354 (8.1%)	601 (13.7%)	
>6 years	2,595	1,699 (65.5%)	195 (7.5%)	701 (27.0%)	
Family History of Lung Cancer					<0.01
Yes	631	454 (71.0%)	40 (6.3%)	137 (21.7%)	
No	8,503	6,445 (75.8%)	732 (8.6%)	1,326 (15.6%)	
Silicosis					<0.01
Yes	444	349 (78.6%)	70 (15.8%)	25 (5.6%)	
No	8,690	6,550 (75.3%)	702 (8.1%)	1,438 (16.6%)	
Tuberculosis					<0.01
Yes	260	181 (69.6%)	40 (15.4%)	39 (15.0%)	
No	8,874	6,718 (75.7%)	732 (8.2%)	1,424 (16.1%)	
Asthma					<0.01
Yes	649	481 (74.1%)	119 (18.3%)	49 (7.6%)	
No	8,485	6,418 (75.6%)	653 (7.7%)	1,414 (16.7%)	
Chronic bronchitis					<0.01
Yes	2,363	1,856 (78.5%)	302 (12.8%)	205 (8.7%)	
No	6,771	5,043 (74.5%)	470 (6.9%)	1,258 (18.6%)	
Pack-Years					<0.01
Mean (IQR)	19.1 (7.8–31.0)	26.4 (14.2–34.5)	18.7 (5.9–25.5)	—	
Occupational Exposure					<0.01
Low group	1,143	690 (60.4%)	51 (4.5%)	402 (35.1%)	
Medium group	2,998	2,243 (74.8%)	201 (6.7%)	554 (18.5%)	
High group	4,993	3,966 (79.4%)	520 (10.4%)	507 (10.2%)	
No. of Lung Cancer	1,324	1,075 (81.2%)	121 (9.1%)	128 (9.7%)	
No. of Person-Years	167,776	124,797	12,314	30,891	

Values were given as n (%) for categorical variables, IQR (Q1Q3): interquartile range. p-value: the differences between the proportions were tested by Chi-square test, and the mean differences were tested by ANOVA between the subgroups. Mann–Whitney U test was carried out for the non-normal distribution data.

TABLE 2 | Smoking status in five consecutive years from 1992 to 1996 in the YTC cohort.

The initial 4 years of follow-up				
	Smoking	**Non-Smoking**	**Total**	**Relapse rate (%)[a]**
At baseline				
Current smokers	6,789	110	6,899	1.6
Never smokers	31	1,432	1,463	2.1
Years since cessation			933	
<1	50	68	118	42.4
1	32	91	123	26.0
2-5	38	129	167	22.8
>5	41	484	525	7.8

[a]*Relapse rate: During the initial 4 years of follow-up, smoking status for persistent change, that is, a change in status that remained for at least 2 years: current smokers quit, never smokers started smoking, and former smokers returned to smoking.*

to abstain from cigarette smoking for 5 years or more. The beneficial effect was nearly similar in the moderately exposed group.

As most women in our cohort were never smokers, a sensitivity test had been conducted and indicated that our analysis was robust (**Table S2**). We further analyzed the lung cancer incidence risk in relation to years of smoking cessation by histological type among the highly exposed group, as shown in **Table 4**. To reduce potential residual confounding by occupational exposure, we excluded miners in lowly and moderately exposed groups based on the quite low sample size of lung cancer cases. Results showed that for squamous cell carcinoma (SQC), the risks showed a significantly

TABLE 3 | Hazard ratio (HR, 95% confidence interval) of lung cancer incidence according to years of smoking cessation among the YTC miners.

Variable	No. of subjects	No. of lung cancer	No. of Person-Years	No. of lung cancer/Person-Years × 10^4	Crude HR (95% CI)	Age-/Sex-Adjusted HR (95% CI)	Full-Adjusted HR (95% CI)[a]
All miners without stratification							
Never Smokers	1,463	128	30,869.7	41.5	0.46 (0.39, 0.56)	0.53 (0.42, 0.67)	0.67 (0.52, 0.85)
Years since cessation	772	100	12,078.9	82.8			
≤1	159	27	2,132.7	126.6	1.49 (1.02, 2.18)	0.98 (0.67, 1.44)	1.03 (0.70, 1.51)
2–5	152	22	2,337.1	94.1	1.08 (0.71, 1.65)	0.85 (0.56, 1.30)	0.85 (0.56, 1.30)
6–10	148	20	2,355.1	84.9	0.98 (0.63, 1.52)	0.68 (0.44, 1.06)	0.66 (0.43, 1.03)
>10	313	31	5,254.0	59.0	0.67 (0.47, 0.95)	0.48 (0.33, 0.68)	0.55 (0.38, 0.79)
Current Smokers	6,899	1,096	124,708.7	87.9	1	1	1
Highly exposed group							
Never Smokers	507	61	9,953.4	61.3	0.49 (0.38, 0.64)	0.57 (0.43, 0.76)	0.71 (0.52, 0.96)
Years since cessation	520	81	7,318.1	110.7			
≤1	97	20	1,104.9	181.0	1.55 (1.00, 2.42)	1.14 (0.73, 1.79)	1.10 (0.70, 1.71)
2–5	107	21	1,473.0	142.6	1.17 (0.76, 1.80)	0.97 (0.63, 1.50)	0.98 (0.63, 1.52)
6–10	103	17	1,478.7	115.0	0.95 (0.59, 1.53)	0.73 (0.45, 1.18)	0.70 (0.43, 1.14)
>10	213	23	3,261.6	70.5	0.58 (0.38, 0.87)	0.46 (0.30, 0.69)	0.53 (0.35, 0.80)
Current Smokers	3,966	803	65,478.1	122.6	1	1	1

[a]*Multiple Cox proportional hazards models were adjusted for age, gender, education, family history of lung cancer, silicosis, tuberculosis, asthma, chronic bronchitis, radon, arsenic, and smoking pack-years. Lowly Exposed Group: low radon and low arsenic; Moderately Exposed Group: low radon–medium arsenic, medium radon–low arsenic, and medium radon–medium arsenic; Highly Exposed Group: high arsenic–low radon, high arsenic–medium radon, high radon–high arsenic, high radon–low arsenic, and high radon–medium arsenic. Radon exposure: low radon: <100 cumulative working level month (WLM), medium radon: ≥100 and <400 WLM, high radon: ≥400 WLM; Arsenic exposure: low arsenic: <40 mg/m^3, medium arsenic: ≥40 and <100 mg/m^3, high arsenic: ≥100 mg/m^3.*

TABLE 4 | Hazard ratio (HR, 95% confidence interval) of lung cancer incidence according to years since smoking cessation by histologic types among highly exposed group.

Variable	No. of subjects	No. of lung cancer	No. of Person-Years	No. of lung cancer/Person-Years × 10^4	Crude HR (95% CI)	Age-/Sex-Adjusted HR (95% CI)	Full-Adjusted HR (95% CI) [a]
Squamous Cell Carcinoma							
Never Smokers	507	13	9,953.4	13.1	0.34 (0.20, 0.60)	0.51 (0.29, 0.89)	0.56 (0.32, 0.99)
Years since cessation	520	26	7,318.1	35.5			
≤5	204	13	2,577.8	50.4	1.24 (0.71, 2.16)	0.94 (0.54, 1.65)	0.92 (0.52, 1.61)
>5	316	13	4,740.3	27.4	0.69 (0.40, 1.21)	0.52 (0.30, 0.91)	0.52 (0.30, 0.92)
Current Smokers	3,966	257	65,478.1	39.2	1	1	1
Adenocarcinoma							
Never Smokers	507	5	9,953.4	5.0	0.37 (0.15, 0.91)	0.55 (0.22, 1.35)	0.72 (0.28, 1.89)
Years since cessation	520	11	7,318.1	15.0			
≤5	204	5	2,577.8	19.4	1.47 (0.60, 3.61)	1.12 (0.45, 2.77)	1.16 (0.47, 2.89)
>5	316	6	4,740.3	12.7	0.95 (0.42, 2.18)	0.71 (0.31, 1.62)	0.79 (0.34, 1.85)
Current Smokers	3,966	88	65,478.1	13.4	1	1	1

[a]*Multiple Cox proportional hazards models were adjusted for age, gender, education, family history of lung cancer, silicosis, tuberculosis, asthma, chronic bronchitis, radon, arsenic, and smoking pack-years. Lowly Exposed Group: low radon and low arsenic; Moderately Exposed Group: low radon–medium arsenic, medium radon–low arsenic, and medium radon–medium arsenic; Highly Exposed Group: high arsenic–low radon, high arsenic–medium radon, high radon–high arsenic, high radon–low arsenic, and high radon–medium arsenic. Radon exposure: low radon: <100 cumulative working level month (WLM), medium radon: ≥100 and <400 WLM, high radon: ≥400 WLM; Arsenic exposure: low arsenic: <40 mg/m^3, medium arsenic: ≥40 and <100 mg/m^3, high arsenic: ≥100 mg/m^3.*

decreasing trend with increasing time since cessation ($p < 0.001$ for trend test), and finally, the risk in those quitting for over 5 years was 0.52 (95% CI: 0.30-0.92), which was similar to the risk in never smokers (HR = 0.56, 95% CI: 0.32–0.99).

DISCUSSION

The YTC cohort with about 27 years of follow-up firstly reported that smoking cessation was associated with a substantial reduction in lung cancer incidence among underground miners exposed to radon and arsenic. For all lung cancer, about a 50% decrease in the risk of lung cancer incidence was shown in nearly 10 years for miners exposed to radon and arsenic. In addition, the long-term beneficial effect was weakened for adenocarcinoma, compared with squamous cell carcinoma.

In male participants among YTC miners, the risk of lung cancer among current smokers was lower than that of Western countries, Japan, Hong Kong, and the rest of mainland China (24–26). It could be due to the fact that the underground mines were relatively closed spaces and the high smoking rate of miners likely resulted in significant passive smoking exposure, even for non-smokers. Therefore, risk in non-smokers working underground in the mines was probably higher than that in non-smokers working in a less confined environment, and use of non-smokers with such exposure misclassification as a reference group would bias the relative risk estimates downward (toward the null). On the other hand, it may be that the effect of smoking is lower due to effect of these occupational exposures. Compelling evidence showed that there was a sub-multiplicative joint effect between occupational exposures (either radon or arsenic exposure) and smoking.

In addition, our results showed that the longer the smoking cessation time at baseline, the lower relapse rates for former smokers during the follow-up period. In cohort studies, smoking status at baseline might change over time, which resulted in bias in the true association between exposure (smoking status) and outcome (lung cancer risks). Therefore, a definition of smoking cessation with stable relapse rate was crucial in the cohort study, and our data showed that the relapse rate of quitting smoking at least 5+ years was as low as 7.8%. Because the definition of former smoker varied by study—quit smoking at least 6 months+ (27), 1+ years (28), 2+ years (10, 29, 30), or 5+ years (9)—more studies were needed to explore the optimal definition of former smokers.

Existing lines of evidence have illustrated a definitive benefit of smoking cessation in relation to lung cancer risks. Notably, smoking cessation is associated with a decrease in relative risk of lung cancer in former smokers compared to current smokers, but the absolute lung cancer risk in former smokers does not decrease from smoking cessation. However, the temporal pattern of this risk after smoking cessation is still controversial. In a previous case–control study in Hong Kong, Lap et al. observed that compared to current smokers, there is a rapid

decrease in lung cancer risk across most histological types of lung cancer within the first 5 years of quitting, and then it almost remained constant. However, Sadik et al. had conducted a meta-analysis and found that a continued progressive reduction in lung cancer risk resulting from smoking cessation would remain at least 15 years (31). In addition, Paul et al. had conducted a pooled analysis and found that relative risk of lung cancer to the current smokers decreased gradually and continuously over years of smoking cessation (32). Similarly, data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial showed that in 30+ pack-year former smokers, smoking abstinence resulted in a gradual decrease in the risk of lung cancer death (33). Results from the National Lung Screening Trial also observed a steady decline in lung cancer death risk with the increase in duration of tobacco abstinence (34).

Generally, most epidemiological studies regarding smoking cessation and lung cancer risk were conducted in the general population, and it has been rarely investigated in occupational epidemiological studies. It appeared that this delayed decrease in lung cancer risk was more common among individuals with occupational exposure. A cohort study conducted among Australian workers exposed to asbestos found that the lung cancer mortality rate ratio among insulators dropped steeply during the first 10 years after quitting smoking (12). A large historical cohort of Chinese silicotics showed that the risk of lung cancer mortality among all silicotics was nearly halved within 20 years since cessation (adjusted HR = 0.54, 95% CI: 0.35–0.83) (11). It is well known that there exist joint effects between radon, arsenic, and smoking, and quitting smoking reduces not only smoking-related lung cancer, but also smoking-radon- and smoking-arsenic-related lung cancer (13, 35, 36). However, to our knowledge, there is no study reporting the effect of smoking cessation on radon- and arsenic exposed populations. Among the YTC miners, our results firstly showed that the benefits of smoking cessation were different in occupational groups exposed to radon and arsenic. For the lowly exposed group, a rapidly decreasing lung cancer incidence risk was shown within the first 5 years of smoking cessation, which was consistent with the moderately exposed group. However, it seemed to take longer years of smoking cessation to achieve the same reduction among the highly exposed group. Importantly, findings from lowly and moderately exposed groups should be viewed as tentative given the low sample size in these two groups, and more studies would be encouraged to focus on this field in the future.

In the YTC, the risk of SQC incidence among workers highly exposed to radon and arsenic was nearly halved after 15 years or more since cessation, but the reduction was smaller for ADC. The benefit of smoking cessation was more prominent for SQC, which was consistent with findings from other studies (10, 24, 27, 37). However, a case–control study in Chinese men showed that the relative risk for SQC decreased by 78% (95% CI: 22%–94%) after a smoker continued to abstain from cigarette smoking for 5 years or more (10). Therefore, it seems to be a delayed reduction of SQC among occupational populations highly exposed to radon and arsenic, compared to the general population.

The strength of this study was that it was a large, prospective population-based cohort that included detailed personal, occupational, and smoking information. There were still some limitations. The closed underground mines and the high smoking rate might have resulted in secondhand smoke exposure to non-smokers, and further biased risk estimates downward (toward the null). Moreover, the histology information was lacking for nearly half of lung cancer cases, which would decrease the statistical power when the analysis was conducted according to histology. Finally, 4.0% (53/1,324) of lung cancer cases was measured by the face-to-face interviews with relatives and workmates of the miners, which may be inaccurate and lead to recall bias. In this study, we had added the exposures (radon and arsenic) together without any weights, which might bias the results due to the different risks of lung cancer by radon and arsenic. Therefore, studies directly comparing the lung cancer risks from radon or arsenic should be conducted in the future.

In conclusion, our study firstly reported that among workers exposed to radon and arsenic, the benefit of smoking cessation is more related to squamous cell lung carcinoma. A tailored smoking cessation strategy is needed among the occupational population exposed to radon and arsenic.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review boards of the National Cancer Center/National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Y-GF and F-HZ had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conception and design: Y-LQ. Data collection: ZS and Y-GF. Analysis and interpretation: ZS. Drafting the article: ZS and X-HJ. Manuscript revision: ZS, Y-GF, and Y-LQ. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Cancer Foundation of China (grant No. CFC2020KYXM001) and Key R & D projects of Science and Technology Department of Sichuan (grant No. 2020YFS0212).

Tianjin Key Medical Discipline (Specialty) Construction Project (grant No. TJLCMS2021-02).

ACKNOWLEDGMENTS

We gratefully acknowledge all participants who have participated in this study. We thank the staff of Office of Gejiu Municipal Leading Group for Cancer Prevention and Control, Gejiu City, Yunnan, China for their assistance in the collection of the follow-up data of the YTC cohort.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.817045/full#supplementary-material

REFERENCES

1. Rongshou Z, Kexin S, Siwei Z, Hongmei Z, Xiaonong Z, Ru C, et al. Analysis on the Prevalence of Malignant Tumors in China in 2015. *Chin J Oncol* (2019) 41:19–28. doi: 10.3760/cma.j.issn.0253-3766.2019.01.008
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
3. Smith CJ, Perfetti TA, Rumple MA, Rodgman A, Doolittle DJ. "IARC Group 2B Carcinogens" Reported in Cigarette Mainstream Smoke. *Food Chem Toxicol* (2001) 39:183–205. doi: 10.1016/S0278-6915(00)00164-2
4. Fingerhut M, Nelson DI, Driscoll T, Concha-Barrientos M, Takala J. The Contribution of Occupational Risks to the Global Burden of Disease: Summary and Next Steps. *La Med Del Lavoro* (2006) 97:313.
5. Ramo DE, Liu H, Prochaska JJ. Tobacco and Marijuana Use Among Adolescents and Young Adults: A Systematic Review of Their Co-Use. *Clin Psychol Rev* (2012) 32:105–21. doi: 10.1016/j.cpr.2011.12.002
6. Li J, Li WX, Bai C, Song Y. Particulate Matter-Induced Epigenetic Changes and Lung Cancer. *Clin Respir J* (2017) 11:539–46. doi: 10.1111/crj.12389
7. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some Drinking-Water Disinfectants and Contaminants, Including Arsenic. *IARC Monogr Eval Carcinog Risks Hum* (2004) 84:1–477.
8. International Agency for Research on Cancer. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Man-Made Mineral Fibres and Radon*. IARC (1988) 43:1–300. Available at: https://monographs.iarc.fr/wp-content/ uploads/2018/06/mono43.
9. Godtfredsen NS. Effect of Smoking Reduction on Lung Cancer Risk. *JAMA* (2005) 294:1505. doi: 10.1001/jama.294.12.1505
10. Tse LA, Yu IT, Qiu H, Au JS, Wang XR, Tam W, et al. Lung Cancer Decreased Sharply in First 5 Years After Smoking Cessation in Chinese Men. *J Thorac Oncol* (2011) 6:1670–6. doi: 10.1097/JTO.0b013e3182217bd4
11. Tse LA, Lin X, Li W, Qiu H, Chan CK, Wang F, et al. Smoking Cessation Sharply Reduced Lung Cancer Mortality in a Historical Cohort of 3185 Chinese Silicotic Workers From 1981 to 2014. *Br J Cancer* (2018) 119:1557–62. doi: 10.1038/s41416-018-0292-6
12. Markowitz SB, Levin SM, Miller A, Morabia A. Asbestos, Asbestosis, Smoking, and Lung Cancer. New Findings From the North American Insulator Cohort. *Am J Respir Crit Care Med* (2013) 188:90–6. doi: 10.1164/rccm.201302-0257OC
13. National Research Council. *Health Effects of Exposure to Radon: BEIR VI*. Washington, DC: National Academies Press (1999). Available at: https://doi.org/10.17226/5499.
14. Mary K, Schubauer-Berigan RDD, Pinkerton LE. Radon Exposure and Mortality Among White and American Indian Uranium Miners: An Update of the Colorado Plateau Cohort. *Am J Epidemiol* (2009) 169:718–30. doi: 10.1093/aje/kwn406
15. Kreuzer M, Fenske N, Schnelzer M, Walsh L. Lung Cancer Risk at Low Radon Exposure Rates in German Uranium Miners. *Brit J Cancer* (2015) 113 (9):1367–9. doi: 10.1038/bjc.2015.324
16. Rencher AC, Carter MW, Mckee DW. A Retrospective Epidemiological Study of Mortality at a Large Western Copper Smelter. *J Occup Environ Med* (1977) 19(11):754–8.
17. Pershagen G, Wall S, Taube A, Linnman L. On the Interaction Between Occupational Arsenic Exposure and Smoking and its Relationship to Lung Cancer. *Scand J Work Environ Health* (1981) 7(4):302–9. doi: 10.5271/sjweh.2544
18. Enterline PE, Marsh GM, Esmen NA, Henderson VL, Paik M. Some Effects of Cigarette Smoking, Arsenic, and SO2 on Mortality Among US Copper Smelter Workers. *J Occup Med* (1987) 29:831–8.
19. Järup L, Pershagen G. Arsenic Exposure, Smoking, and Lung Cancer in Smelter Workers—A Case-Control Study. *Am J Epidemiol* (1991) 134:545–51. doi: 10.1093/oxfordjournals.aje.a116128
20. Qiao YL, Taylor PR, Yao SX, Schatzkin A, Mao BL, Lubin J, et al. Relation of Radon Exposure and Tobacco Use to Lung Cancer Among Tin Miners in Yunnan Province, China. *Am J Ind Med* (1989) 16:511–21. doi: 10.1002/ajim.4700160504
21. Taylor PR, Qiao YL, Schatzkin A, Yao SX, Lubin J, Mao BL, et al. Relation of Arsenic Exposure to Lung Cancer Among Tin Miners in Yunnan Province, China. *Br J Ind Med* (1989) 46:881–6. doi: 10.1136/oem.46.12.881
22. Qiao YL, Taylor PR, Yao SX, Erozan YS, Luo XC, Barrett MJ, et al. Risk Factors and Early Detection of Lung Cancer in a Cohort of Chinese Tin Miners. *Ann Epidemiol* (1997) 7:533–41. doi: 10.1016/S1047-2797(97)00115-4
23. Taylor PR, Qiao YL, Schatzkin A, Yao SX, Lubin J, Mao BL, et al. Relation of Arsenic Exposure to Lung Cancer Among Tin Miners in Yunnan Province, China. *Occup Environ Med* (1989) 46:881–6. doi: 10.1136/oem.46.12.881
24. Pesch B, Kendzia B, Gustavsson P, Jöckel KH, Johnen G, Pohlabeln H, et al. Cigarette Smoking and Lung Cancer–Relative Risk Estimates for the Major Histological Types From a Pooled Analysis of Case-Control Studies. *Int J Cancer* (2012) 131:1210–9. doi: 10.1002/ijc.27339
25. Seki T, Nishino Y, Tanji F, Maemondo M, Takahashi S, Sato I, et al. Cigarette Smoking and Lung Cancer Risk According to Histologic Type in Japanese Men and Women. *Cancer Sci* (2013) 104:1515–22. doi: 10.1111/cas.12273
26. Chen ZM, Peto R, Iona A, Guo Y, Chen YP, Bian Z, et al. Emerging Tobacco-Related Cancer Risks in China: A Nationwide, Prospective Study of 0.5 Million Adults. *Cancer* (2015) 121 Suppl 17:3097–106. doi: 10.1002/cncr.29560
27. Pohlabeln H, Jöckel KH, Müller KM. The Relation Between Various Histological Types of Lung Cancer and the Number of Years Since Cessation of Smoking. *Lung Cancer* (1997) 18:223–9. doi: 10.1016/S0169-5002(97)00067-6
28. Leitzmann MF, Koebnick C, Abnet CC, Freedman ND, Park Y, Hollenbeck A, et al. Prospective Study of Physical Activity and Lung Cancer by Histologic Type in Current, Former, and Never Smokers. *Am J Epidemiol* (2009) 169:542–53. doi: 10.1093/aje/kwn371
29. Crispo A, Brennan P, Jöckel KH, Schaffrath-Rosario A, Wichmann HE, Nyberg F, et al. The Cumulative Risk of Lung Cancer Among Current, Ex- and Never-Smokers in European Men. *Br J Cancer* (2004) 91:1280–6. doi: 10.1038/sj.bjc.6602078
30. Mei Z, Wang L-M, Chong LY, Li X-Y, Yong J, Nan H, et al. A 2010 Survey on Smoking and Cessation Behaviors Among Chinese Adults. *Chin J Prev Med* (2012) 46(5):404–8. doi: 10.3760/cma.j.issn.0253-9624.2012.05.006
31. Putter H, Fiocco M, Geskus RB. Tutorial in Biostatistics: Competing Risks and Multi-State Models. *Stat Med* (2007) 26:2389–430. doi: 10.1002/sim.2712
32. Wakai K, Marugame T, Kuriyama S, Sobue T, Tamakoshi A, Satoh H, et al. Decrease in Risk of Lung Cancer Death in Japanese Men After Smoking Cessation by Age at Quitting: Pooled Analysis of Three Large-Scale Cohort Studies. *Cancer Sci* (2007) 98:584–9. doi: 10.1111/j.1349-7006.2007.00423.x
33. Pinsky PF, Zhu CS, Kramer BS. Lung Cancer Risk by Years Since Quitting in 30+ Pack Year Smokers. *J Med Screen* (2015) 22:151–7. doi: 10.1177/0969141315579119
34. Tanner NT, Kanodra NM, Gebregziabher M, Payne E, Halbert CH, Warren GW, et al. The Association Between Smoking Abstinence and Mortality in the National Lung Screening Trial. *Am J Respir Crit Care Med* (2016) 193:534–41. doi: 10.1164/rccm.201507-1420OC

35. Lubin JH, Boice JD Jr., Edling C, Hornung RW, Howe GR, Kunz E, et al. Lung Cancer in Radon-Exposed Miners and Estimation of Risk From Indoor Exposure. *J Natl Cancer Inst* (1995) 87:817–27. doi: 10.1093/jnci/87.11.817
36. Su Z, Wei M-N, Jia X-H, Fan Y-G, Zhao F-H, Zhou Q-H, et al. Arsenic, Tobacco Use, and Lung Cancer: An Occupational Cohort With 27 Follow-Up Years. *Environ Res* (2021) 206:112611. doi: 10.1016/j.envres.2021.112611
37. Khuder SA, Mutgi AB. Effect of Smoking Cessation on Major Histologic Types of Lung Cancer. *Chest* (2001) 120:1577–83. doi: 10.1378/chest.120.5.1577

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fonc.2022.858855

Cysteinyl Leukotriene Receptor Antagonists Associated With a Decreased Incidence of Cancer: A Retrospective Cohort Study

Ha Young Jang, In-Wha Kim and Jung Mi Oh**

College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, South Korea

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Nosheen Masood, Fatima Jinnah Women University, Pakistan
Pornpun Vivithanaporn, Mahidol University, Thailand
Kant Sangpairoj, Faculty of medicine, Thammasat University, Thailand

***Correspondence:**
In-Wha Kim
iwkim2@hanmail.net
Jung Mi Oh
jmoh@snu.ac.kr

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *20 January 2022*
Accepted: *16 March 2022*
Published: *07 April 2022*

Citation:
Jang HY, Kim I-W and Oh JM (2022) Cysteinyl Leukotriene Receptor Antagonists Associated With a Decreased Incidence of Cancer: A Retrospective Cohort Study. Front. Oncol. 12:858855. doi: 10.3389/fonc.2022.858855

Aim: Cysteinyl leukotrienes receptor antagonists (LTRAs) are promising chemoprevention options to target cysteinyl leukotriene signaling in cancer. However, only a number of randomized clinical trials (RCTs) or observational studies have been conducted to date; thus, the effect of LTRAs on patients is yet to be elucidated. Using insurance claim data, we aimed to evaluate whether LTRAs have cancer preventive effects by observing patients who took LTRAs.

Method: Patients diagnosed with asthma, allergic rhinitis, chronic cough, and have no history of cancer were followed-up from 2005 to 2017. Cox proportional hazard regression analysis was conducted to estimate the hazard ratios (HRs) for cancer risk of LTRA users.

Result: We followed-up (median: 5.6 years) 188,906 matched patients (94,453 LTRA users and 94,453 non-users). LTRA use was associated with a decreased risk of cancer (adjusted HR [aHR] = 0.85, 95% confidence interval [CI] = 0.83–0.87). The cancer risk showed a tendency to decrease rapidly when LTRAs were used in high dose (aHR = 0.56, 95% CI = 0.40–0.79) or for longer durations of more than 3 years (aHR = 0.68, 95% CI = 0.60–0.76) and 5 years (aHR = 0.33, 95% CI = 0.26–0.42). The greater preventive effects of LTRAs were also observed in patients with specific risk factors related to sex, age, smoking, and the presence of comorbidities.

Conclusion: In this study, we found that LTRA use was associated with a decreased risk of cancer. The high dose and long duration of the use of LTRAs correlated with a lower cancer risk. Since LTRAs are not yet used for the prevention or treatment of cancer, our findings could be used for developing a new chemo-regimen or designing feasible RCTs.

Keywords: cysteinyl leukotriene receptor antagonists, cancer, cancer prevention, drug repurposing, observational study

INTRODUCTION

Cancer is the leading cause of death in Korea, and the mortality rate of this disease continuously to increase annually (1). As the cancer incidence continues to rise, the importance of cancer prevention is being emphasized. Cancer treatment is expensive as well as developing effective anticancer drugs. Thus, if cancer is successfully prevented, the overall medical cost can be reduced. Moreover, it is also challenging to plan cancer prevention strategies through clinical trials in terms of its duration and cost. Cancer prevention clinical trials take more than 5-10 years to complete and usually require thousands of participants. The estimated cost for large clinical trials involving more than 10,000 people is approximately $100 to $200 million (2). Despite decade-long efforts to find effective cure, candidates for anticancer drugs are usually discontinued during the phase 3 of the clinical trials due to problems, such as efficacy and toxicity (3). As the results of these trials do not always lead to successful cancer prevention strategies, there is an urgent need for identifying alternative drug therapies effective in preventing cancer. Drug repurposing is the process of searching for new indications for drugs that already exist in the market (4). Since this method is based on previously accumulated research and development data, the new drug development process can be accelerated, cutting costs at the same time (5). Recently, many studies on drug repurposing are being conducted based on genome, phenome, and insurance claim data (6).

Inflammation is a critical part in the pathogenesis of cancer, and the correlation of high levels of cysteinyl leukotrienes (CysLT) and CysLT1 receptor (CYsLTR) with various types of cancer have been reported several times in *in-vitro* studies (7–12). CYsLTR antagonists (LTRAs), including montelukast, pranlukast, and zafirlukast, have been widely used for treating asthma, allergic rhinitis, or chronic cough (13), and are the most promising chemoprevention options to target CysLT signaling in cancer. In addition to CysLT1 signaling, montelukast essentially induces apoptosis in cancer cells while zafirlukast is found to be involved in the cancer cell cycle (14, 15). Moreover, the role of LTRAs could also be associated with cancer metastasis, showing cell migration and invasion were suppressed in glioblastoma cells (14), colon cancer cells (16), skin cancer cells (17), and 5-FU-resistant colon cancer cells (18). However, the chemopreventive effects of LTRAs described above are all reported in *in-vitro* studies. Thus, it is still questionable whether the same effects can be observed in people taking LTRAs, especially since only limited randomized clinical trials (RCTs) and observational studies for humans are available (19). To observe the cancer-preventing effects of LTRAs in humans, a long-term follow-up study with a sufficiently large cohort size is essential. Therefore, using insurance claim data, we aimed to evaluate whether LTRAs have cancer prevention effects in a real-world setting by observing patients who took LTRAs.

MATERIALS AND METHODS

Study Design and Sources

This study used a cohort study design and analyzed the health insurance data officially provided by the Korean National Health Insurance Service (KNHIS) (20). The insurance data included the patients' demographic, diagnosis, procedure, and prescription data. Additionally, physical examination data that were linked to the KNHIS data were used. Physical examination information included the body mass index (BMI), smoking status, alcohol consumption, and exercise data. The requirement for the written informed consent from the participants was waived and all participants were anonymized by a randomized identification number. This study was approved by the institutional review board (IRB) of Seoul National University (IRB No. E1901/003-004).

Study Population

To evaluate the effect of LTRA use on the prevention of cancer, patients diagnosed with asthma, allergic rhinitis, or chronic cough more than twice from 2005 to 2011 were included. Diagnosis of each disease was identified by the recorded diagnostic code of J45.x, J30.x, and R05.x for asthma, allergic rhinitis, and chronic cough in the claim, respectively. Patients who met the following criteria were excluded: diagnosed with asthma, allergic rhinitis, or chronic cough between 2002 and 2004; diagnosed with cancer before each patient's index date; received LTRAs before being diagnosed with asthma, allergic rhinitis, or chronic cough; whose follow-up period is less than 1 year; whose day of LTRA use is less than 30 days.

Ascertainment of Exposure

The LTRAs involved in this study include montelukast, pranlukast, and zafirlukast based on the anatomical therapeutic chemical (ATC) classification system. Information of the administered dose, frequency, and duration of the use of LTRAs were retrieved from the KNHIS database. Patients with no history of LTRA use were included in the non-user group. For the LTRA users, each daily dose was calculated by multiplying the number of tablets to be taken each day by the dose of each tablet, and this was converted to the defined daily dose (DDD), which is assigned by the World Health Organization's Collaborating Center (WHOCC) for Drug Statistics Methodology (www.whocc.no/atc_ddd_index) (21). The cumulated dose was defined as the sum of multiplying the prescribed duration by the defined daily dose (DDD) of LTRAs.

Ascertainment of Cancer

Individuals were followed-up until 2017, and outcomes were recorded from the individual's index date. Primary endpoint of the study was cancer. Cancer event was defined based on the International Classification of Diseases-10 (ICD-10) codes (C00-C97). Cancer with the top 5 mortality rates (lung, hepatic, colorectal, stomach, pancreatic) and additional cancer types (breast, urological, skin, and brain/central nervous system cancer) were defined as secondary endpoints (1).

Confounding Variables

Baseline characteristics, potentially influencing the study outcomes were included. These include demographic information, such as age at enrollment, sex, index year, region, and economic status. Region information was also collected by

dividing the patients into special metropolitan city, metropolitan city, and province based on the patients' insurance payment regions. Economic status of the enrolled participants was assessed based on income-related insurance payment. Concomitant asthma, anti-allergy medications, and initial diagnosis (asthma, allergic rhinitis, or chronic cough) within 1 year of index date were evaluated. Comorbidity burden was measured using the updated Charlson comorbidity index (CCI) to classify the level of comorbidity up to 1 year of index date (22). Furthermore, information on the smoking status and alcohol intake from questionnaire data and the BMI from physical examination data were collected.

Statistical Analysis

Statistical analyses were performed for the intention-to-treat population. In the LTRA user group, if the date of LTRA initiation differed from the time of diagnosis, the patients would have periods during which cancer could not have been affected by treatment (immortal time). Therefore, each patient's index date was defined as the very first date when LTRAs were prescribed for the LTRA users. The index date of non-users was then matched with the index date of the LTRA users. Patients were followed-up until the earliest onset of cancer, the date of the last follow-up, or the end of the study period. To adjust the effect of confounding variables between the LTRA user and non-user groups, propensity score matching was done. Propensity score was estimated by logistic regression with variables, including age, index year, region, economic status, co-medications, initial diagnosis, smoking status, alcohol intake, and BMI. LTRA users were matched 1:1 to non-users with the greedy 5 to 1 digit matching algorithm (23). Subsequently, the distribution of the propensity score before and after matching was inspected and the distribution of baseline covariates was evaluated with standardized difference. Standardized difference of over 0.1 was regarded as a sign of imbalance (24).

Cox proportional hazard regression was used to estimate the hazard ratio (HR) of LTRAs for cancer risk, with 95% confidence interval (CI). The confounding factors used were the age at enrollment, sex, index year, region, economic status, concomitant asthma/anti-allergy medications, initial diagnosis, CCI, smoking status, alcohol intake, and BMI. To test the robustness of our model, sensitivity analyses were performed. To prevent the LTRAs exposure factor from affecting the main outcomes, we applied a different exposure definition. In our original study design, the LTRAs exposure was defined as the sum of doses of the prescribed medications. In the sensitivity analysis, a new gap concept was defined to see the continuous use of LTRAs; if the gap between prescription refills was <30 days or at 50% of each prescription period, the patient was considered to have continued LTRA use. If the gap exceeded the predefined threshold, it was considered as patients have stopped and have not taken LTRAs any longer. Another sensitivity analysis was conducted by narrowing the index date between 2008 to 2011, and any changes in the risk of cancer were evaluated by calculating the HRs. Analyses were done with SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

Demographics

Among all the patients diagnosed with asthma, allergic rhinitis, or chronic cough two or more times between 2005 and 2011 (n = 4,387,602), a total of 2,632,224 newly diagnosed patients with these conditions without a cancer history were identified (**Figure 1**). After excluding the patients who do not meet the predefined inclusion criteria, the eligible study cohort included 1,786,168 patients (208,323 LTRA users and 1,577,845 non-users). LTRA users took more co-medications and had higher CCI scores. The proportion of patients who were diagnosed with asthma was higher in LTRA users (82.2%) than non-users (53.3%). After the propensity score matching, 94,453 LTRA users were matched with 94,453 non-users. The above difference (co-medications, CCI, and initial diagnosis) was reduced, and standardized differences were below 0.1 for all covariates (**Table 1**). The median length of follow-up was 5.6 years (5.5 and 5.7 years for non-users and LTRA users, respectively). The median duration of LTRAs prescription during follow-up (65 days, interquartile range: 41-150 days) and mean age of patients [56.4 years; men: 42.6% (n = 80,533)] were shown. The most frequently used DDD were intermediate doses (64.1%), followed by low doses (35.5%), and high doses (0.4%).

Risk of Cancer in LTRAs Users

The median time of the first onset of cancer events was 3.4 years. The incidence rates of all recorded cancer types were shown in **Supplementary Table S1**. The use of LTRAs showed a significantly decreased risk of overall cancers (adjusted HR [aHR] = 0.85, 95% CI = 0.83–0.87). When examining each type of cancer, hepatic cancer (aHR = 0.73, 95% CI = 0.68–0.79), colorectal cancer (aHR = 0.83, 95% CI = 0.76–0.91), gastric cancer (aHR = 0.69, 95% CI = 0.62–0.76), breast cancer (aHR = 0.77, 95% CI = 0.71–0.83), and urological cancer (aHR = 0.92, 95% CI = 0.86–0.97) were significantly associated with LTRA use. In contrast, LTRAs showed no significant effect on lung, pancreatic, skin, and brain/central nervous system cancers (**Table 2**).

Risk of Cancers by LTRAs Dose, Duration, and Cumulative Dose

When examining the cancer risk in terms of LTRA dose, the low (aHR = 0.89, 95% CI = 0.86–0.92) and intermediate doses (aHR = 0.86, 95% CI = 0.83–0.89) showed similar aHRs to the original results (**Table 3**). The aHR was also observed to be significantly lowered when the high dose was used (aHR = 0.56, 95% CI = 0.40–0.79). When the period of use of LTRAs was analyzed, the cancer risk showed a tendency to rapidly decrease when LTRAs were used for more than 3 years (aHR = 0.68, 95% CI = 0.60–0.76). Furthermore, the aHR decreased to 0.33 (95% CI = 0.26–0.42) when LTRA usage exceeds 5 years. A similar pattern was observed when the analysis was performed according to the cumulative dose obtained through the multiplication of dose and duration (cumulative DDD*year [cDY]). A significant

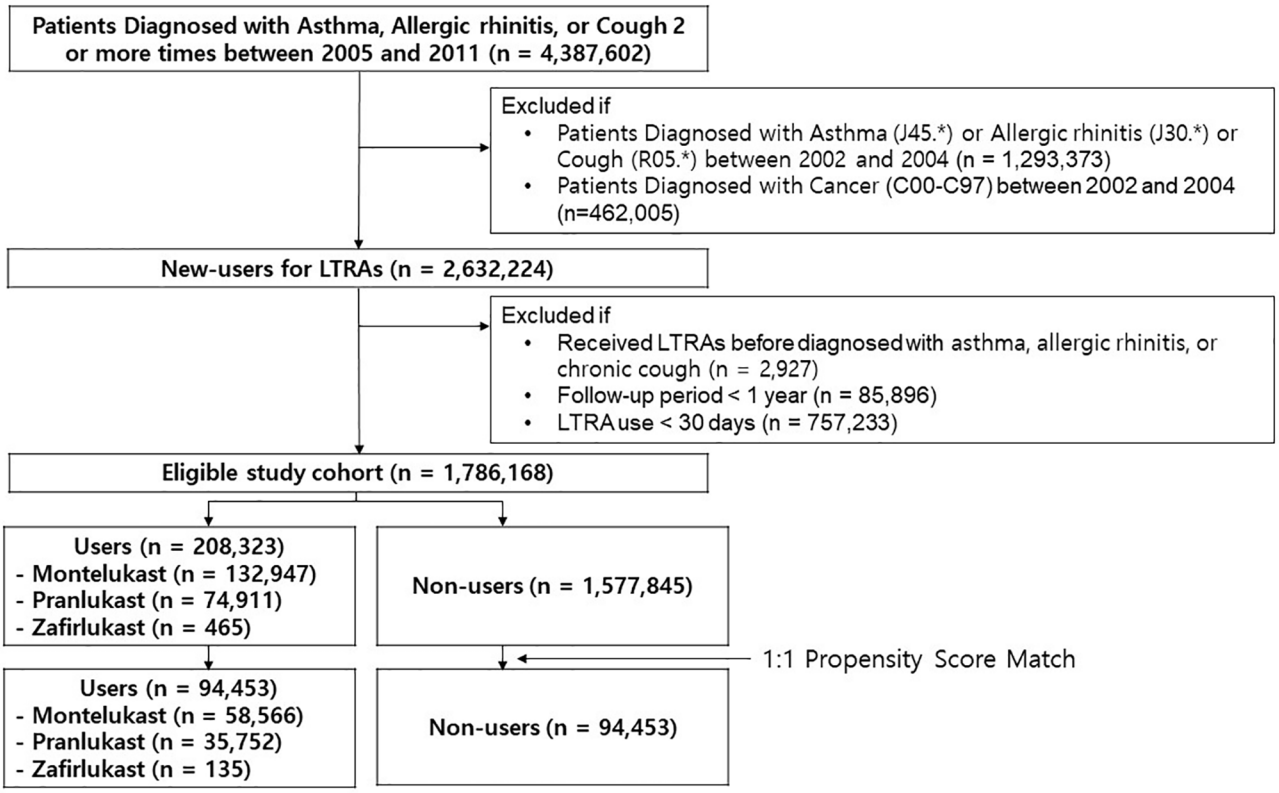


FIGURE 1 | Study flow chart. LTRAs, Cysteinyl leukotrienes receptor antagonist.

decrease in aHR was shown when the cumulative dose was more than 5 cDY (aHR = 0.53, 95% CI = 0.47–0.60).

Sensitivity Analyses

The cancer prevention effect of LTRAs was the same after the gap change to 30 days and at 50% proportion of permissible gap. LTRA usage still significantly lowered the risk of cancer, and similar results were observed across all cancer types (lung, hepatic, colorectal, gastric, pancreatic, breast, urological, skin, and brain/central nervous system cancer) (**Supplementary Table S2**). In the analysis for cancer risk by narrowing the index period between 2008 to 2011, the same results for all cancer types were observed (**Supplementary Table S3**).

Subgroup Analyses

The greater preventive effects of LTRAs were observed in men (aHR = 0.78, 95% CI = 0.75–0.81), patients aged >65 years (aHR = 0.75, 95% CI = 0.71–0.79), and with a history of smoking or still currently smoking (aHR = 0.81, 95% CI = 0.78–0.85) compared to women, patients aged ≤65 years, and those who never smoked, respectively (**Figure 2**). LTRA use in patients with CCI scores of 0 showed no significant association with cancer (aHR = 1.01, 95% CI = 0.89–1.14); however, with higher CCI scores, the HR gradually decrease from 1.05 (95% CI 0.99–1.11) (CCI score: 1) to 0.78 (95% CI 0.75–0.80) (CCI score: 3). No significant differences in aHR were observed according to the patients' alcohol intake, initial diagnosis, economic status, and region.

DISCUSSION

Our study analyzed patients who are using LTRAs through a long follow-up study. To our knowledge, this is the first research that consider the demographic information, co-medications, underlying comorbidities, and the patients' physical examination data, including smoking status, alcohol intake, and BMI, while using a sufficiently large sample size. Our study results found that LTRA use was associated with an overall decreased risk of cancer. In addition, by dividing the dose and period of LTRA use into several subgroups, our study could identify the amount of dose and duration that may significantly lower the risk of cancer.

A previous cohort study also showed that the use of LTRAs significantly decreased the overall cancer risk, specifically for lung, colorectal, and breast cancer (25). The same trends were also found in our study; however, the magnitude of the reduced

TABLE 1 | Baseline characteristics.

Characteristics	Non-users (N=94,453)	LTRAs users (N=94,453)	STD
Sex (male)	40,515 (42.8)	40,018 (42.4)	-0.005
Age (year)	51.4 ± 11.2	51.3 ± 12.4	0.014
BMI (kg/m^2)	23.9 ± 3.2	23.9 ± 3.5	0.001
Drink (times/week)	0.9 ± 1.5	0.9 ± 1.5	-0.005
Economic status[a]			
1	10929 (11.6)	11045 (11.7)	0.029
2	14083 (14.9)	14201 (15.0)	
3	22333 (23.6)	22042 (23.3)	
4	23318 (24.7)	23474 (24.9)	
5	23790 (25.2)	23691 (25.1)	
Comorbidities			
Asthma	72014 (76.2)	72262 (76.5)	0.041
Allergic rhinitis	16085 (17.0)	15729 (16.7)	
Chronic cough	6354 (6.7)	6462 (6.8)	
Index year			
2008	8871 (9.4)	8978 (9.5)	0.051
2009	12961 (13.7)	12857 (13.6)	
2010	12806 (13.6)	12760 (13.5)	
2011	14097 (14.9)	14306 (15.1)	
2012	15933 (16.9)	15910 (16.8)	
2013	12605 (13.4)	12559 (13.3)	
2014	9938 (10.5)	10077 (10.7)	
2015	7242 (7.7)	7006 (7.4)	
Charlson comorbidity index			
0	7539 (8.0)	7360 (7.8)	0
1	25567 (27.1)	25633 (27.1)	
2	9370 (9.9)	9362 (9.9)	
3	51977 (55.0)	52098 (55.2)	
Smoking			
Never	66822 (70.8)	67058 (71.0)	0
History of smoking	8708 (9.2)	8518 (9.0)	
Current smoking	18923 (20.0)	18877 (20.0)	
Co-medications			
Xanthines	48267 (51.1)	47859 (50.7)	-0.009
β-Blockers	56710 (60.0)	56803 (60.1)	0.002
Anti-cholinergics	7874 (8.3)	7378 (7.8)	-0.019
Systemic steroids	79022 (83.7)	79608 (84.3)	0.017
Region			
Special metropolitan city	49025 (51.9)	49054 (51.9)	0.025
Metropolitan city	21916 (23.2)	21982 (23.3)	
Province	23512 (24.9)	23417 (24.8)	

Values are represented as mean ± standard deviation or number (%); LTRAs, Cysteinyl leukotrienes receptor antagonist; STD, standardized difference.
[a]*Economic status was assessed based on income-related insurance payment; BMI, body mass index.*

risk was smaller than the previous reported study. This result can be attributed to differences in the sample size and in the use of various covariates. In the work of Tsai et al., the number of patients after the propensity score matching was 25,110 (4,185 in the taking group, 20,925 in the non-taking group), which was much smaller than the 188,906 participants in our study. In addition, their study did not consider the variables related to lifestyle (e.g., smoking status, alcohol intake), which are major risk factors of cancer.

We found that the use of LTRAs had a significant preventive effect on overall cancers, which was consistent with other previous findings. Many studies have reported that LTRAs are effective not only for treatment (14, 15, 26, 27), including cancer metastasis (14, 16, 17, 28), but also for prevention (11, 25, 29–33), so it seems that LTRAs can be used in various stages of cancer. First, LTRAs inhibit the growth and/or induce apoptosis of a large series of human cancer cell lines. LTRAs inhibit growth of glioblastomas cells, by decreasing expression of B-cell lymphoma 2 (Bcl-2) protein and reducing the phosphorylation of extracellular signal-regulated kinase 1/2 (14). In breast cancer cells, apoptosis was also induced (15). A similar mechanism was found in colon cancer. In addition to significant reductions in cell proliferation, adhesion and colony formation, the induction of cell cycle arrest and apoptosis were observed in a dose-dependent manner (26). Montelukast induced down-regulation of Bcl-2, up-regulation of Bcl-2 homologous antagonist/killer, and nuclear translocation of apoptosis-inducing factors in lung cancer cells (27). Second, LTRAs could inhibit metastasis of cancer by preventing tumor cell migration through both cerebral and peripheral capillaries (14). Matrix metallopeptidase-9 (MMP-9) degrades extracellular

TABLE 2 | Hazard ratios for each cancer components.

	Events	Person-year	Hazard ratio (95% CI)	
			Unadjusted	Adjusted
All Cancer				
Non-users	11369	520292	–	–
LTRAs	10399	536725	0.88 (0.86 – 0.91)	0.85 (0.83 – 0.87)
Lung Cancer				
Non-users	989	551209	–	–
LTRAs	1201	560047	1.19 (1.09 – 1.29)	1.06 (0.94 – 1.16)
Liver Cancer				
Non-users	1681	548448	–	–
LTRAs	1271	559331	0.74 (0.69 – 0.79)	0.73 (0.68 – 0.79)
Colorectal Cancer				
Non-users	1133	550157	–	–
LTRAs	1017	559934	0.88 (0.81 – 0.96)	0.83 (0.76 – 0.91)
Stomach Cancer				
Non-users	819	551045	–	–
LTRAs	625	560859	0.75 (0.67 – 0.83)	0.69 (0.62 – 0.76)
Pancreas Cancer				
Non-users	672	551969	–	–
LTRAs	641	561173	0.93 (0.84 – 1.04)	0.91 (0.81 – 1.01)
Breast Cancer				
Non-users	1433	549177	–	–
LTRAs	1181	559698	0.80 (0.74 – 0.87)	0.77 (0.71 – 0.83)
Urological Cancer				
Non-users	2146	547338	–	–
LTRAs	2191	557066	1.00 (0.94 – 1.06)	0.92 (0.86 – 0.97)
Skin Cancer				
Non-users	516	550132	–	–
LTRAs	534	558712	1.02 (0.90 – 1.15)	1.00 (0.88 – 1.14)
Brain and Central Nervous System Cancer				
Non-users	175	548347	–	–
LTRAs	150	554163	0.85 (0.68 – 1.05)	0.83 (0.67 – 1.03)

Hazard ratio was adjusted for age at enrollment, sex, index year, region, economic status, concomitant asthma/anti-allergy medications, initial diagnosis, charlson comorbidity index, smoking status, alcohol intake, and body mass index. CI, confidence interval; LTRAs, Cysteinyl leukotrienes receptor antagonist.

TABLE 3 | Hazard ratios for cancer according to dose, duration, and cumulative dose of cysteinyl leukotriene receptor antagonists.

	Events	Person-years	Adjusted Hazard ratio (95% CI)
Dose			
Non-users	11369	520292	–
<0.5 DDD	4439	228313	0.89 (0.86 – 0.92)
0.5–1.0 DDD	5926	305943	0.86 (0.83 – 0.89)
≥1.0 DDD	34	2469	0.56 (0.40 – 0.79)
Duration			
Non-users	11369	520292	–
<0.5 year	7689	419707	0.85 (0.82 – 0.87)
0.5–1 year	1086	50909	0.87 (0.82 – 0.93)
1–3 year	1269	45512	1.02 (0.97 – 1.09)
3–5 year	281	13656	0.68 (0.60 – 0.76)
≥5 year	74	6941	0.33 (0.26 – 0.42)
Cumulative dose			
Non-users	11369	520292	–
<0.5 cDY	6854	376505	0.84 (0.81 – 0.87)
0.5–1 cDY	1426	68835	0.89 (0.84 – 0.94)
1–3 cDY	1352	56193	0.95 (0.89 – 1.00)
3–5 cDY	466	16221	1.03 (0.94 – 1.13)
≥5 cDY	301	18971	0.53 (0.47 – 0.60)

*Hazard ratio was adjusted for age at enrollment, sex, index year, region, economic status, concomitant asthma/anti-allergy medications, initial diagnosis, Charlson comorbidity index, smoking status, alcohol intake, and body mass index. cDY, cumulative defined daily dose*year; CI, confidence interval; DDD, defined daily dose; LTRAs, Cysteinyl leukotrienes receptor antagonists.*

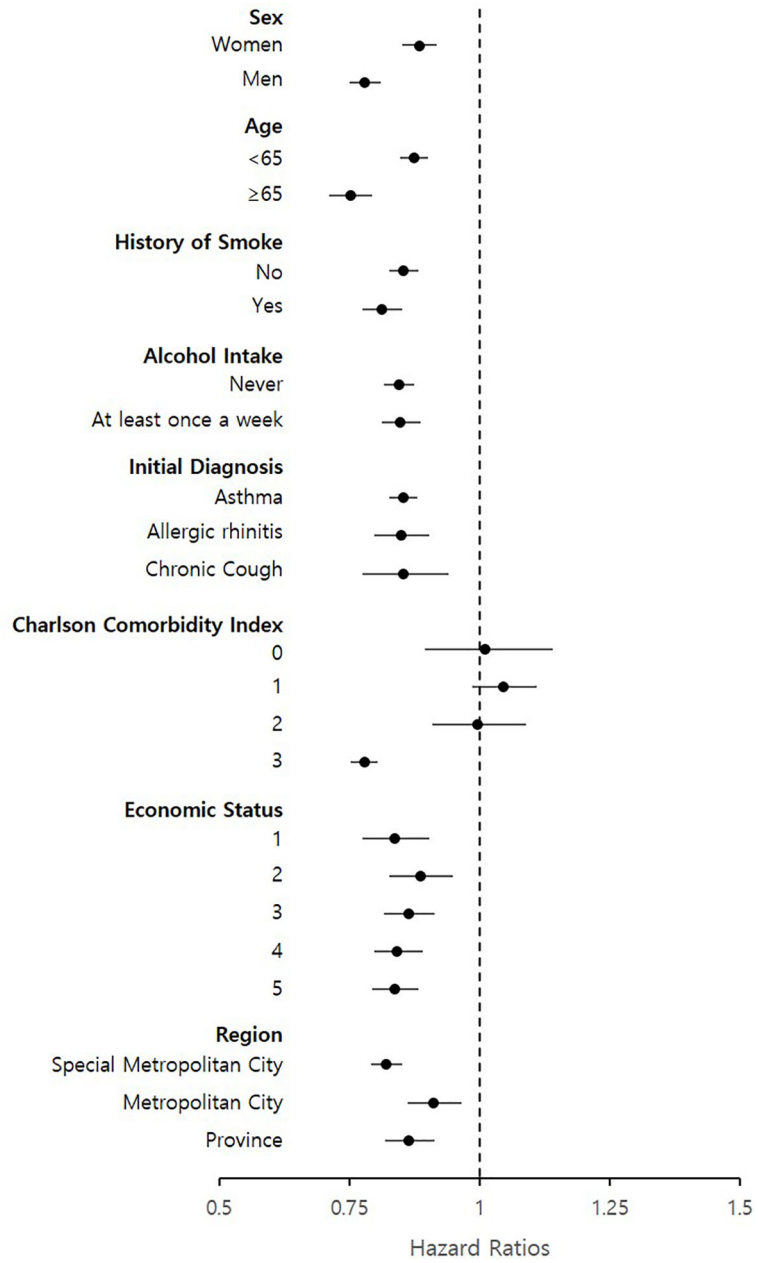


FIGURE 2 | Subgroup analysis of hazard ratios for cancer events based on patient's sex, age history of smoke, alcohol intake, initial diagnosis, charlson comorbidity index, economic status and region.

matrix proteins and was increased in colon cancer patients. The MMP-9 expression and activity were reduced by montelukast (16). LTRAs inhibited epidermal growth factor-induced T cell lymphoma invasion and metastasis inducing protein 1 expression in skin cancer cells (17). There seems to be a difference in roles of preventive mechanisms within the LTRAs. Pranlukast can inhibit tumor cell migration through both the brain and peripheral capillaries, whereas montelukast inhibits tumor cell migration only in the peripheral capillaries (28). The preventive effect of LTRAs has been reported in several *in-vitro* and *in-vivo* studies for certain cancers, including colorectal (29), gastric (11), and pancreatic cancer (30). A previous cohort study also showed similar results, reporting that the risks of breast, colorectal, and liver cancers were significantly reduced (25). However, a non-significant association between lung cancer and LTRA use was found in our study, while other groups have

reported its cancer risk reduction effect (25, 31). Three studies also showed that LTRAs reduced the risk of metastatic lung cancer, but not of lung cancer itself (28, 32, 33).

Despite these efforts, there have not been reports on any definite association between LTRAs and a specific type of cancer yet. The pathogenesis of cancer appears to be multifactorial, and such findings may have arisen due to differences in the study samples, study designs, or statistical methods. Tsai et al. (2016) also showed that the use of LTRAs was an independent protecting factor for overall cancers, reporting an HR of 0.31 (95% CI: 0.24–0.39). The magnitude of reduced risk was found to be smaller in our study (HR 0.85, 95% CI = 0.83–0.87), which might be due to larger sample size and the use of additional covariates. For instance, the patient's smoking status had a high HR range, 1.16 (against liver cancer) to 1.67 (against lung cancer), implying that the smoking covariate is a large proportion in our cox proportional hazard regression model.

In our study, the analysis of dose and duration of LTRAs use is noteworthy. Most LTRA prescriptions (99.6%) provided for the patients in this study were low (<0.5 DDD) or intermediate (0.5 ≤ DDD < 1.0), and only a few proportions were high (0.4%). Our results showed that overall cancer risk was rapidly lowered when LTRAs were used in high doses. In the duration analysis, >3 years of LTRA use correlated with a much lower HR for cancers. LTRAs are usually considered as safe during long-term administration even at doses substantially higher than the recommended dose (34). Therefore, this suggests that future studies should consider a higher dose and longer duration when prescribing LTRAs to be able to secure its anti-cancer property without having to worry about its side effects. However, recently, neuropsychiatric events were reported in post-marketing surveillance and resulted in safety alert in 2008 and a black box warning in 2020. Additionally, conflicting reports on the association between LTRAs and neuropsychiatric events have been published (35, 36). Therefore, it is necessary to pay attention to these precautions. The results of our study could also be used in the design of clinical trials. For instance, RCTs have been conducted with zileuton, a 5-lipoxygenase inhibitor that shares a similar mechanism with LTRAs, as an adjuvant agent to conventional chemotherapy for lung cancer patients (37). With this, new and improved RCTs can be conducted using LTRAs as an addition to existing anticancer therapies.

In our subgroup analysis, notable results were also observed in specific patient groups. The greater preventive effects of LTRAs in lowering the risk of cancer were observed in the following: in men, patients aged >65 years, patients with a history of smoking or are currently smoking, and those with high CCI scores. Considering that men, aged patients, smoking, and the presence of various comorbidities are well-known risk factors, LTRAs may contribute to lowering the cancer risks in patients with these particular characteristics. For the design of realistic and feasible clinical trials, the selection of specific patient groups with the above-mentioned risk factors may be beneficial and more effective.

There are several limitations encountered in our study. Due to the nature of the real-world data, the purpose of prescribing LTRAs to the patients was not for cancer prevention. Moreover, our study does not include an active comparator, and therefore it may be susceptible to selection bias. However, to reduce bias, as many variables were collected and matched to minimize the differences between groups. But note that there may still be some residual confounding after bias reduction. It was impossible to specify the stage/subtype of cancer because the disease information provided by the ICD-10 code was limited. We also suggest that some caution should be exercised when interpreting our results. There have been several studies showing that the use of LTRAs are also effective in reducing the risk of lung cancer, but the results in our study were not statistically significant (25, 31). Considering that baseline comorbidities, such as asthma, can have a significant effect on the occurrence of lung cancer (38, 39), this study may not have completely ruled out the effects of other comorbid diseases on cancer because it used CCI score as an indirect measure of various disease severity. Likewise, our study used a retrospective cohort design and not all information are included and available in the KNHIS data. Therefore, although we adjusted for all possible confounders, there still might be residual confounding factors present during our analyses.

The findings of our study suggest that the use of LTRAs was associated with a decreased risk of overall cancer. The high dose and long duration of LTRA use correlated with the lowered risk. The greater preventive effects of LTRAs were also observed in patients with specific risk factors related to sex, age, smoking, and the presence of comorbidities. As LTRAs have not yet been used for the prevention or treatment of cancer, our findings could be used for developing a new chemo-regimen or in designing feasible RCTs. For future studies, further research is needed to elucidate the specific mechanism and clinical significance of our results.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Data that can view all the records of a patient are difficult to share due to the policy of the NHIS. It can only be viewed in anonymized form when analyzed. Therefore, if there is a request for original data, the statistical data obtained after the desired statistical processing on the server will be shared. Requests to access these datasets should be directed to National Health Insurance Service, nhiss.nhis.or.kr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional review board (IRB) of Seoul National University (IRB No. E1901/003-004). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

HJ contributed to the conception and design of the study, data acquisition, analysis and interpretation of results, drafting, and

revision of the manuscript. JO and I-WK contributed to the conception and design of the study, analysis and interpretation of results, and revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Research Foundation of Korea grant funded by the Korean government (MSIT) (no. NRF-2018R1A2B6001859).

ACKNOWLEDGMENTS

This study used National Health Insurance Service (NHIS) data (NHIS-2019-1-324) from the NHIS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.858855/full#supplementary-material

REFERENCES

1. Korea VSDS, Shin H-Y, Kim J, Lee S, Park MS, Park S, et al. Cause-Of-Death Statistics in 2018 in the Republic of Korea. *J Korean Med Assoc* (2020) 63 (5):286–97. doi: 10.5124/jkma.2020.63.5.286
2. Guess HA, Rudnick SA. Use of Cost-Effectiveness Analysis in Planning Cancer Chemoprophylaxis Trials. *Control Clin Trials* (1983) 4(2):89–100. doi: 10.1016/s0197-2456(83)80016-6
3. Williams R. Discontinued in 2013: Oncology Drugs. *Expert Opin Investig Drugs* (2015) 24(1):95–110. doi: 10.1517/13543784.2015.971154
4. Langedijk J, Mantel-Teeuwisse AK, Slijkerman DS, Schutjens MH. Drug Repositioning and Repurposing: Terminology and Definitions in Literature. *Drug Discov Today* (2015) 20(8):1027–34. doi: 10.1016/j.drudis.2015.05.001
5. Ashburn TT, Thor KB. Drug Repositioning: Identifying and Developing New Uses for Existing Drugs. *Nat Rev Drug Discov* (2004) 3(8):673–83. doi: 10.1038/nrd1468
6. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A Survey of Current Trends in Computational Drug Repositioning. *Brief Bioinform* (2016) 17 (1):2–12. doi: 10.1093/bib/bbv020
7. Zhang WP, Hu H, Zhang L, Ding W, Yao HT, Chen KD, et al. Expression of Cysteinyl Leukotriene Receptor 1 in Human Traumatic Brain Injury and Brain Tumors. *Neurosci Lett* (2004) 363(3):247–51. doi: 10.1016/j.neulet.2004.03.088
8. Magnusson C, Mezhybovska M, Lorinc E, Fernebro E, Nilbert M, Sjolander A. Low Expression of CysLT1R and High Expression of CysLT2R Mediate Good Prognosis in Colorectal Cancer. *Eur J Cancer* (2010) 46(4):826–35. doi: 10.1016/j.ejca.2009.12.022
9. Matsuyama M, Yoshimura R. Cysteinyl-Leukotriene1 Receptor is a Potent Target for the Prevention and Treatment of Human Urological Cancer. *Mol Med Rep* (2010) 3(2):245–51. doi: 10.3892/mmr_00000247
10. Magnusson C, Liu J, Ehrnstrom R, Manjer J, Jirstrom K, Andersson T, et al. Cysteinyl Leukotriene Receptor Expression Pattern Affects Migration of Breast Cancer Cells and Survival of Breast Cancer Patients. *Int J Cancer* (2011) 129(1):9–22. doi: 10.1002/ijc.25648
11. Venerito M, Kuester D, Harms C, Schubert D, Wex T, Malfertheiner P. Upregulation of Leukotriene Receptors in Gastric Cancer. *Cancers (Basel)* (2011) 3(3):3156–68. doi: 10.3390/cancers3033156
12. Slater K, Heeran AB, Garcia-Mulero S, Kalirai H, Sanz-Pamplona R, Rahman A, et al. High Cysteinyl Leukotriene Receptor 1 Expression Correlates With Poor Survival of Uveal Melanoma Patients and Cognate Antagonist Drugs Modulate the Growth, Cancer Secretome, and Metabolism of Uveal Melanoma Cells. *Cancers (Basel)* (2020) 12(10):2950. doi: 10.3390/cancers12102950
13. Yokomizo T, Nakamura M, Shimizu T. Leukotriene Receptors as Potential Therapeutic Targets. *J Clin Invest* (2018) 128(7):2691–701. doi: 10.1172/JCI97946
14. Piromkraipak P, Parakaw T, Phuagkhaopong S, Srihirun S, Chongthammakun S, Chaithirayanon K, et al. Cysteinyl Leukotriene Receptor Antagonists Induce Apoptosis and Inhibit Proliferation of Human Glioblastoma Cells by Downregulating B-Cell Lymphoma 2 and Inducing Cell Cycle Arrest. *Can J Physiol Pharmacol* (2018) 96(8):798–806. doi: 10.1139/cjpp-2017-0757
15. Suknuntha K, Yubolphan R, Krueaprasertkul K, Srihirun S, Sibmooh N, Vivithanaporn P. Leukotriene Receptor Antagonists Inhibit Mitogenic Activity in Triple Negative Breast Cancer Cells. *Asian Pac J Cancer Prev* (2018) 19(3):833–7. doi: 10.22034/APJCP.2018.19.3.833
16. Vinnakota K, Zhang Y, Selvanesan BC, Topi G, Salim T, Sand-Dejmek J, et al. M2-Like Macrophages Induce Colon Cancer Cell Invasion *via* Matrix Metalloproteinases. *J Cell Physiol* (2017) 232(12):3468–80. doi: 10.1002/jcp.25808
17. Magi S, Takemoto Y, Kobayashi H, Kasamatsu M, Akita T, Tanaka A, et al. 5-Lipoxygenase and Cysteinyl Leukotriene Receptor 1 Regulate Epidermal Growth Factor-Induced Cell Migration Through Tiam1 Upregulation and Rac1 Activation. *Cancer Sci* (2014) 105(3):290–6. doi: 10.1111/cas.12340
18. Satapathy SR, Sjolander A. Cysteinyl Leukotriene Receptor 1 Promotes 5-Fluorouracil Resistance and Resistance-Derived Stemness in Colon Cancer Cells. *Cancer Lett* (2020) 488:50–62. doi: 10.1016/j.canlet.2020.05.023
19. Saier L, Peyruchaud O. Emerging Role of Cysteinyl LTs in Cancer. *Br J Pharmacol* (2021) 1–20. doi: 10.1111/bph.15402
20. Cheol Seong S, Kim YY, Khang YH, Heon Park J, Kang HJ, Lee H, et al. Data Resource Profile: The National Health Information Database of the National Health Insurance Service in South Korea. *Int J Epidemiol* (2017) 46(3):799–800. doi: 10.1093/ije/dyw253
21. WHOCC. *ATC Classifcation Index With DDDs*. Oslo, Norway: WHO Collaborating Centre for Drug Statistics Methodology (2016). Available at: www.whocc.no/atc_ddd_index.
22. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *Am J Epidemiol* (2011) 173(6):676–82. doi: 10.1093/aje/kwq433
23. Parsons LS. *Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques* (2001). Available at: http://www2.sas.com/proceedings/sugi26/p214-26.pdf (Accessed 2017-01-01).
24. Zhang Z, Kim HJ, Lonjon G, Zhu Ywritten on behalf of, A.M.E. Big-Data Clinical Trial Collaborative Group. Balance Diagnostics After Propensity Score Matching. *Ann Transl Med* (2019) 7(1):16. doi: 10.21037/atm.2018.12.10
25. Tsai MJ, Wu PH, Sheu CC, Hsu YL, Chang WA, Hung JY, et al. Cysteinyl 26. Leukotriene Receptor Antagonists Decrease Cancer Risk in Asthma Patients. *Sci Rep* (2016) 6:23979. doi: 10.1038/srep23979
26. Savari S, Liu M, Zhang Y, Sime W, Sjolander A. CysLT(1)R Antagonists Inhibit Tumor Growth in a Xenograft Model of Colon Cancer. *PloS One* (2013) 8(9):e73466. doi: 10.1371/journal.pone.0073466
27. Tsai MJ, Chang WA, Tsai PH, Wu CY, Ho YW, Yen MC, et al. Montelukast Induces Apoptosis-Inducing Factor-Mediated Cell Death of Lung Cancer Cells. *Int J Mol Sci* (2017) 18(7):1353. doi: 10.3390/ijms18071353
28. Nozaki M, Yoshikawa M, Ishitani K, Kobayashi H, Houkin K, Imai K, et al. Cysteinyl Leukotriene Receptor Antagonists Inhibit Tumor Metastasis by Inhibiting Capillary Permeability. *Keio J Med* (2010) 59(1):10–8. doi: 10.2302/kjm.59.10
29. Burke L, Butler CT, Murphy A, Moran B, Gallagher WM, O'Sullivan J, et al. Evaluation of Cysteinyl Leukotriene Signaling as a Therapeutic Target for Colorectal Cancer. *Front Cell Dev Biol* (2016) 4:103. doi: 10.3389/fcell.2016.00103
30. Hennig R, Ding XZ, Tong WG, Schneider MB, Standop J, Friess H, et al. 5-Lipoxygenase and Leukotriene B(4) Receptor Are Expressed in Human

Pancreatic Cancers But Not in Pancreatic Ducts in Normal Tissue. *Am J Pathol* (2002) 161(2):421–8. doi: 10.1016/S0002-9440(10)64198-3
31. Kahnt AS, Rorsch F, Diehl O, Hofmann B, Lehmann C, Steinbrink SD, et al. Cysteinyl Leukotriene-Receptor-1 Antagonists Interfere With PGE2 Synthesis by Inhibiting mPGES-1 Activity. *Biochem Pharmacol* (2013) 86(2):286–96. doi: 10.1016/j.bcp.2013.05.005
32. Lee KS, Kim SR, Park HS, Jin GY, Lee YC. Cysteinyl Leukotriene Receptor Antagonist Regulates Vascular Permeability by Reducing Vascular Endothelial Growth Factor Expression. *J Allergy Clin Immunol* (2004) 114 (5):1093–9. doi: 10.1016/j.jaci.2004.07.039.32
33. Gunning WT, Kramer PM, Steele VE, Pereira MA. Chemoprevention by Lipoxygenase and Leukotriene Pathway Inhibitors of Vinyl Carbamate-Induced Lung Tumors in Mice. *Cancer Res* (2002) 62(15):4199–201.
34. Gulati A. *Review of the Role of Anti Leukotrienes in the Therapy of Allergic Rhinitis in Children* (2013). Available at: https://www.who.int/selection_medicines/committees/expert/19/applications/LTRA_28_C_Ad.pdf (Accessed 2021.05.24).
35. Umetsu R, Tanaka M, Nakayama Y, Kato Y, Ueda N, Nishibata Y, et al. Neuropsychiatric Adverse Events of Montelukast: An Analysis of Real-World Datasets and Drug-Gene Interaction Network. *Front Pharmacol* (2021) 12:764279. doi: 10.3389/fphar.2021.764279
36. Fox CW, Khaw CL, Gerke AK, Lund BC. Montelukast and Neuropsychiatric Events - A Sequence Symmetry Analysis. *J Asthma* (2021) 1–7. doi: 10.1080/02770903.2021.2018705
37. Edelman M. *Carboplatin and Gemcitabine Combined With Celecoxib and/or Zileuton in Treating Patients With Advanced Non-Small Cell Lung Cancer* (2003). Available at: https://clinicaltrials.gov/ct2/show/NCT00070486 (Accessed 2021-05-24).
38. Koh WP, Yuan JM, Wang R, Seow A, Lee HP, Yu MC. Chronic Rhinosinusitis and Risk of Lung Cancer in the Singapore Chinese Health Study. *Int J Cancer* (2008) 123(6):1398–402. doi: 10.1002/ijc.23623
39. Qu YL, Liu J, Zhang LX, Wu CM, Chu AJ, Wen BL, et al. Asthma and the Risk of Lung Cancer: A Meta-Analysis. *Oncotarget* (2017) 8(7):11614–20. doi: 10.18632/oncotarget.14595

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fonc.2022.852718

Analysis of Tryptophan and Its Main Metabolite Kynurenine and the Risk of Multiple Cancers Based on the Bidirectional Mendelian Randomization Analysis

*Ran Li†, Xuanyang Wang†, Yuntao Zhang, Xiaoqing Xu, Lulu Wang, Chunbo Wei, Lin Liu, Ziqi Wang and Ying Li**

Department of Nutrition and Food Hygiene, the National Key Discipline, School of Public Health, Harbin Medical University, Harbin, China

Background: Tryptophan and its metabolites have been found related to various cancers, but the direction of this relationship is still unclear. The purpose of this study is to explore the causal associations of tryptophan and kynurenine with multiple cancers based on the bidirectional Mendelian randomization analysis.

Methods: The data of a genome-wide association study meta-analysis on 7,824 individuals was used to explore the genetic variants strongly associated with tryptophan and kynurenine. Genetic instruments of four specific cancers were obtained from available summary-level data of 323,590 European participants. Bidirectional Mendelian randomization analysis was conducted to examine possible causality. Sensitivity analysis was performed to test heterogeneity and horizontal pleiotropy. COX regression analysis was conducted to explore associations between dietary tryptophan and cancer mortality in NHANES 1988-1994.

Results: No evidence of any causal association of tryptophan and kynurenine with the risk of four specific cancers was shown, except for weak correlations were suggested between lung or prostate cancer and kynurenine. Multiple sensitivity analyses generated similar results. Our findings from COX regression analysis were consistent with the above results.

Conclusions: Our study did not find any causal relationship between tryptophan and kynurenine and multiple cancers. The associations still need further research.

Keywords: tryptophan, kynurenine, multiple cancers, causation, Mendelian randomization analysis

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Andreas Hadjisavvas, The Cyprus Institute of Neurology and Genetics, Cyprus
Ilona Sadok, The John Paul II Catholic University of Lublin, Poland
Dietmar Fuchs, Innsbruck Medical University, Austria

***Correspondence:**
Ying Li
liying_helen@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *11 January 2022*
Accepted: *15 March 2022*
Published: *14 April 2022*

Citation:
Li R, Wang X, Zhang Y, Xu X, Wang L, Wei C, Liu L, Wang Z and Li Y (2022) Analysis of Tryptophan and its Main Metabolite Kynurenine and the Risk of Multiple Cancers Based on the Bidirectional Mendelian Randomization Analysis. Front. Oncol. 12:852718. doi: 10.3389/fonc.2022.852718

1 INTRODUCTION

As an essential amino acid for the human body, tryptophan and a series of intermediate products of its metabolic pathways (serotonin pathway, kynurenine pathway and indole pathway) have become the therapeutic targets for depression, schizophrenia, neurodegenerative diseases, autoimmune diseases and cancers (1), such as some of the main rate-limiting enzymes (2) tryptophan-2,3-dioxygenase (TDO), kynurenine monooxygenase (KMO), indoleamine-2,3-dioxygenase 1 (IDO1) and indoleamine-2,3-dioxygenase 2 (IDO2).

Most existing studies have shown that tryptophan metabolic pathways including tryptophan degradation, kynurenine synthesis and overactivation of some major rate-limiting enzymes can promote tumor progression by inhibiting anti-tumor immune responses, limiting tumor immune infiltration and enhancing the malignant characteristics of cancer cells (3). Case-control studies have suggested that tryptophan metabolism pathways play a role in regulating regulatory T cells and in the infiltration of immune cells in cancer (4). The inhibitory effect of tryptophan metabolism pathways on immune cells is believed to be achieved by increasing the immunosuppressive catabolites of tryptophan and reducing tryptophan (5). *In vitro* experiments have found that higher levels of kynurenine, the main metabolite of tryptophan, were suggested to increase the proliferation and migration ability of cancer cells and help tumors avoid immune surveillance by reducing the activity of natural killer cells, dendritic cells or proliferating T cells (6). In addition, animal experiments have also indicated that enzymes involved in tryptophan metabolism are expressed in a variety of cancers. IDO1 is expressed in about 58% of human tumors and is related to the adverse clinical outcomes of various cancers, including melanoma, gynecological cancer and hematological malignancies (7).

However, the above associations have not been replicated in other studies. Some studies suggested that the immunomodulatory properties of the tryptophan metabolism pathways were mainly results of the influence of metabolites of the kynurenine pathway, rather than results of the reduction of tryptophan (8). Although the levels of systemic tryptophan in patients with lung cancer (9), malignant glioma (10), malignant melanoma (11), rectal cancer (12) and gynecological-related cancers (13) showed a downward trend, the elevation of the metabolites of the kynurenine pathway in the blood was rarely observed, which may be due to the small local changes of kynurenine and its downstream metabolites in tumor microenvironment.

Moreover, the link between tryptophan metabolism pathways and cancer has prompted a lot of researches on treatments targeting the kynurenine pathway, especially by inhibiting the key enzymes including TDO, IDO1 and KMO (1). Although current clinical trials have found some of these key enzyme inhibitors achieved the expected effects in early cancer immunotherapy, the results of phase III trials were negative. Given the mixed results, the unclear causal relationship between tryptophan and its main metabolite kynurenine and cancer still needs to be clarified.

In general, various cancers have been reported to be related to the changes of tryptophan and its metabolites in human body, such as lung cancer (9), breast cancer (14–16), colon cancer (17), rectal cancer (12), ovarian cancer (18), prostate cancer (19, 20), malignant glioma (10), malignant melanoma (11) and T cells leukemia (21), etc. However, no uniform convincing conclusion has been drawn so far.

Mendelian randomization (MR) is a causal inference method that uses genetic variations as exposure tools to estimate the causal influences of exposures on outcomes, based on the Mendelian Law of Independence. It overcomes the inherent confounding factors of general researches and provides reasonable temporality of causal inference (22).

In order to explore the causal associations between changes in the tryptophan metabolic pathways, including tryptophan and its main metabolite kynurenine, and the risk of site-specific cancers such as breast, lung, prostate and ovarian cancers, a bidirectional MR analysis was conducted in this study. To evaluate the presumed causal relationship, tryptophan and kynurenine existed in either plasma or serum were considered to be exposure factors and the risk of site-specific cancers was considered to be outcome in forward MR analysis. During reverse MR analysis, the risk of breast, lung, prostate and ovarian cancer was selected as exposure factor and the plasma or serum tryptophan and kynurenine concentrations were chosen as outcomes. Throughout the bidirectional MR analysis, SNPs strongly associated with the selected exposure factors ($P < 5\times10^{-8}$) were used as genetic instruments. Meanwhile, we used NHANES 1988-1994 (NHANES III) data to analyze the association of dietary tryptophan intake and cancer mortality.

2 MATERIALS AND METHODS

2.1 Bidirectional Mendelian Randomization Analysis

2.1.1 Study Design Overview

Figure 1 provides an overview of the participating studies and overall design of the MR analysis performed. We have identified SNPs that have a strong correlation with target exposure in published public data, and then explored whether there is a potential causal relationship between them and the corresponding outcomes. Briefly, we conducted bidirectional MR analysis twice, one to evaluate the potential causal association between tryptophan and cancer, and the other to evaluate the potential causal relationship of kynurenine and cancer.

2.1.2 Selection of Genetic Instruments Strongly Associated With Tryptophan or Kynurenine

Summary statistics of a meta-analysis of Genome-wide Association Studies (GWAS) were obtained. From this study, SNPs strongly associated with tryptophan or kynurenine, at a statistically significance level ($P < 5\times10^{-8}$) were identified, by genotyping 7,824 adult individuals from 2 European population studies (23) (KORA-TwinsUK studies). Then pairwise-linkage disequilibrium (LD) clumping with a clumping window of 10 MB and an r^2 cutoff of 0.001 was applied to ensure independence among genetic instruments. To evaluate the weak instrument bias, F-statistic for each SNP was calculated. SNPs with low statistical power were removed (24) (F-statistics < 10).

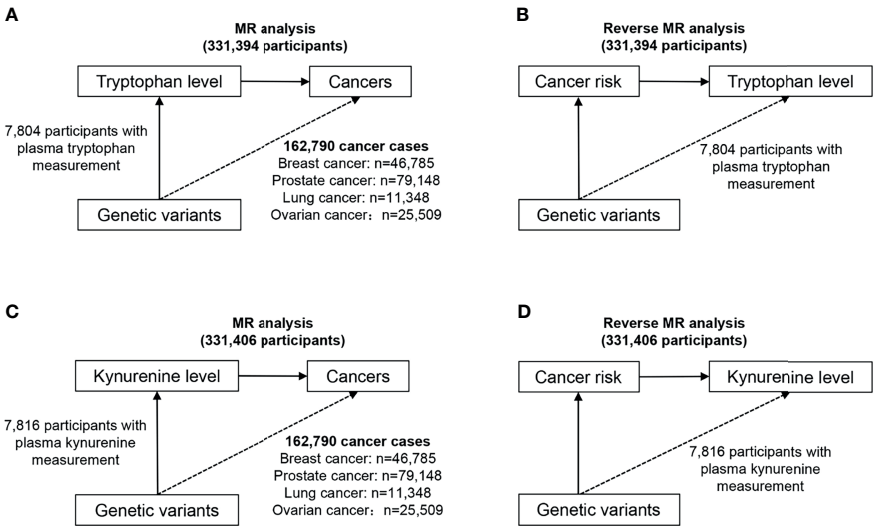


FIGURE 1 | Study design overview of the bidirectional MR analysis. **(A)** Design of the MR analysis of the causal association of circulating tryptophan levels with the risk of site-specific cancers (upper left). **(B)** Design of the reverse MR analysis of the causal association of site-specific cancers with circulating tryptophan levels (upper right). **(C)** Design of the MR analysis of the causal association of circulating kynurenine levels with the risk of site-specific cancers (lower left). **(D)** Design of the reverse MR analysis of the causal association of site-specific cancers with circulating kynurenine levels (lower right).

To make sure the effects of SNPs on the exposure correspond to the same allele as their effects on outcome, the matching of effect alleles of each SNP between the summary statistics of the exposure and the outcome was examined using the harmonise_data function (25). Finally, we selected 18 tryptophan-related SNPs and 4 kynurenine-related SNPs as genetic instruments for the MR analysis. The details of each SNP are described in **Supplementary Table S1**.

2.1.3 Selection of Genetic Instruments Strongly Associated With Breast Cancer, Lung Cancer, Prostate Cancer or Ovarian Cancer

Instrumental variables associated with breast cancer, lung cancer, prostate cancer or ovarian cancer were selected from the summary statistics of a meta-analysis of GWAS with numbers of cases ranging from 11,348 (lung cancer) to 79,148 (prostate cancer). Publicly available summary-level data for these cancers were obtained from the Breast Cancer Association Consortium (BCAC), International Lung Cancer Consortium (ILCCO), the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) and the Ovarian Cancer National Alliance (OCAC) respectively. Briefly, the BCAC and PRACTICAL consortiums aim to identify genes that are related to the risk of breast and prostate cancer by combining data from many studies. The ILCCO was established in 2004, with the goal of sharing compatible data from lung cancer epidemiology studies around the world to maximize statistical power. The OCAC consortium was founded in 2005 dedicated to foster collaborative efforts to discover and validate associations between genetic polymorphisms and the risk of ovarian cancer (18). After pairwise-linkage disequilibrium (LD) clumping and matching of coding alleles between exposure and outcome, we obtained 30 SNPs for breast cancer, 3 SNPs for lung cancer, 89 SNPs for prostate cancer and 7 SNPs for ovarian cancer. The details of each SNP are described in **Supplementary Table S2**.

2.1.4 Outcome Data Sources

Publicly available summary-level data for breast, lung, prostate and ovarian cancer were obtained from the BCAC, ILCCO, PRACTICAL and OCAC using the MR-Base database. Summary-level data for tryptophan and kynurenine were obtained from the KORA-TwinsUK studies up to 7824 adult individuals of European descent (26, 27). Detailed information on the above sources are described in **Supplementary Tables S3–S5**.

2.1.5 Statistical Analysis

Based on the publicly available GWAS summary statistics we retrieved, a bidirectional MR analysis was conducted. Firstly, we performed a MR analysis using the inverse-variance-weighted (IVW) method (28) as our primary MR method to assess the association between genetically predicted circulating tryptophan levels and cancers. To reduce the possibility that the genetic instruments of exposure affect the outcome independently, the following sensitivity analysis methods were chosen: the maximum likelihood method, the simple median method, the weighted median method (29) and the penalised weighted median. Secondly, we carried out a reverse MR analysis to examine the potential causal association of site-specific cancers with circulating tryptophan levels. The IVW method was also treated as the primary approach. We conducted sensitivity analysis using the MR-Egger method, the weighted median method, the simple mode method and the weighted mode method to evaluate the

possible violation of the MR assumptions and the robustness of the results. MR-Egger regression (30), MR-PRESSO global test (31) and Cochran,s Q test were executed to detect the degree of heterogeneity and horizontal pleiotropy among estimates of SNPs in each analysis. The above statistical methods for the bidirectional MR analysis were also used to explore the relationship between circulating kynurenine levels and site-specific cancers.

Statistical analysis was conducted in R 4.1.1 and MR Base platform (32) (http://app.mrbase.org/). All *P*-values were two-tailed and associations were considered statistically significant at $P < 0.05$.

2.2 Cox Proportional Hazards Model

2.2.1 Study Population

A total of 16,678 adults met the inclusion and exclusion criteria in NHANES III were selected in our study. The NHANES III survey was conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) between 1988 and 1994, which was designed to examine the health and nutritional status of the noninstitutionalized U.S. population (33). It contains two parts of data, interviews and examinations, based on demographic, socioeconomic, dietary, health-related questions, physiological measurements, laboratory tests and other information administered by highly trained medical personnel (34). All procedures were approved by CDC's Institutional Review Board (IRB) and all study subjects provided written informed consent (35). Participants were included if they were aged 18 years and above and were excluded if they had missing values on dietary tryptophan intake and/or all-cancer mortality information (n=2,920).

2.2.2 Measurements

Each participant underwent anthropometric measurements, provided a blood sample and completed a detailed questionnaire on sociodemographic, lifestyle and health-related factors (36). Dietary tryptophan intake, considered as an exposure variable in this study was estimated by a 24-hour recall methodology, collected through an automated interview with the Dietary Data Collection (DDC) system (37). Information on age (years), gender (male, female), education (years), poverty income ratio (continuous), healthy eating index score (continuous), race (Non-Hispanic White, Non-Hispanic Black, Mexican-American and other race), smoker (yes, no), drinker (yes, no), regular exercise (yes, no), diabetes (yes, no), hypertension (yes, no) and cancer (yes, no) was based on self-report during the questionnaire portion of the NHANES III survey. Body mass index (BMI) was calculated as the ratio of weight (kg) to the square of height (m). Poverty income ratio is the ratio of the midpoint of observed family income category to the official poverty threshold (38). C-reactive protein level was measured by high-sensitivity latex-enhanced nephelometry by CDC (39). Smokers were defined as adults who have smoked 100 cigarettes in their lifetime and who currently smoke cigarettes. Drinkers were defined as individuals who reported having at least 12 drinks in the last 12 months. Hypercholesterolemia was defined as a cholesterol level greater than 239 mg/dL (40).

To determine the final mortality status of every participant, multiple sources of information were utilized by NCHS, including the National Death Index (NDI), the Second Longitudinal Study of Aging (LSOA II), the Centers for Medicare and Medicaid Services (CMS) and death certificates (41). The outcome was all-cancer mortality status ascertained by NDI (42). Death due to cancer was defined as ICD-10 coding C00-C97 (43).

2.2.3 Statistical Analysis

According to the analysis guidelines downloaded from NHANES III (44), mean, proportion and confidence interval (CI) of variables were calculated, considering the complex, stratified sampling design by applying weights, strata and sampling unit values to produce estimates of the U.S. population. Baseline characteristics are described by quintiles of dietary tryptophan intake separately for participants with cancer and non-cancer subjects. Continuous variables and categorical variables are described as mean (95% CI) and percentage (95% CI) respectively. General linear models (for continuous variables) and chi-square tests (for categorical variables) were conducted to assess univariate relations among different groups.

Cox proportional hazards (CPH) models were used to evaluate the association of dietary tryptophan intake with all-cancer mortality by calculating hazard ratios (HRs) and its 95% CIs. Survival time was defined as the months between NHANES III interview date and death or census date.

To check the PH assumption of the Cox regression models, a graphical method based on the Kaplan-Meier test was adopted (45). According to the present analysis, dietary tryptophan intake satisfied the PH assumption ($P > 0.05$). Potential confounders we selected were well-established or biological interest factors.

Statistical analysis was conducted in R 4.1.1 and MR Base platform (32) (http://app.mrbase.org/). All *P*-values were two-tailed and associations were considered statistically significant at $P < 0.05$.

2.3 Data Availability

The summary statistics for tryptophan and kynurenine GWAS by Shin et al. (23) are available at http://mips.helmholtz-muenchen.de/proj/GWAS/gwas/index.php. The breast (90), lung (91), prostate (92) and ovarian (93) cancer GWAS summary data are derived from https://gwas.mrcieu.ac.uk/. The data of NHANES III are available at https://wwwn.cdc.gov/nchs/nhanes/nhanes3/default.aspx.

3 RESULTS

3.1 Information on the Selected SNPs and the Population Involved in the Study

General selection process was reflected by the Manhattan plots of the SNPs strongly associated with circulating tryptophan and kynurenine levels (**Supplementary Figure S1**). By drawing quantile-quantile (Q-Q) plots and calculating genomic inflation factors, the conclusion was that the selected SNPs and their corresponding traits were significantly related (**Supplementary Figure S2**). Briefly, 18 tryptophan-related SNPs were identified, explaining 3.80% of the circulating tryptophan levels' variance and 4 kynurenine-related SNPs were selected, explaining 1.19% of

circulating kynurenine levels' variance. The strongest signal identified in association to tryptophan was rs13122250 ($P = 8.95\times10^{-12}$) on chromosome 4, which has not been investigated yet. The second strongest signal was rs1016522 ($P = 1.59\times10^{-10}$), identified within the HMHB1 gene, which is responsible for the generation of immune response after recognition by specific T cells (50), and involved in adaptive immune response, cellular response to tumor necrosis factor and positive regulation of interferon-gamma production (51). We have identified other genomic loci with less obvious linkage to tryptophan, such as TGFBR3 (rs284191, $P = 1.97\times10^{-9}$), ERGIC1 (rs1559063, $P = 7.82\times10^{-9}$), DGKB (rs38271, $P = 1.19\times10^{-8}$), FTO (rs2111118, $P = 1.21\times10^{-8}$), ZPR1 (rs603446, $P = 1.38\times10^{-8}$), EDIL3 (rs1373962, $P = 2.71\times10^{-8}$), P3H2 (rs710580, $P = 3.57\times10^{-8}$), GREB1 (rs7584842, $P = 4.15\times10^{-8}$). TGFBR3 is involved in immune response (52), while FTO (53) and P3H2 (54) influence tumor occurrence and development. As for kynurenine, the strongest signal identified was rs8051149 within the SLC7A5 gene ($P = 9.07\times10^{-26}$), which is responsible for L-tryptophan transmembrane transport (55), the positive regulation of cytokine production in immune response, and the positive regulation of interferon-gamma production (56). Other associations with no obvious link to kynurenine include SH2B3 (rs3184504, $P = 6.05\times10^{-18}$), which is also associated with tumor occurrence and development (57) and IDO2 (rs10085935, $P = 3.33\times10^{-9}$). IDO2 catalyzes the first rate limiting step of the tryptophan catabolism and kynurenine pathway (58) and is involved in immune regulation, however, it may not have a significant role in tryptophan-related tumoral resistance (59). These SNPs regarded here as instrumental variables, have been verified by previous studies (60–63) (**Supplementary Table S6**). Detailed characteristics of the SNPs strongly associated with site-specific cancers were summarized in **Supplementary Table S2**. Detailed information of the metabolites (tryptophan and kynurenine) and population data included in the bidirectional MR analysis were described in **Supplementary Tables S3–S5**.

The baseline characteristics by quintiles of dietary tryptophan intake for participants were described in **Supplementary Table S7** separately. Participants with lower dietary tryptophan intake level were more likely to be older, have less years in education, be poorer, have higher C-reactive protein level, have lower healthy eating index score, be females, be non-Hispanic whites, smoke less, drink less, have less regular exercise, be hypertensive patients, have cancers and higher cancer mortality.

3.2 Results of Bidirectional MR Analysis of Circulating Tryptophan Levels and Site-Specific Cancers

In MR analysis, genetic predisposition to a lower circulating tryptophan level was not significantly associated with the risk of breast cancer (β 0.57; 95% CI -0.36-1.50, $P = 0.23$), lung cancer (β 0.08; 95% CI -1.54-1.71, $P = 0.92$), prostate cancer (β -0.92; 95% CI -2.04-0.20, $P = 0.11$) and ovarian cancer (β 1.39; 95% CI -0.43-3.20, $P = 0.14$) (**Figure 2**). Since significant heterogeneity for the associations of circulating tryptophan levels with prostate cancer and ovarian cancer were obtained by Cochran's Q test, a multiplicative random effects model (inverse variance weighted regression) was adopted to re-estimate causal effects again (prostate cancer, $P = 0.11$; ovarian cancer, $P = 0.14$), which indicated similar results. Because MR-Egger regression showed significant horizontal pleiotropy for the associations of circulating tryptophan levels with prostate cancer and ovarian cancer, the results of MR Egger based on sensitivity analysis were used to estimate the causal relationships (prostate cancer: β 1.18;

Outcome	Cases/Participants	β(95%CI)	P-value	Nsnp	MR-Egger regression	MR-PRESSO global test	Cochran's Q test
Breast cancer	46785/42892			18	0.383	0.368	0.451
Inverse variance weighted		0.572(-0.360~1.504)	0.229				
Weighted median		0.410(-0.882~1.702)	0.534				
Simple median		0.215(-1.060~1.491)	0.741				
Maximum likelihood		0.587(-0.357~1.530)	0.223				
Penalised weighted median		0.410(-0.865~1.685)	0.528				
Lung cancer	11348/15861			18	0.483	0.338	0.401
Inverse variance weighted		0.081(-1.544~1.707)	0.922				
Weighted median		0.219(-2.025~2.463)	0.848				
Simple median		0.303(-1.928~2.533)	0.790				
Maximum likelihood		0.084(-1.528~1.696)	0.918				
Penalised weighted median		0.219(-2.049~2.487)	0.850				
Prostate cancer	79148/61106			18	0.001*	0.115	0.001
Inverse variance weighted		-0.922(-2.041~0.197)	0.106				
Weighted median		-0.707(-1.774~0.359)	0.194				
Simple median		-0.898(-1.881~0.084)	0.073				
Maximum likelihood		-0.921(-1.678~-0.163)	0.017*				
Penalised weighted median		-0.704(-1.765~0.358)	0.194				
Ovarian cancer	25509/40941			18	0.001*	0.135	0.002
Inverse variance weighted		1.385(-0.430~3.200)	0.135				
Weighted median		1.540(-0.314~3.394)	0.104				
Simple median		1.945(0.113~3.777)	0.037*				
Maximum likelihood		1.453(0.208~2.697)	0.022*				
Penalised weighted median		2.318(0.404~4.231)	0.018*				

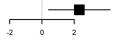


FIGURE 2 | Forest plot for associations between circulating tryptophan levels and the risk of site-specific cancers. β, beta value; CI, confidence interval; Nsnp, number of the SNPs; *P < 0.05.

95% CI -14.19-16.55, $P = 0.88$) (ovarian cancer: β -2.36; 95% CI -27.48-22.77, $P = 0.88$), which were consistent with the results of the IVW method (30).

During reverse MR analysis, the causal effects of site-specific cancers on circulating tryptophan levels were generally not significant (**Figure 3**). Due to the existence of horizontal pleiotropic and heterogeneity suggested by MR-Egger regression and Cochran's Q test, the weighted median method was conducted to verify the association of prostate cancer with circulating tryptophan levels, which displayed consistent results compared with IVW approach (β -0.003; 95% CI -0.008-0.001, $P = 0.12$). The forest plots and scatter plots for the causal effects of exposures on corresponding outcomes were exhibited in **Supplementary Figures S3–S10**.

3.3 Results of Bidirectional MR Analysis of Circulating Kynurenine Levels and Site-Specific Cancers

Using IVW method as our primary method, we found the estimates of the causal relationship between circulating kynurenine levels and the risk of breast cancer (β -0.37; 95% CI -0.90-0.16, $P = 0.17$), lung cancer (β -0.15; 95% CI -2.56-2.27, $P = 0.91$), prostate cancer (β -0.24; 95% CI -0.78-0.30, $P = 0.38$) and ovarian cancer (β 0.59; 95% CI -0.15-1.33, $P = 0.12$) were not significant in MR analysis (**Figure 4**). Just as described before, inverse variance weighted regression should be regarded as a main method during MR analysis of circulating kynurenine levels with lung cancer (ovarian cancer, $P = 0.91$), which suggested no significant associations of circulating kynurenine levels with lung cancer.

The reverse MR analysis showed lung cancer, but not other cancers, had significant genetic correlation with circulating kynurenine levels (**Figure 5**). Considering there were only three instrumental variables for lung cancer explaining 0.39% of the variance of the risk of lung cancer, the precision of this estimate may be relatively limited. Since the *P*-value of MR-PRESSO global test on the reverse MR analysis of prostate cancer and circulating kynurenine levels reached the significant level, MR Egger based on sensitivity analysis was used to estimate causal effects again (β 0.012; 95% CI 0.001-0.023, $P = 0.03$), which indicated there might be a weak correlation between prostate cancer and circulating kynurenine levels. The forest plots and scatter plots of causal effects from exposures to corresponding outcomes were shown in **Supplementary Figures S11–S18**.

3.4 Observational Study of Dietary Tryptophan Intake With All-Cancer Mortality

Supplementary Table S8 presents the association between dietary tryptophan intake and all-cancer mortality. During the analysis of the overall sample, model 1 adjusting for age, gender and race did not show any association in quintile 2 (HR 1.004; 95% CI 0.739-1.365), quintile 3 (HR 0.997; 95% CI 0.703-1.413), quintile 4 (HR 1.137; 95% CI 0.825-1.567) and quintile 5 (HR 1.154; 95% CI 0.834-1.597) (quintile 1 was considered the reference group, $P_{trend} = 0.75$).With adjustments for age, gender, race, education, poverty income ratio, smoker, drinker and regular exercise in model 2, HRs (95% CIs) in quintile 2 (HR 1.066; 95% CI 0.788-1.442), quintile 3 (HR 0.994; 95% CI 0.703-1.413), quintile 4 (HR 1.163; 95% CI 0.856-1.581) and quintile 5 (HR 1.233; 95% CI 0.942-1.612) suggested no statistically significant ($P_{trend} = 0.57$). After further adjusting for healthy eating index score, C-reactive protein and diabetes in model 3, HRs (95% CIs) for all-cancer mortality were 1.151 (0.847-1.563) in quintile 2, 1.078 (0.757-1.533) in quintile 3, 1.205 (0.881-1.648) in quintile 4 and 1.160 (0.872-1.545) in quintile

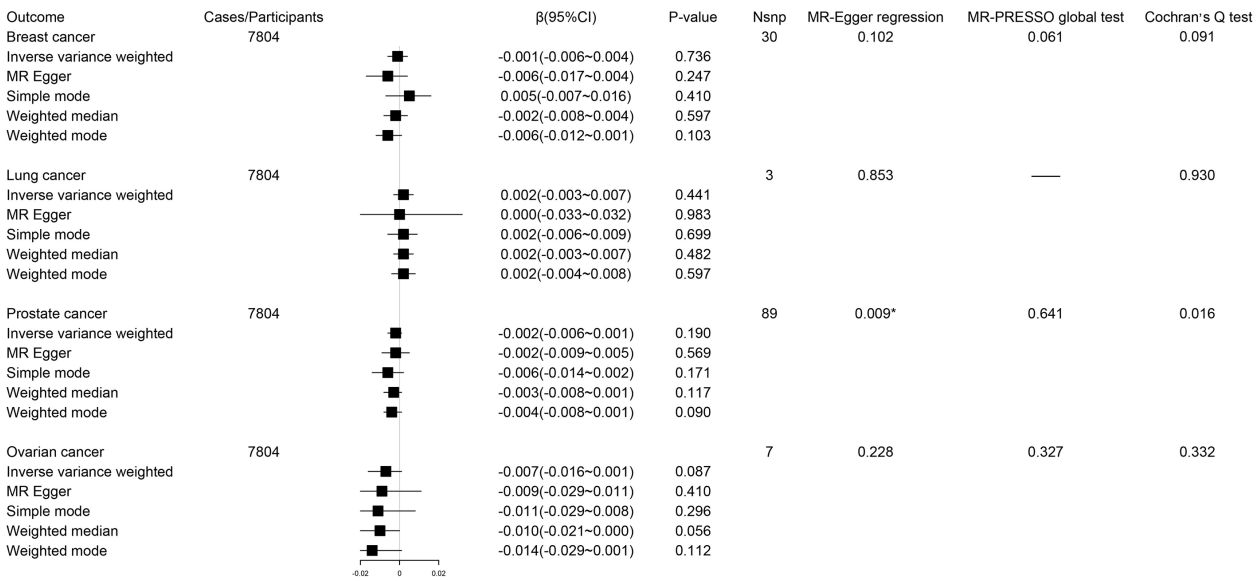


FIGURE 3 | Forest plot for associations between site-specific cancers and circulating tryptophan levels. β, beta value; CI, confidence interval; Nsnp, number of the SNPs; *P < 0.05.

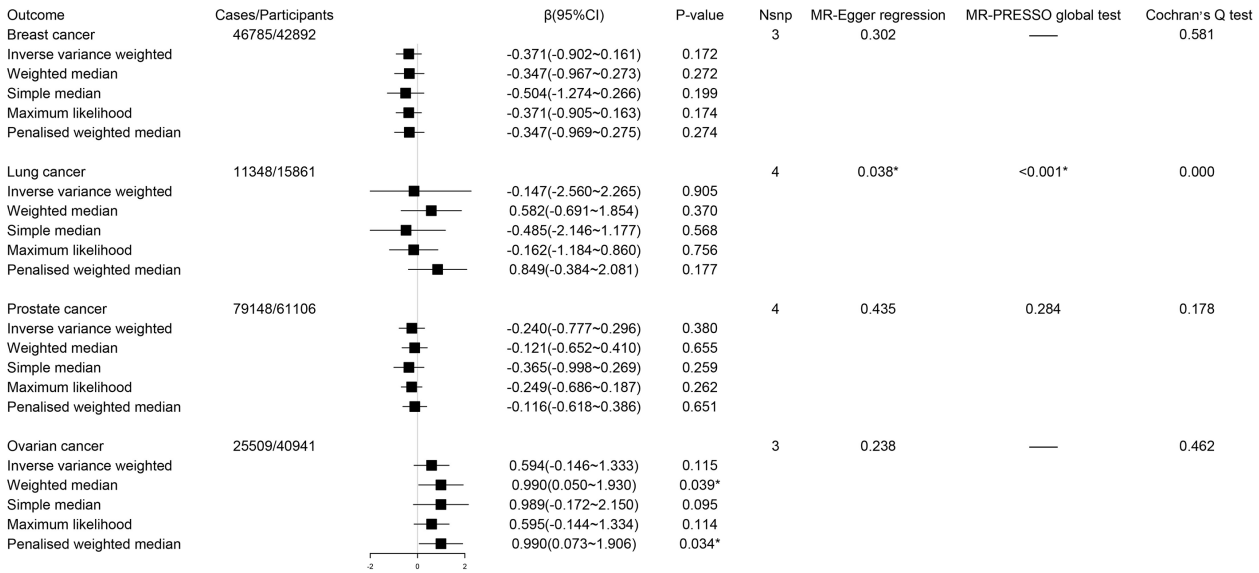


FIGURE 4 | Forest plot for associations between circulating kynurenine levels and the risk of site-specific cancers. β, beta value; CI, confidence interval; Nsnp, number of the SNPs; *P < 0.05.

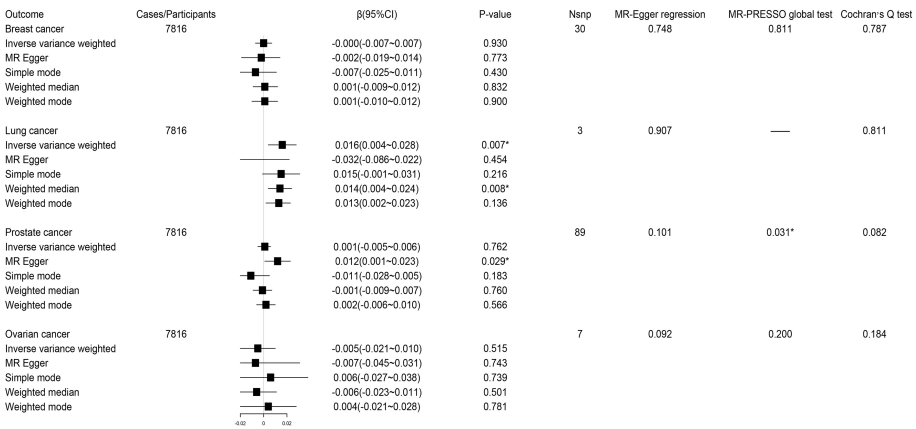


FIGURE 5 | Forest plot for associations between site-specific cancers and circulating kynurenine levels. β, beta value; CI, confidence interval; Nsnp, number of the SNPs; *P < 0.05.

5, which still remained not significant ($P_{trend} = 0.80$). Furthermore, the results of the COX regression model analysis conducted in cancer and non-cancer population were consistent with that of the whole study population. Generally speaking, hazard ratios and 95% CIs for all-cancer mortality by quintiles of dietary tryptophan intake did not show any significant effects.

4 DISCUSSION

In this study, the results of the MR analysis following sensitivity analysis, did not provide any evidence to prove the causal relationship between tryptophan and kynurenine and the risk of site-specific cancers. Moreover, after statistical analysis of dietary tryptophan intake and all-cancer mortality in NHANES III, we did not find any possible connection.

Cancer, which is characterized by uncontrolled growth and metastasis, is a general term for a group of multiple diseases that can affect any part of the body, and it is the second leading cause of death worldwide. The occurrence and development of cancer is closely related to the immune system. The theory of cancer immune monitoring proves the interaction between cancer and the immune system (64). A large amount of evidence shows that both the innate and the acquired immune response can identify

and eliminate tumors and mutated cancer cells are easily recognized and eliminated by the human immune system.

In recent years, more and more factors thought to have anti-tumor immunosuppressive effects have been discovered. The promoting effect of the reduction of tryptophan on the risk of cancer has been found in some observational studies, which basically involves the most types of cancer.

Two points of view have been proposed to explain how reduction of tryptophan plays an important role in immune process. One view is that the reduction of tryptophan inhibits the proliferation of immune cells by significantly reducing the content of tryptophan (12, 65, 66). Another point of view supports that it is not tryptophan itself which is responsible for this effect, but rather the changes of some metabolites involved in the tryptophan catabolism pathways that play a leading role in immunosuppression (67, 68). A generally accepted view is that tryptophan can be catabolized by immune cells and cancer cells at sites of immunodominance, inflammation, and tumorigenesis.

The body suppresses the production of antigen-specific T cells and limits excessive immune responses by depleting tryptophan and accumulating tryptophan catabolites with immunosuppressive effects. For example, after observing the chronic inflammation C57BL/6 strain in mice for a week, which was produced by threating them with phorbol myristate acetate, researchers found that the decrease in tryptophan was related to colon cancer and other inflammation-driven cancers (17). After measuring and analyzing the levels of tryptophan metabolites in 80 patients with colorectal cancer, it was found that the decrease in plasma tryptophan concentrations was associated with more advanced cancers (69). Another cross-sectional study of 200 patients with T cells leukemia/lymphoma designed to screen for immune activation-related biomarkers showed that lower concentration of plasma tryptophan was associated with shorter survival time in cancer patients (21). In addition, the viewpoint that the common phenomenon of decreased tryptophan and increased kynurenine concentrations in peripheral blood predicting enhanced tryptophan catabolism in cancer patients is related to the activation of pro-inflammatory cytokines, tumor progression and the occurrence of adverse clinical outcomes is proposed in more and more researches (12).

However, after excluding confounding factors such as socio-economic, diet and lifestyle, our study did not find any clear causal association between circulating tryptophan levels and the risk of site-specific cancers. We speculate the reduction in concentrations of tryptophan may not have enough effects on the immune response to tumors to change the immune system's role in the development of cancer. Alternatively other substances of the tryptophan degradation pathway affect the occurrence and development of cancer, rather than the decrease of tryptophan itself.

This is also consistent with other previous studies, in which the association between tryptophan concentrations and the occurrence and development of cancer was not significant. Studies on the changes in tryptophan metabolism pathways during pregnancy and infection found that tryptophan metabolites were the key regulators that regulate the behavior of immune cell behavior, and decrease in tryptophan was only an accompanying phenomenon indicating the changes in pathways (8). Although in most *in vitro* experiments, increasing the concentration of tryptophan in culture medium can restore the growth of cancer cells, bacteria or parasites, there are still many potential reasons that can be used to suspect this view. One reason is that tryptophan depletion experiments performed under cell culture conditions can not fully represent the internal environment of the body associated with infection. Another is that most bacteria can synthesize the required tryptophan by themselves, which means the effect of local tryptophan reduction may be amplified.

In the study of the relationship between dietary tryptophan intake and the risk of cancer death in NHANES III and in consistency with the MR analysis based on the whole study population and the cancer-affected population, we came to the conclusion that dietary tryptophan intake has little effect on the occurrence and development of cancer death after adjusting for age, gender, race, education, poverty income ratio, smoking, drinking, regular exercise, healthy eating index score, serum C-reactive protein and diabetes. The reason may be that the demand for tryptophan is multifaceted, such as protein synthesis, neuron protection, the maintenance of signal pathways, immune tolerance and the synthesis of nicotinamide adenine dinucleotide etc. Circulating kynurenine is primarily derived from endogenous tryptophan catabolism and current data on the presence and content of kynurenine in food are unexpectedly sparse, and thus it was not possible to conduct an association analysis between dietary kynurenine intake and cancer mortality based on NHANES data.

The decrease in circulating tryptophan concentrations in the human body is believed to be mainly caused by the enhancement of the tryptophan catabolism (3). Generally speaking, in the process of tryptophan catabolism, circulating tryptophan concentrations continues to decrease, and its downstream metabolites are continuously produced at the same time. Based on the above description, we speculate that the regulation of immune response in the tumor microenvironment produced by tryptophan catabolism pathways, is mostly due to the impact of substances of the dominant metabolic pathways, other than tryptophan.

Tryptophan is an important precursor of biologically active metabolites including tryptamine, serotonin, melatonin, kynurenine, kynurenic acid, quinolinic acid and nicotinamide adenine dinucleotide, which are mainly produced through three different metabolic pathways: serotonin pathway, kynurenine pathway and indole pathway (70, 71). Among them, more than 95% of free tryptophan is degraded *via* the kynurenine pathway (72).

More and more evidence from multiple laboratories indicate that the increase of kynurenine and its metabolites with immunomodulatory properties is the main mechanism of promoting immune tolerance in the tryptophan catabolism. Disturbances of the kynurenine pathway are thought to be related to central nervous system diseases, malignant tumors, inflammatory bowel diseases and cardiovascular diseases. Kynurenine can reduce the activity of natural killer cells,

dendritic cells and T cells. Kynurenic acid can promote monocyte extravasation and control the release of cytokines (6).

As an endogenous pro-tumor proliferation ligand, kynurenine can bind to aromatic hydrocarbon receptors and activate aromatic hydrocarbon receptors to exert its biological effects, which implies that the high levels of kynurenine may increase the proliferation and migration of cancer cells and help tumors escape from immune surveillance (73). During *in vitro* experiments, exogenous addition of kynurenine, 3-hydroxykynurenine, 3-hydroxyanthranilic acid and quinolinic acid inhibited the proliferation of cultured T cells and induced them to apoptosis (5, 74).

However, our study did not find any evidence on the association of circulating kynurenine levels with the risk of site-specific cancers, which may be due to the slight alteration of kynurenine and its downstream metabolites in the tumor microenvironment. Similar to the effect of decreased tryptophan on anti-tumor immune response in the body, the effect of increased kynurenine on immune system may also be little. During a number of observational studies, although patients with gynecological cancer (13), T cells leukemia (21), colorectal cancer (12), malignant melanoma (11), malignant glioma (3, 10) and lung cancer (9) had reduced systemic tryptophan levels, no increase in the concentration of kynurenine pathway metabolites in the blood was observed.

After summarizing and analyzing all the results of this research, we put forward two guesses: 1) the changes of tryptophan and its main metabolite kynurenine are related to the immune response, but the inhibitory effects of the decrease of tryptophan and the increase of kynurenine on the anti-tumor immune response may not be enough to affect the occurrence and development of cancer; 2) there is a strong causal link between other certain substances in tryptophan metabolism pathways and the risk of cancer. The changes in circulating tryptophan and kynurenine are only accompanying phenomena of the progress of these pathways.

Preclinical studies have shown TDO and IDO, the main rate-limiting enzymes that can regulate kynurenine pathway in patients with malignant tumors, can regulate the tryptophan-mediated tumor immune escape response through depleting tryptophan and accumulating kynurenine in the tumor microenvironment (1). But whether and how the decrease of tryptophan or the increase of kynurenine promotes T cells-mediated tumor rejection *in vivo* remains to be studied. Existing studies provide some possible mechanisms. For example, the reduction of tryptophan leads to anergy and apoptosis of T cells through the general amino acid control non-derepressible 2 (75) (GCN2) and the integrated stress response (76) (IRS) pathways and an increase in kynurenine inhibits T cells differentiation through the aryl hydrocarbon receptor (3) (AHR) pathway. The fact that a variety of human tumors express TDO and IDO indicates the therapeutic potential of targeted drugs to inhibit TDO or IDO in the process of cancer treatment (77–79). However, current clinical trials have shown although some of the key enzyme inhibitors have achieved the expected effects in early cancer immunotherapy, the results of phase III trial are negative, which suggests we lack a precise understanding of the exact downstream mechanism of immunosuppression related to tryptophan metabolism.

Components of the tryptophan catabolism that were previously associated to cancer, have been found to interact with pathways of the tumor microenvironment (80, 81). IDO for example, has been reported to be associated with changes in the complement pathway of the tumor microenvironment (82), while interferons which are potent inducers of immunomodulatory responses are mediated by IDO (83). IDO also regulates the activation of tumor suppressive regulatory T cells in the tumor microenvironment (84). Another important immunosuppressive cell population of the tumor microenvironment, the myeloid-derived suppressor cells (MDSCs) are recruited to tumors by an IDO-indirect mechanism (85). In conclusion, many physiological processes are capable of inducing IDO, and multiple factors may limit IDO expression and thus regulate IDO activity in physiological environments. Therefore, we should also acknowledge that interactions between genetics and environment may still increase the risk of cancer in association to the tryptophan and kynurenine pathways.

In addition, the important role of serotonin pathway in cancer progression and anti-tumor immune response is being confirmed by more and more researches. Serotonin is an inflammatory mediator (86) related to the proliferation and invasion of various cancer cells (87). Studies on triple-negative breast cancer have shown serotonin promotes the invasion and proliferation of tumor cells through its receptor subtype 5-HT_7 (88). In certain cancers and gliomas, serotonin has shown to promote tumor growth and survival (89). In animal experiments, serotonin regulates the expression of specific serotonin receptors in cancer cells through a process called serotonylation and up-regulates the expression level of programmed cell death ligand 1 (PD-L1), which is related to the suppression of the immune system (90).

Our study has several strengths. First, to our knowledge, this is the first bidirectional MR analysis on the relationship between circulating tryptophan or kynurenine and site-specific cancers, which strengthens the causal inference through diminishing residual confounding and other biases. Second, we used the summary statistics of large-scale GWASs. Third, to examine the possible associations, NHANES III data and four kinds of site-specific cancers from different data sources were chosen as our outcomes to increase the statistical power to detect weak associations.

There are still several limitations in our study. The major limitation is that only circulating tryptophan and kynurenine has been studied in the present study, however, other metabolites involved in the tryptophan metabolic pathways have not been studied yet. In addition, the size of the populations used to select genetic instruments strongly associated with tryptophan or kynurenine may not be large enough, which may affect the choice of instrumental variables. Third, despite several large-scale genetic consortia we utilized, the variation of site-specific cancers explained by the SNPs was still relatively small, which may limit the statistical power and precision for the MR analysis. Lastly, our results are mainly based on participants of European ancestry and may not be applicable to other ethnic populations. To date,

most GWAS performed are primarily conducted on European populations. Although GWAS in Asian and Latin American populations are increasingly being conducted, the population data available generally suffer from insufficient sample sizes and limited geographic distribution of the population. In addition, differential frequencies of genetic variants that exist between populations with different genetic backgrounds can lead to spurious associations between genetic variants and outcomes. Therefore, currently no additional data of diverse ethnic backgrounds can be used in this study. We expect that future GWAS development, will allow this application.

Based on the "common disease-common variant" hypothesis, GWAS have been extensively conducted to dissect the genetic components of complex diseases and quantitative traits (90). However, the identified disease-associated common variants can only explain small part of the corresponding disease heritability. Since, the MR analysis approach relies on GWAS, this inevitably leads to missing heritability of lower frequency variants (91). Association studies of less common variants, include adaptive burden tests, variance-component tests, combined burden and variance-component tests, combined association in the presence of linkage test, sum of powered score test and exponential combination test (92, 93). We believe these methods can be implemented in the future to explore missing heritability, fill the gaps of the MR analysis approach and augment current findings.

In summary, this MR analysis did not find evidence to support the causal relationship between circulating tryptophan or kynurenine concentration and cancer. Given the existing results, whether changes in tryptophan metabolism pathways may influence the risk of cancer needs further and broader researches such as clarifying the effects of circulating tryptophan or kynurenine on the immune response of the body and carrying out research on the relationship between other metabolites in the tryptophan metabolic pathways and cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

All studies included in this analysis were approved by local review boards and performed in accordance with the Declaration of Helsinki. All participants gave written informed consent to participate in the study.

AUTHOR CONTRIBUTIONS

YL is the guarantor of this work and has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. YL and RL conceived the study design. XW, YZ and ZW did the statistical analysis. XW, RL, CW and LL repeated and validated the statistical analysis. XW, RL, YZ, LL and LW wrote the manuscript. All authors provided critical insights of the manuscript. All authors contributed to the article and approved the submitted version

FUNDING

All authors were supported by funding from National Natural Science Foundation (82030100 to YL).

ACKNOWLEDGMENTS

The authors thank all the researchers for sharing the data of genome-wide association study (GWAS) meta-analyses on exposures and outcomes. The authors also acknowledge volunteers, patients and medical personnel involved in the participating studies for their contribution.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.852718/full#supplementary-material

REFERENCES

1. Platten M, Nollen EAA, Röhrig UF, Fallarino F, Opitz CA. Tryptophan Metabolism as a Common Therapeutic Target in Cancer, Neurodegeneration and Beyond. *Nat Rev Drug Discovery* (2019) 18(5):379–401. doi: 10.1038/s41573-019-0016-5
2. Schwarcz R, Stone TW. The Kynurenine Pathway and the Brain: Challenges, Controversies and Promises. *Neuropharmacology* (2017) 112(Pt B):237–47. doi: 10.1016/j.neuropharm.2016.08.003
3. Opitz CA, Litzenburger UM, Sahm F, Ott M, Tritschler I, Trump S, et al. An Endogenous Tumour-Promoting Ligand of the Human Aryl Hydrocarbon Receptor. *Nature* (2011) 478(7368):197–203. doi: 10.1038/nature10491
4. Ino K, Yamamoto E, Shibata K, Kajiyama H, Yoshida N, Terauchi M, et al. Inverse Correlation Between Tumoral Indoleamine 2,3-Dioxygenase Expression and Tumor-Infiltrating Lymphocytes in Endometrial Cancer: Its Association With Disease Progression and Survival. *Clin Cancer Res* (2008) 14 (8):2310–7. doi: 10.1158/1078-0432.Ccr-07-4144
5. Frumento G, Rotondo R, Tonetti M, Damonte G, Benatti U, Ferrara GB. Tryptophan-Derived Catabolites Are Responsible for Inhibition of T and Natural Killer Cell Proliferation Induced by Indoleamine 2,3-Dioxygenase. *J Exp Med* (2002) 196(4):459–68. doi: 10.1084/jem.20020121
6. Cervenka I, Agudelo LZ, Ruas JL. Kynurenines: Tryptophan's Metabolites in Exercise, Inflammation, and Mental Health. *Sci (New York NY)* (2017) 357 (6349):eaaf9794. doi: 10.1126/science.aaf9794
7. Théate I, van Baren N, Pilotte L, Moulin P, Larrieu P, Renauld JC, et al. Extensive Profiling of the Expression of the Indoleamine 2,3-Dioxygenase 1 Protein in Normal and Tumoral Human Tissues. *Cancer Immunol Res* (2015) 3(2):161–72. doi: 10.1158/2326-6066.Cir-14-0137
8. Badawy AA, Namboodiri AM, Moffett JR. The End of the Road for the Tryptophan Depletion Concept in Pregnancy and Infection. *Clin Sci (London Engl 1979)* (2016) 130(15):1327–33. doi: 10.1042/cs20160153
9. Suzuki Y, Suda T, Furuhashi K, Suzuki M, Fujie M, Hahimoto D, et al. Increased Serum Kynurenine/Tryptophan Ratio Correlates With Disease

Progression in Lung Cancer. *Lung Cancer (Amsterdam Netherlands)* (2010) 67(3):361–5. doi: 10.1016/j.lungcan.2009.05.001
10. Zhai L, Dey M, Lauing KL, Gritsina G, Kaur R, Lukas RV, et al. The Kynurenine to Tryptophan Ratio as a Prognostic Tool for Glioblastoma Patients Enrolling in Immunotherapy. *J Clin Neurosci* (2015) 22(12):1964–8. doi: 10.1016/j.jocn.2015.06.018
11. Weinlich G, Murr C, Richardsen L, Winkler C, Fuchs D. Decreased Serum Tryptophan Concentration Predicts Poor Prognosis in Malignant Melanoma Patients. *Dermatol (Basel Switzerland)* (2007) 214(1):8–14. doi: 10.1159/000096906
12. Huang A, Fuchs D, Widner B, Glover C, Henderson DC, Allen-Mersh TG. Serum Tryptophan Decrease Correlates With Immune Activation and Impaired Quality of Life in Colorectal Cancer. *Br J Cancer* (2002) 86 (11):1691–6. doi: 10.1038/sj.bjc.6600336
13. Schroecksnadel K, Winkler C, Fuith LC, Fuchs D. Tryptophan Degradation in Patients With Gynecological Cancer Correlates With Immune Activation. *Cancer Lett* (2005) 223(2):323–9. doi: 10.1016/j.canlet.2004.10.033
14. Liu Y, Zhou N, Zhou L, Wang J, Zhou Y, Zhang T, et al. Il-2 Regulates Tumor-Reactive Cd8(+) T Cell Exhaustion by Activating the Aryl Hydrocarbon Receptor. *Nat Immunol* (2021) 22(3):358–69. doi: 10.1038/s41590-020-00850-9
15. Lu J, Liu X, Liao YP, Wang X, Ahmed A, Jiang W, et al. Breast Cancer Chemo-Immunotherapy Through Liposomal Delivery of an Immunogenic Cell Death Stimulus Plus Interference in the Ido-1 Pathway. *ACS Nano* (2018) 12 (11):11041–61. doi: 10.1021/acsnano.8b05189
16. Liu Q, Zhai J, Kong X, Wang X, Wang Z, Fang Y, et al. Comprehensive Analysis of the Expression and Prognosis for Tdo2 in Breast Cancer. *Mol Ther Oncolyt* (2020) 17:153–68. doi: 10.1016/j.omto.2020.03.013
17. Muller AJ, Sharma MD, Chandler PR, Duhadaway JB, Everhart ME, Johnson BA3rd, et al. Chronic Inflammation That Facilitates Tumor Progression Creates Local Immune Suppression by Inducing Indoleamine 2,3 Dioxygenase. *Proc Natl Acad Sci USA* (2008) 105(44):17073–8. doi: 10.1073/pnas.0806173105
18. Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM, et al. Association Between Endometriosis and Risk of Histological Subtypes of Ovarian Cancer: A Pooled Analysis of Case-Control Studies. *Lancet Oncol* (2012) 13(4):385–94. doi: 10.1016/s1470-2045(11)70404-1
19. Pichler R, Fritz J, Heidegger I, Steiner E, Culig Z, Klocker H, et al. Predictive and Prognostic Role of Serum Neopterin and Tryptophan Breakdown in Prostate Cancer. *Cancer Sci* (2017) 108(4):663–70. doi: 10.1111/cas.13171
20. Lee SH, Mahendran R, Tham SM, Thamboo TP, Chionh BJ, Lim YX, et al. Tryptophan-Kynurenine Ratio as a Biomarker of Bladder Cancer. *BJU Int* (2021) 127(4):445–53. doi: 10.1111/bju.15205
21. Giusti RM, Maloney EM, Hanchard B, Morgan OS, Steinberg SM, Wachter H, et al. Differential Patterns of Serum Biomarkers of Immune Activation in Human T-Cell Lymphotropic Virus Type I-Associated Myelopathy/Tropical Spastic Paraparesis, and Adult T-Cell Leukemia/Lymphoma. *Cancer Epidemiol Biomarkers Prev* (1996) 5(9):699–704.
22. Davey Smith G, Ebrahim S. What Can Mendelian Randomisation Tell Us About Modifiable Behavioural and Environmental Exposures? *BMJ (Clinical Res ed)* (2005) 330(7499):1076–9. doi: 10.1136/bmj.330.7499.1076
23. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An Atlas of Genetic Influences on Human Blood Metabolites. *Nat Genet* (2014) 46(6):543–50. doi: 10.1038/ng.2982
24. Chen L, Yang H, Li H, He C, Yang L, Lv G. Novel Insights Into Modifiable Risk Factors of Cholelithiasis: A Mendelian Randomization Study. *Hepatol (Baltimore Md)* (2021) 75(4):785–96. doi: 10.1002/hep.32183
25. Yu Z, Coresh J, Qi G, Grams M, Boerwinkle E, Snieder H, et al. A Bidirectional Mendelian Randomization Study Supports Causal Effects of Kidney Function On Blood Pressure. *Kidney Int* (2020) 98(3):708–16. doi: 10.1016/j.kint.2020.04.044
26. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The Uk Adult Twin Registry (Twinsuk Resource). *Twin Res Hum Genet* (2013) 16(1):144–9. doi: 10.1017/thg.2012.89
27. Wichmann HE, Gieger C, Illig T. Kora-Gen–Resource for Population Genetics, Controls and a Broad Spectrum of Disease Phenotypes. *Gesundheitswesen* (2005) 67(Suppl 1):S26–30. doi: 10.1055/s-2005-858226
28. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-Analysis of Genetic Studies Using Mendelian Randomization–a Multivariate Approach. *Stat Med* (2005) 24(14):2241–54. doi: 10.1002/sim.2100
29. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization With Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* (2016) 40(4):304–14. doi: 10.1002/gepi.21965
30. Burgess S, Thompson SG. Interpreting Findings From Mendelian Randomization Using the Mr-Egger Method. *Eur J Epidemiol* (2017) 32 (5):377–89. doi: 10.1007/s10654-017-0255-x
31. Verbanck M, Chen CY, Neale B, Do R. Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred From Mendelian Randomization Between Complex Traits and Diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7
32. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The Mr-Base Platform Supports Systematic Causal Inference Across the Human Phenome. *eLife* (2018) 7:e34408. doi: 10.7554/eLife.34408
33. Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94. Series 1: Programs and Collection Procedures. *Vital Health Stat 1* (1994) (32):1–407.
34. Shan Z, Rehm CD, Rogers G, Ruan M, Wang DD, Hu FB, et al. Trends in Dietary Carbohydrate, Protein, and Fat Intake and Diet Quality Among Us Adults, 1999-2016. *Jama* (2019) 322(12):1178–87. doi: 10.1001/jama.2019.13771
35. National Center for Health Statistics. *Research Ethics Review Board (Erb) Approval* (Accessed November 2, 2021).
36. Goyal A, Terry MB, Jin Z, Siegel AB. C-Reactive Protein and Colorectal Cancer Mortality in U.S. Adults. *Cancer Epidemiol Biomarkers Prev* (2014) 23 (8):1609–18. doi: 10.1158/1055-9965.Epi-13-0577
37. Briefel RR, Sempos CT. Dietary Methodology Workshop for the Third National Health and Nutrition Examination Survey. *Vital Health Stat Ser 4 Documents Committee Rep* (1992) 27):1–108.
38. Skelton JA, Cook SR, Auinger P, Klein JD, Barlow SE. Prevalence and Trends of Severe Obesity Among Us Children and Adolescents. *Acad Pediatr* (2009) 9 (5):322–9. doi: 10.1016/j.acap.2009.04.005
39. Shen G, Huang JY, Huang YQ, Feng YQ. The Relationship Between Telomere Length and Cancer Mortality: Data From the 1999-2002 National Healthy and Nutrition Examination Survey (Nhanes). *J Nutr Health Aging* (2020) 24(1):9–15. doi: 10.1007/s12603-019-1265-z
40. National Institutes of Health. *High Blood Cholesterol* (Accessed November 5, 2021).
41. Centers for Disease Control and Prevention. *National Center for Health Statistics* (Accessed November 6, 2021).
42. Centers for Disease Control and Prevention. *National Health and Nutrition Examination Survey Data Documentation, Codebook, and Frequencies* (Accessed November 7, 2021).
43. Anderson RN, Miniño AM, Hoyert DL, Rosenberg HM. Comparability of Cause of Death Between Icd-9 and Icd-10: Preliminary Estimates. *Natl Vital Stat Rep* (2001) 49(2):1–32.
44. National Center for Health Statistics. *Nhanes Iii (1988-1994) -Survey Methods and Analytic Guidelines* (Accessed November 7, 2021).
45. Hess KR. Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression. *Stat Med* (1995) 14(15):1707–23. doi: 10.1002/sim.4780141510
46. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association Analysis Identifies 65 New Breast Cancer Risk Loci. *Nature* (2017) 551(7678):92–4. doi: 10.1038/nature24284
47. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare Variants of Large Effect in Brca2 and Chek2 Affect Risk of Lung Cancer. *Nat Genet* (2014) 46(7):736–41. doi: 10.1038/ng.3002
48. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association Analyses of More Than 140,000 Men Identify 63 New Prostate Cancer Susceptibility Loci. *Nat Genet* (2018) 50(7):928–36. doi: 10.1038/s41588-018-0142-8
49. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 New Susceptibility Loci for Different Histotypes of Epithelial Ovarian Cancer. *Nat Genet* (2017) 49(5):680–91. doi: 10.1038/ng.3826
50. Spierings E, Wieles B, Goulmy E. Minor Histocompatibility Antigens–Big in Tumour Therapy. *Trends Immunol* (2004) 25(2):56–60. doi: 10.1016/j.it.2003.12.002

51. Dolstra H, Fredrix H, Preijers F, Goulmy E, Figdor CG, de Witte TM, et al. Recognition of a B Cell Leukemia-Associated Minor Histocompatibility Antigen by Ctl. *J Immunol (Baltimore Md 1950)* (1997) 158(2):560–5.
52. Adewoye AH, Nolan VG, Ma Q, Baldwin C, Wyszynski DF, Farrell JJ, et al. Association of Polymorphisms of Igf1r and Genes in the Transforming Growth Factor- Beta /Bone Morphogenetic Protein Pathway With Bacteremia in Sickle Cell Anemia. *Clin Infect Dis* (2006) 43(5):593–8. doi: 10.1086/506356
53. Li Z, Weng H, Su R, Weng X, Zuo Z, Li C, et al. Fto Plays an Oncogenic Role in Acute Myeloid Leukemia as a N(6)-Methyladenosine Rna Demethylase. *Cancer Cell* (2017) 31(1):127–41. doi: 10.1016/j.ccell.2016.11.017
54. Shah R, Smith P, Purdie C, Quinlan P, Baker L, Aman P, et al. The Prolyl 3-Hydroxylases P3h2 and P3h3 Are Novel Targets for Epigenetic Silencing in Breast Cancer. *Br J Cancer* (2009) 100(10):1687–96. doi: 10.1038/sj.bjc.6605042
55. Yan R, Zhao X, Lei J, Zhou Q. Structure of the Human Lat1-4f2hc Heteromeric Amino Acid Transporter Complex. *Nature* (2019) 568 (7750):127–30. doi: 10.1038/s41586-019-1011-z
56. Hayashi K, Jutabha P, Endou H, Sagara H, Anzai N. Lat1 Is a Critical Transporter of Essential Amino Acids for Immune Reactions in Activated Human T Cells. *J Immunol (Baltimore Md 1950)* (2013) 191(8):4080–5. doi: 10.4049/jimmunol.1300923
57. Willman CL. Sh2b3: A New Leukemia Predisposition Gene. *Blood* (2013) 122 (14):2293–5. doi: 10.1182/blood-2013-08-519843
58. Metz R, Duhadaway JB, Kamasani U, Laury-Kleintop L, Muller AJ, Prendergast GC. Novel Tryptophan Catabolic Enzyme Ido2 Is the Preferred Biochemical Target of the Antitumor Indoleamine 2,3-Dioxygenase Inhibitory Compound D-1-Methyl-Tryptophan. *Cancer Res* (2007) 67 (15):7082–7. doi: 10.1158/0008-5472.Can-07-1872
59. van Baren N, Van den Eynde BJ. Tryptophan-Degrading Enzymes in Tumoral Immune Resistance. *Front Immunol* (2015) 6:34. doi: 10.3389/fimmu.2015.00034
60. Li M, Kwok MK, Fong SSM, Schooling CM. Effects of Tryptophan, Serotonin, and Kynurenine on Ischemic Heart Diseases and Its Risk Factors: A Mendelian Randomization Study. *Eur J Clin Nutr* (2020) 74(4):613–21. doi: 10.1038/s41430-020-0588-5
61. Li M, Kwok MK, Fong SSM, Schooling CM. Indoleamine 2,3-Dioxygenase and Ischemic Heart Disease: A Mendelian Randomization Study. *Sci Rep* (2019) 9(1):8491. doi: 10.1038/s41598-019-44819-7
62. Bagheri M, Wang C, Shi M, Manouchehri A, Murray KT, Murphy MB, et al. The Genetic Architecture of Plasma Kynurenine Includes Cardiometabolic Disease Mechanisms Associated With the Sh2b3 Gene. *Sci Rep* (2021) 11 (1):15652. doi: 10.1038/s41598-021-95154-9
63. Cheng Y, Li Y, Benkowitz P, Lamina C, Köttgen A, Sekula P. The Relationship Between Blood Metabolites of the Tryptophan Pathway and Kidney Function: A Bidirectional Mendelian Randomization Analysis. *Sci Rep* (2020) 10 (1):12675. doi: 10.1038/s41598-020-69559-x
64. Wahid B, Ali A, Rafique S, Waqar M, Wasim M, Wahid K, et al. An Overview of Cancer Immunotherapeutic Strategies. *Immunotherapy* (2018) 10(11):999–1010. doi: 10.2217/imt-2018-0002
65. Lanser L, Kink P, Egger EM, Willenbacher W, Fuchs D, Weiss G, et al. Inflammation-Induced Tryptophan Breakdown Is Related With Anemia, Fatigue, and Depression in Cancer. *Front Immunol* (2020) 11:249. doi: 10.3389/fimmu.2020.00249
66. Sudar-Milovanovic E, Gluvic Z, Obradovic M, Zaric B, Isenovic ER. Tryptophan Metabolism in Atherosclerosis and Diabetes. *Curr Med Chem* (2022) 29(1):99–113. doi: 10.2174/0929867328666210714153649
67. Ala M. Tryptophan Metabolites Modulate Inflammatory Bowel Disease and Colorectal Cancer by Affecting Immune System. *Int Rev Immunol* (2021), 1–20. doi: 10.1080/08830185.2021.1954638
68. Hunt C, Macedo ECT, Suchting R, de Dios C, Cuellar Leal VA, Soares JC, et al. Effect of Immune Activation on the Kynurenine Pathway and Depression Symptoms - a Systematic Review and Meta-Analysis. *Neurosci Biobehav Rev* (2020) 118:514–23. doi: 10.1016/j.neubiorev.2020.08.010
69. Li Y, Hu N, Yang D, Oxenkrug G, Yang Q. Regulating the Balance Between the Kynurenine and Serotonin Pathways of Tryptophan Metabolism. *FEBS J* (2017) 284(6):948–66. doi: 10.1111/febs.14026
70. Le Floc'h N, Otten W, Merlot E. Tryptophan Metabolism, From Nutrition to Potential Therapeutic Applications. *Amino Acids* (2011) 41(5):1195–205. doi: 10.1007/s00726-010-0752-7
71. van der Goot AT, Nollen EA. Tryptophan Metabolism: Entering the Field of Aging and Age-Related Pathologies. *Trends Mol Med* (2013) 19(6):336–44. doi: 10.1016/j.molmed.2013.02.007
72. Stone TW, Stoy N, Darlington LG. An Expanding Range of Targets for Kynurenine Metabolites of Tryptophan. *Trends Pharmacol Sci* (2013) 34 (2):136–43. doi: 10.1016/j.tips.2012.09.006
73. Fallarino F, Grohmann U, Vacca C, Bianchi R, Orabona C, Spreca A, et al. T Cell Apoptosis by Tryptophan Catabolism. *Cell Death Differ* (2002) 9 (10):1069–77. doi: 10.1038/sj.cdd.4401073
74. Terness P, Bauer TM, Röse L, Dufter C, Watzlik A, Simon H, et al. Inhibition of Allogeneic T Cell Proliferation by Indoleamine 2,3-Dioxygenase-Expressing Dendritic Cells: Mediation of Suppression by Tryptophan Metabolites. *J Exp Med* (2002) 196(4):447–57. doi: 10.1084/jem.20020052
75. Munn DH, Sharma MD, Baban B, Harding HP, Zhang Y, Ron D, et al. Gcn2 Kinase in T Cells Mediates Proliferative Arrest and Anergy Induction in Response to Indoleamine 2,3-Dioxygenase. *Immunity* (2005) 22(5):633–42. doi: 10.1016/j.immuni.2005.03.013
76. Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M, et al. Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells. *Mol Cell* (2000) 6(5):1099–108. doi: 10.1016/s1097-2765(00)00108-8
77. Bessede A, Gargaro M, Pallotta MT, Matino D, Servillo G, Brunacci C, et al. Aryl Hydrocarbon Receptor Control of a Disease Tolerance Defence Pathway. *Nature* (2014) 511(7508):184–90. doi: 10.1038/nature13323
78. Muller AJ, DuHadaway JB, Donover PS, Sutanto-Ward E, Prendergast GC. Inhibition of Indoleamine 2,3-Dioxygenase, an Immunoregulatory Target of the Cancer Suppression Gene Bin1, Potentiates Cancer Chemotherapy. *Nat Med* (2005) 11(3):312–9. doi: 10.1038/nm1196
79. Pilotte L, Larrieu P, Stroobant V, Colau D, Dolusic E, Frédérick R, et al. Reversal of Tumoral Immune Resistance by Inhibition of Tryptophan 2,3-Dioxygenase. *Proc Natl Acad Sci USA* (2012) 109(7):2497–502. doi: 10.1073/pnas.1113873109
80. Munn DH, Mellor AL. Ido in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance. *Trends Immunol* (2016) 37(3):193–207. doi: 10.1016/j.it.2016.01.002
81. Muller AJ, Manfredi MG, Zakharia Y, Prendergast GC. Inhibiting Ido Pathways to Treat Cancer: Lessons From the Echo-301 Trial and Beyond. *Semin Immunopathol* (2019) 41(1):41–8. doi: 10.1007/s00281-018-0702-0
82. Li M, Bolduc AR, Hoda MN, Gamble DN, Dolisca SB, Bolduc AK, et al. The Indoleamine 2,3-Dioxygenase Pathway Controls Complement-Dependent Enhancement of Chemo-Radiation Therapy Against Murine Glioblastoma. *J Immunother Cancer* (2014) 2:21. doi: 10.1186/2051-1426-2-21
83. Orabona C, Puccetti P, Vacca C, Bicciato S, Luchini A, Fallarino F, et al. Toward the Identification of a Tolerogenic Signature in Ido-Competent Dendritic Cells. *Blood* (2006) 107(7):2846–54. doi: 10.1182/blood-2005-10-4077
84. Nishikawa H, Sakaguchi S. Regulatory T Cells in Cancer Immunotherapy. *Curr Opin Immunol* (2014) 27:1–7. doi: 10.1016/j.coi.2013.12.005
85. Smith C, Chang MY, Parker KH, Beury DW, DuHadaway JB, Flick HE, et al. Ido Is a Nodal Pathogenic Driver of Lung Cancer and Metastasis Development. *Cancer Discov* (2012) 2(8):722–35. doi: 10.1158/2159-8290.Cd-12-0014
86. Shajib MS, Khan WI. The Role of Serotonin and Its Receptors in Activation of Immune Responses and Inflammation. *Acta Physiol (Oxford England)* (2015) 213(3):561–74. doi: 10.1111/apha.12430
87. Sarrouilhe D, Mesnil M. Serotonin and Human Cancer: A Critical View. *Biochimie* (2019) 161:46–50. doi: 10.1016/j.biochi.2018.06.016
88. Gautam J, Banskota S, Regmi SC, Ahn S, Jeon YH, Jeong H, et al. Tryptophan Hydroxylase 1 and 5-Ht(7) Receptor Preferentially Expressed in Triple-Negative Breast Cancer Promote Cancer Progression Through Autocrine Serotonin Signaling. *Mol Cancer* (2016) 15(1):75. doi: 10.1186/s12943-016-0559-6
89. Sarrouilhe D, Clarhaut J, Defamie N, Mesnil M. Serotonin and Cancer: What Is the Link? *Curr Mol Med* (2015) 15(1):62–77. doi: 10.2174/1566524015666150114113411

90. Fatima S, Shi X, Lin Z, Chen GQ, Pan XH, Wu JC, et al. 5-Hydroxytryptamine Promotes Hepatocellular Carcinoma Proliferation by Influencing B-Catenin. *Mol Oncol* (2016) 10(2):195–212. doi: 10.1016/j.molonc.2015.09.008
91. Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of Gwas Discovery. *Am J Hum Genet* (2012) 90(1):7–24. doi: 10.1016/j.ajhg.2011.11.029
92. Maher B. Personal Genomes: The Case of the Missing Heritability. *Nature* (2008) 456(7218):18–21. doi: 10.1038/456018a
93. Raychaudhuri S. Mapping Rare and Common Causal Alleles for Complex Human Diseases. *Cell* (2011) 147(1):57–69. doi: 10.1016/j.cell.2011.09.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 June 2022
doi: 10.3389/fonc.2022.902643

Mortality Rate of Lymphoma in China, 2013–2020

Weiping Liu[1], Jinlei Qi[2], Jiangmei Liu[2], Yuqin Song[1], Lijun Wang[2], Maigeng Zhou[2], Jun Ma[3]* and Jun Zhu[1]**

[1] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Lymphoma, Peking University Cancer Hospital and Institute, Beijing, China, [2] National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, [3] Department of Hematology & Oncology, Harbin Institute of Hematology & Oncology, Harbin, China

Lymphoma is a malignant disease that threatens human health and imposes a significant burden on the society burden; however, there are limited accurate mortality data on lymphoma in China. The present study aimed to analyse lymphoma-associated mortality at the national and provincial levels in mainland China. Mortality data of lymphoma was extracted from the disease surveillance system of the Chinese Center for Disease Control and Prevention. Mortality was represented by the number of deaths, crude mortality rate, and age-standardized mortality rate. Temporal trends in mortality rates were examined using the fitting joinpoint models. Lymphoma accounted for 31,225 deaths in 2020, of which 1,838 and 29,387 were due to Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL), respectively. The age-standardized mortality rate per 100,000 population was 1.76 for lymphoma, 0.10 for HL, and 1.66 for NHL. The mortality rate increased with age, reaching a peak in the age group of 80–84 years for HL and over 85 years for NHL. Moreover, the death risk due to lymphoma was approximately 1.5–2 times greater in males than in females in all age groups. The mortality rate was higher in eastern China than in central and western China, indicating a heterogeneous distribution at the provincial level. During 2013–2020, the mortality rate of lymphoma decreased by 1.85% (–22.94% for HL and –0.14% for NHL). In conclusion, the mortality of lymphoma varied by sex, age, and regions, which highlighted the need of establish differentiated strategy for disease control and prevention.

Keywords: lymphoma, Hodgkin disease, non-Hodgkin lymphoma, epidemiology, mortality

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Wenbin Liu, Second Military Medical University, China
Xiaoxue Liu, Wuhan University, China

****Correspondence:***
Maigeng Zhou
zhoumaigeng@ncncd.chinacdc.cn
Jun Ma
majun0322@126.com
Jun Zhu
zhu-jun2017@outlook.com

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *23 March 2022*
Accepted: *17 May 2022*
Published: *07 June 2022*

Citation:
Liu W, Qi J, Liu J, Song Y, Wang L, Zhou M, Ma J and Zhu J (2022) Mortality Rate of Lymphoma in China, 2013–2020. Front. Oncol. 12:902643. doi: 10.3389/fonc.2022.902643

INTRODUCTION

Lymphoma is a malignant disease that threatens human health. A systematic analysis based on the GLOBOCAN 2020 study from the International Agency for Research on Cancer revealed 83,087 incident cases and 23,376 deaths due to Hodgkin lymphoma (HL) and 544,352 incident cases and 259,793 deaths due to non-Hodgkin lymphoma (NHL), which accounted for 3.2% of all cancer cases and 2.8% of all cancer deaths worldwide (1). Compared with the results from the GLOBOCAN 2018 study, the incident cases increased by 3.9% and deaths decreased by 10.7% for HL, whereas the incident case and deaths increased by 6.8% and 4.5% for NHL, respectively (2).

China has a lower burden of lymphoma than the western countries. For example, there were an estimated 136,960 new cases of lymphoid malignancy with an age-standardized incidence rate of 34.4 per 100,000 population in the United States in 2016 (3). The new cases and age-standardized incidence rate per 100,000 population were 8,500 and 2.7 for HL, and 125,850 and 31.1 for NHL, respectively (3). During the same period, there were an estimated 6,900 and 68,500 incident cases with an age-standardized incidence rate of 0.46 and 4.29 per 100,000 population for HL and NHL in China (4). Notably, the mortality rate of lymphoma and myeloma showed a significant upward trend with an annual increase of 4.5% from 2004 to 2016 (5). In 2017, the incidence and mortality rates of NHL ranked 14th 12th, while the incidence and mortality rates of HL ranked 31st among all cancers in China (6).

Due to the low incidence rate of HL, NHL, and multiple myeloma, these three diseases are commonly grouped into the same classification in China. For example, the National Central Cancer Registry of China estimated that there were 52,100 deaths due to lymphoma and myeloma in 2015, but the accurate mortality data of lymphoma alone were not determined (7). In the current study, we conducted a comprehensive analysis of lymphoma-associated mortality at the national and provincial levels in mainland China.

METHODS

The study was approved by the Ethics Committee of the Chinese Center for Disease Control and Prevention (Beijing, China).

Data Sources

Mortality data were collected from the Chinese Center for Disease Control and Prevention–Disease Surveillance Points (CDC–DSP) system. This system consists of 605 surveillance points and covers a population of 323.8 million (24.3% of the total population of the country) across 31 provinces.

National age-specific population data were obtained from the National Bureau of Statistics of China (http://data.stats.gov.cn). The 2010 census population data of China were used to determine the age-standardized mortality rates by Chinese standard population (ASMRC). The Segi's population was used to calculate the age-standardized mortality rate worldwide (ASMRW) (8).

Data Collection

The daily death records from 1 January 2013 to 31 December 2020 were collected from the CDC-DSP system. International Classification of Diseases- 10 codes were used to identify HL (C81–C81.99) and NHL (C82–C86.6, C96–C96.9).

Quality Control

The quality control procedures for CDC-DSP system includes annual training of standard work flow, random checking of the accuracy of disease classification and duplication, which is done at the county, province, and national levels. The underreporting rate was evaluated by retrospective survey every 3 years, which was 9.4% in the recent period of 2015–2017.

Classification

The geographic unit includes 31 provinces, municipalities and autonomous regions in mainland China, which was referred to as provinces in the present study. The geographic area was divided into eastern China (including Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan), central China (including Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan), western China (including Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia and Xinjiang) (9). Urban/rural classification was made according to administrative characteristics (county as rural and district in cities as urban) (10).

Statistical Analysis

The mortality rates were calculated by the method of the following formula: estimated mortality rates = reported mortality rates/(1 - underreporting rates). The estimated deaths due to lymphoma were generated by the sum of the products of the age-specific mortality rates and the corresponding population in each stratum. Temporal trends in mortality rates from 2013 to 2020 were examined by fitting joinpoint models (version 4.6.0.0; National Cancer Institute). Changes were represented by the average annual percent change (AAPC) and their corresponding 95% confidence interval (CI) over the entire period. The term "increase" or "decrease" was used to describe the trends when the slope was statistically significant. For nonstatistically significant trends, the term "stable" was used.

The changes in the number of deaths between 2013 and 2020 was attributed to population growth, population structure, and age-specific mortality rate. The decomposition analysis used two counterfactual scenarios to calculate the number of deaths. The first scenario assumed that the total population grew but the population structure and age-specific mortality rate remained unchanged from 2013 to 2020. The difference between the number of deaths observed in 2013 and the first scenario was the change in the number of deaths exclusively attributable to population growth. The second scenario assumed that the total population grew and the population structure changed, but the age-specific mortality rate remained unchanged from 2013 to 2020. The difference between the first and the second scenario was the change in the number of deaths exclusively attributable to population ageing. The difference between the second scenario and the number of deaths observed in 2020 was the change in the number of deaths exclusively attributable to age-specific mortality rate.

RESULTS

Expected Deaths and the Mortality Rate of Lymphoma in 2020

There were an estimated 31,225 deaths due to lymphoma, with a crude mortality rate of 2.26 per 100,000 population. The ASMRC and ASMRW per 100,000 population were 1.76 and 1.35,

respectively (**Table 1**). For HL, the number of deaths was 1,838, with crude mortality rate, ASMRC and ASMRW of 0.13, 0.10 and 0.08 per 100,000 population, respectively. For NHL, the number of deaths was 29,387, with crude mortality rate, ASMRC and ASMRW of 2.13, 1.66 and 1.27 per 100,000 population, respectively.

Mortality Rates of Lymphoma Stratified by Age and Sex in 2020

In total, the age-specific mortality rate of lymphoma increased with age and reached a peak (18.04 per 100,000 population) in the age group of over 85 years (**Figure 1A**; **Table 2**). An upward trend in mortality rate with age was observed in both HL and NHL. The peak mortality rate was observed in the age group of 80–84 years for HL and over 85 years for NHL.

The risk of death due to lymphoma was approximately 1.5–2 times greater in males than in females in all age groups. For HL, the age-specific mortality rate was less than one per 100,000 population in those younger than 75 years in males, and in all age groups in females (**Figure 1B**). For NHL, the age-specific mortality rate gradually increased and reached a maximum in males over 85 years (24.68 per 100,000 population) and females aged 80–84 years (13.44 per 100,000 population, **Figure 1C**).

Mortality Rates of Lymphoma Stratified by Regions and Provinces in 2020

In total, a heterogeneous distribution of mortality rates was observed (**Appendix Figures S1**, **S2**). Eastern China had a higher crude mortality rate and ASMRC than central and western China (**Appendix Table S1**; **Figures 2A, B**).

At the provincial level, the crude mortality rate was highest in Hubei (4.39 per 100,000 population), Tianjin (3.86 per 100,000 population), and Liaoning (3.53 per 100,000 population), while was lowest in Ningxia (0.68 per 100,000 population), Tibet (1.14 per 100,000 population), and Qinghai (1.29 per 100,000 population) (**Figure 2A**). In contrast, the highest ASMRC was observed in Hubei (3.21 per 100,000 population), Fujian (3.17 per 100,000 population), and Tianjin (2.77 per 100,000 population), while the lowest ASMRC was seen in Ningxia (0.69 per 100,000 population), Hebei (1.04 per 100,000 population), and Jilin (1.15 per 100,000 population, **Figure 2B**).

Trends in Mortality of Lymphoma From 2013 to 2020

The mortality rate decreased by 1.85% during 2013–2020 (−22.94% for HL and −0.14% for NHL, **Table 3**). The change was attributed to three factors: population growth (5.13%), population aging (21.57%), and age-specific mortality rate (−28.55%, **Table 4**).

TABLE 1 | Mortality rate of lymphoma by sex and residence in China, 2020.

	Sex	Crude rate($1/10^5$)	ASMRC($1/10^5$)	ASMRW($1/10^5$)
Lymphoma				
All	Both	2.26	1.76	1.35
	Male	2.73	2.26	1.73
	Female	1.78	1.30	0.98
Urban	Both	2.20	1.72	1.30
	Male	2.63	2.21	1.68
	Female	1.76	1.27	0.95
Rural	Both	2.30	1.79	1.38
	Male	2.79	2.30	1.77
	Female	1.79	1.31	1.00
Hodgkin lymphoma				
All	Both	0.13	0.10	0.08
	Male	0.16	0.14	0.10
	Female	0.10	0.08	0.06
Urban	Both	0.12	0.09	0.07
	Male	0.14	0.12	0.09
	Female	0.09	0.07	0.06
Rural	Both	0.14	0.11	0.09
	Male	0.18	0.15	0.11
	Female	0.11	0.08	0.06
Non-Hodgkin lymphoma				
All	Both	2.13	1.66	1.27
	Male	2.56	2.12	1.63
	Female	1.68	1.22	0.92
Urban	Both	2.09	1.62	1.23
	Male	2.49	2.08	1.59
	Female	1.67	1.20	0.90
Rural	Both	2.16	1.68	1.29
	Male	2.61	2.15	1.66
	Female	1.69	1.24	0.94

ASMRC, age-standardized mortality rate adjusted by the Chinese standard population; ASMRW, age-standardized mortality rate adjusted by the world standard population.

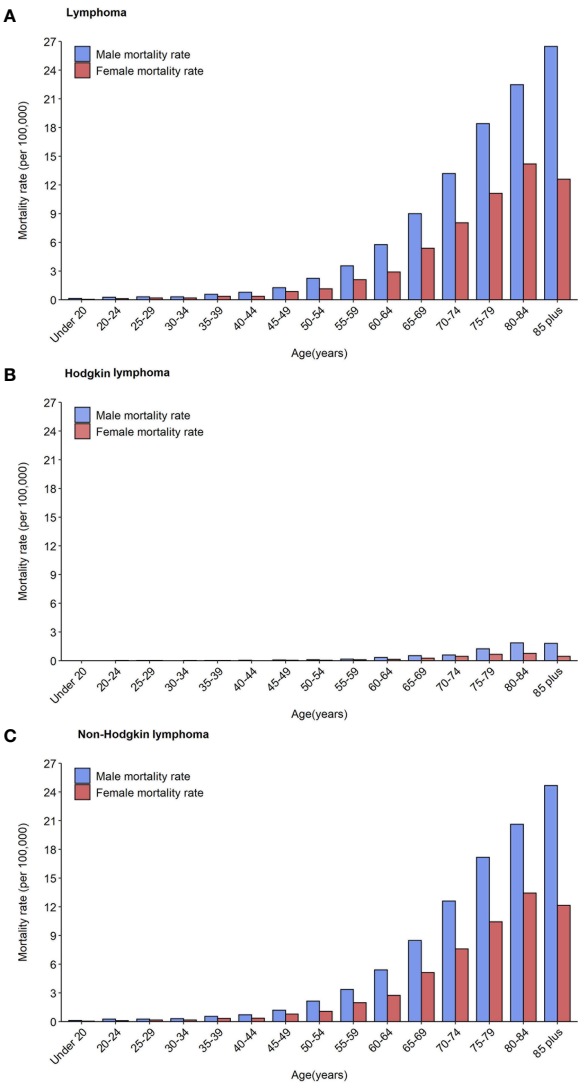


FIGURE 1 | Mortality rates of lymphoma **(A)**, Hodgkin lymphoma **(B)** and non-Hodgkin lymphoma **(C)** by age groups and sex in 2020.

The ASMRC per 100,000 population decreased from 2.30 in 2013 to 1.76 in 2020, which resulted in an AAPC of −3.6% (95% CI: −5.6% to −1.5%). In terms of residence variation, the ASMRC showed a decrease of 0.66 per 100,000 population with an AAPC of −4.7% in urban areas, and a decrease of 0.47 per 100,000 population with an AAPC of −2.5% in rural areas (**Appendix Table S2**, **Figure 3A**). All regions including eastern, central and western China showed a downward trend in the ASMRC (**Appendix Table S3**; **Figure 3B**). Moreover, the ASMRC of both HL and NHL in all areas showed a significant downward trend (**Table 3**).

DISCUSSION

The present study is the most comprehensive evaluation of lymphoma mortality based on the CDC-DSP system. The 605 surveillance points in this system were selected using an iterative method of multistage stratification and had a good national and regional representativeness at both national and provincial levels (10). We determined the temporal downward trend in lymphoma mortality during the past decade. Moreover, we explored the demographic and geographical differences in lymphoma mortality.

Older individuals often have a higher cancer burden. For example, the cancer burden was lowest in those aged 35–39 years and highest in those aged 80–84 years in China (11). Similarly, the probability of developing NHL in those older than 70 years was 6–7 fold higher than that in those younger than 60 years, which was 3–4 fold higher than that in the age group of 60–69 years in the United States (12). Moreover, NHL contributed to 4% of cancer deaths in males and 5% of cancer deaths in females aged over 85 years in 2019 (13). Consistent with previous studies,

TABLE 2 | Mortality rate of lymphoma stratified by age, sex, region and residence, 2020.

Age groups	Gender			Residence		Region		
	Both	Male	Female	Urban	Rural	Eastern	Central	Western
Lymphoma								
0~	0.10	0.15	0.05	0.10	0.11	0.09	0.10	0.13
20~	0.21	0.28	0.14	0.22	0.21	0.27	0.12	0.22
25~	0.26	0.32	0.20	0.15	0.37	0.23	0.28	0.30
30~	0.25	0.31	0.19	0.16	0.34	0.22	0.26	0.32
35~	0.47	0.58	0.36	0.42	0.52	0.49	0.36	0.55
40~	0.59	0.79	0.37	0.47	0.68	0.64	0.63	0.44
45~	1.08	1.28	0.87	0.86	1.24	1.09	1.04	1.10
50~	1.71	2.26	1.14	1.19	2.08	1.59	1.88	1.70
55~	2.83	3.55	2.10	2.48	3.09	2.95	2.85	2.55
60~	4.35	5.78	2.90	4.00	4.58	4.80	4.47	3.36
65~	7.17	9.02	5.40	7.02	7.26	7.72	7.43	5.78
70~	10.55	13.20	8.06	10.33	10.70	12.38	9.79	8.18
75~	14.54	18.42	11.13	15.85	13.60	16.78	14.14	10.81
80~	17.81	22.47	14.20	21.01	15.47	19.32	16.60	15.95
85~	18.04	26.49	12.61	23.75	13.85	22.14	13.95	14.04
Hodgkin lymphoma								
0~	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.02
20~	0.02	0.01	0.02	0.04	0.00	0.03	0.00	0.02
25~	0.04	0.04	0.03	0.02	0.05	0.03	0.03	0.04
30~	0.01	0.00	0.03	0.00	0.03	0.01	0.03	0.00
35~	0.03	0.03	0.03	0.04	0.02	0.02	0.05	0.02
40~	0.03	0.06	0.00	0.02	0.04	0.03	0.03	0.03
45~	0.08	0.09	0.07	0.05	0.09	0.06	0.09	0.09
50~	0.09	0.10	0.07	0.05	0.12	0.06	0.12	0.10
55~	0.15	0.19	0.11	0.09	0.20	0.12	0.21	0.13
60~	0.25	0.35	0.15	0.25	0.25	0.24	0.25	0.27
65~	0.39	0.53	0.26	0.36	0.42	0.36	0.43	0.42
70~	0.53	0.61	0.46	0.55	0.52	0.57	0.57	0.41
75~	0.95	1.25	0.68	0.97	0.93	0.95	1.01	0.86
80~	1.24	1.86	0.77	1.31	1.20	1.16	1.72	0.80
85~	0.99	1.81	0.47	0.84	1.10	1.26	0.90	0.51
Non-Hodgkin lymphoma								
0~	0.10	0.14	0.05	0.10	0.10	0.09	0.09	0.12
20~	0.20	0.27	0.12	0.18	0.21	0.25	0.12	0.20
25~	0.23	0.28	0.17	0.13	0.32	0.20	0.25	0.26
30~	0.24	0.31	0.17	0.16	0.31	0.21	0.23	0.32
35~	0.44	0.55	0.33	0.38	0.50	0.47	0.31	0.53
40~	0.55	0.73	0.37	0.45	0.64	0.61	0.60	0.41
45~	1.00	1.20	0.80	0.81	1.14	1.03	0.96	1.01
50~	1.62	2.15	1.07	1.15	1.96	1.54	1.76	1.60
55~	2.68	3.37	1.99	2.39	2.90	2.83	2.64	2.42
60~	4.09	5.42	2.75	3.75	4.33	4.55	4.21	3.09
65~	6.77	8.49	5.13	6.67	6.84	7.36	7.01	5.36
70~	10.02	12.59	7.60	9.78	10.18	11.81	9.21	7.77
75~	13.59	17.17	10.45	14.88	12.67	15.83	13.13	9.95
80~	16.57	20.62	13.44	19.71	14.27	18.15	14.88	15.15
85~	17.04	24.68	12.14	22.91	12.74	20.88	13.05	13.52

the mortality of lymphoma showed an upward trend with age and reached a peak in the age group of over 85 years in the present study. Notably, the mortality rate was very low (less than one per 100,000 population) in the age groups of < 45 years and was very high (more than 10 per 100,000 population) in the age groups of > 70 years. This phenomenon may be partly explained by a higher age-specific incidence rate, poor chemotherapy tolerance, complications due to therapy, and lower survival rates in the elderly (14–17). These findings highlight the need to develop differentiated disease control and prevention strategies for different age-specific populations.

The level of economic development played an important role in the heterogeneous geographical distribution of the cancer burden. A cohort study involving 497,693 participants aged between 35 and 74 years showed that the standardized mortality rates were higher in rural areas (241.2 per 100,000 person-years) than in urban areas (183.5 per 100,000 person-years) (18). Moreover, the cancer mortality rate was 44% higher

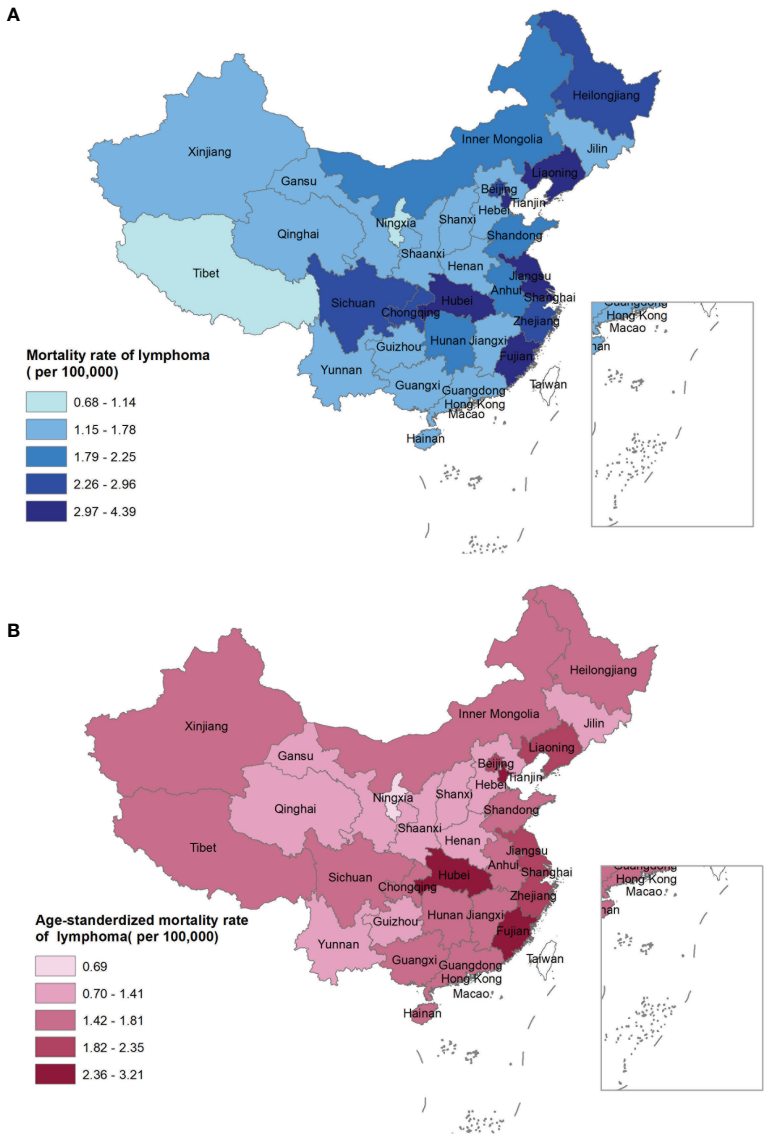


FIGURE 2 | Crude mortality rate **(A)** and age-standardized mortality rate of China **(B)** of lymphoma for both sexes by provinces in 2020, China.

in rural men aged 30–34 years and 44% higher in rural women aged 15–19 years than in their urban counterparts (19). In the present study, rural areas had a higher mortality rate of lymphoma than the urban areas, especially with a 22% higher mortality rate of HL. However, the incidence rates of lymphoid neoplasms in rural areas were lower than those in urban areas (3.4 vs 4.7 per 100,000 population) in China (20). This urban–rural discordance may be partly explained by an increase in deaths due to insufficient access to health services and poor survival in rural areas (21–24). Therefore, these findings support the establishment of an official medical system to reduce the burden of lymphoma, especially in rural areas.

Importantly, there was a significant decrease in lymphoma mortality during 2013–2020. In particular, despite the increased incidence due to population growth and aging, the HL mortality rate decreased by about a quarter, due to a 50% reduction in the age-specific mortality rates. This was associated with a better prognosis due to advances in anti-lymphoma treatment (25, 26). For example, a study involving 3,760 lymphoma patients showed that the 5- and 10-year overall survival rates for classic HL were 80% and 71%, respectively, and the 5-year overall survival rate for classic HL increased by 25% during the past two decades (55.4% in 1996–2000 vs 79.0% in 2010–2015) (27). Similarly, immunochemotherapy with rituximab improved the prognosis of B-cell lymphoma. For example, the introduction of rituximab into therapy for diffuse large B-cell lymphoma led to better survival outcomes compared to chemotherapy alone (28–30). However, the 5-year relative survival rate of lymphoid

TABLE 3 | Mortality rate and average annual percentage change of lymphoma by sex in China, 2013-2020.

	Mortality ratein 2013	Mortality ratein 2020	AAPC(95% CI, %)	*P* value
Lymphoma				
Male				
Crude rate ($1/10^5$)	3.05	2.73	-0.6 (-2.5 to 1.3)	0.458
ASMRC ($1/10^5$)	3.03	2.26	-3.1 (-5.1 to -1.1)	0.009
ASMRW ($1/10^5$)	2.36	1.73	-4.0 (-6.7 to -1.2)	0.005
Female				
Crude rate ($1/10^5$)	1.77	1.78	0.3 (-1.8 to 2.5)	0.737
ASMRC ($1/10^5$)	1.61	1.30	-3.3 (-6.3 to -0.3)	0.034
ASMRW ($1/10^5$)	1.27	0.98	-3.9 (-6.8 to -0.8)	0.014
Hodgkin lymphoma				
Male				
Crude rate ($1/10^5$)	0.23	0.16	-4.7 (-7.4 to -1.8)	0.007
ASMRC ($1/10^5$)	0.23	0.14	-7.1 (-9.6 to -4.5)	0.001
ASMRW ($1/10^5$)	0.18	0.10	-7.6 (-9.9 to -5.2)	< 0.001
Female				
Crude rate ($1/10^5$)	0.13	0.10	-4.0 (-7.2 to -0.8)	0.023
ASMRC ($1/10^5$)	0.12	0.08	-6.7 (-9.6 to -3.7)	0.002
ASMRW ($1/10^5$)	0.10	0.06	-6.9 (-9.4 to -4.3)	0.001
Non-Hodgkin lymphoma				
Male				
Crude rate ($1/10^5$)	2.82	2.56	-0.3 (-2.3 to 1.7)	0.724
ASMRC ($1/10^5$)	2.80	2.12	-3.4 (-6.3 to -0.4)	0.026
ASMRW ($1/10^5$)	2.18	1.63	-3.7 (-6.6 to -0.7)	0.015
Female				
Crude rate ($1/10^5$)	1.64	1.68	0.6 (-1.6 to 2.9)	0.517
ASMRC ($1/10^5$)	1.49	1.22	-2.6 (-5.9 to 0.8)	0.131
ASMRW ($1/10^5$)	1.17	0.92	-3.6 (-6.6 to -0.6)	0.019

AAPC, average annual percentage change; CI, confidence interval

malignancies was only 38.3% in China, which was markedly lower than that in the western countries (70% or higher) (31). Careful attention should be paid to the increase in the mortality rate of both HL and NHL due to the population in China. Therefore, further studies should focus on evaluating the impact of the promotion of standardized diagnosis and treatment procedures on lymphoma burden in the absence of prevention measures.

This study has several limitations. First, the data were extracted according to the International Classification of Diseases- 10 codes from the CDC-DSP database, the mortality rates of lymphoma subtypes were not assessed according to the lymphoid tumor classifications of the World Health Organization. Second, the mortality rate of HL was very low (< 0.1 per 100,000 population) in the age groups less than 55 years at the national level, which may lead to underestimation or overestimation of the results at the provincial level. Third, socioeconomic factors such as sociodemographic index and human development index, were not used to evaluate the attribution to the change in disease burden.

In conclusion, the present study determined the spatiotemporal characteristics of lymphoma mortality using nationally representative data from China. The mortality rate was higher in males and older individuals. Moreover, rural areas had higher mortality rates than urban areas. An encouraging downward trend was observed, especially in HL mortality. Moreover, the present

TABLE 4 | Decomposition of changes in lymphoma deaths from 2013 to 2020.

	Lymphoma	Hodgkin lymphoma	Non-Hodgkin lymphoma
Both			
Observed number of people in 2013	7,813*	586*	7,227*
Number expected with 2020 population, 2013 population age structure, and 2013 deaths	8,213	616	7,597
Number expected with 2020 population, 2020 population age structure, and 2013 deaths	9,898	746	9,151
Observed number of people in 2020	7,667*	451*	7,216*
Percentage change from 2013 due to population growth	5.13	5.13	5.13
Percentage change from 2013 due to population ageing	21.57	22.32	21.51
Percentage change from 2013 due to change in age-specific mortality rate	-28.55	-50.39	-26.78
Observed percentage change from 2013 to 2020	-1.85	-22.94	-0.14
Male			
Observed number of people in 2013	5,032*	382*	4,650*
Number expected with 2020 population, 2013 population age structure, and 2013 deaths	5,271	400	4,871
Number expected with 2020 population, 2020 population age structure, and 2013 deaths	6,297	483	5,813
Observed number of people in 2020	4,713*	285*	4,428*
Percentage change from 2013 due to population growth	4.73	4.73	4.73
Percentage change from 2013 due to population ageing	20.40	21.87	20.28
Percentage change from 2013 due to change in age-specific mortality rate	-31.49	-52.08	-29.80
Observed percentage change from 2013 to 2020	-6.36	-25.48	-4.79
Female			
Observed number of people in 2013	2,780*	204*	2,576*
Number expected with 2020 population, 2013 population age structure, and 2013 deaths	2,933	215	2,718
Number expected with 2020 population, 2020 population age structure, and 2013 deaths	3,567	260	3,308
Observed number of people in 2020	2,955*	167*	2,788*
Percentage change from 2013 due to population growth	5.54	5.54	5.54
Percentage change from 2013 due to population ageing	22.81	21.98	22.88
Percentage change from 2013 due to change in age-specific mortality rate	-22.03	-45.65	-20.16
Observed percentage change from 2013 to 2020	6.33	-18.13	8.26

**Death number was based on the disease surveillance points system of Chinese Center for Disease Control and Prevention.*

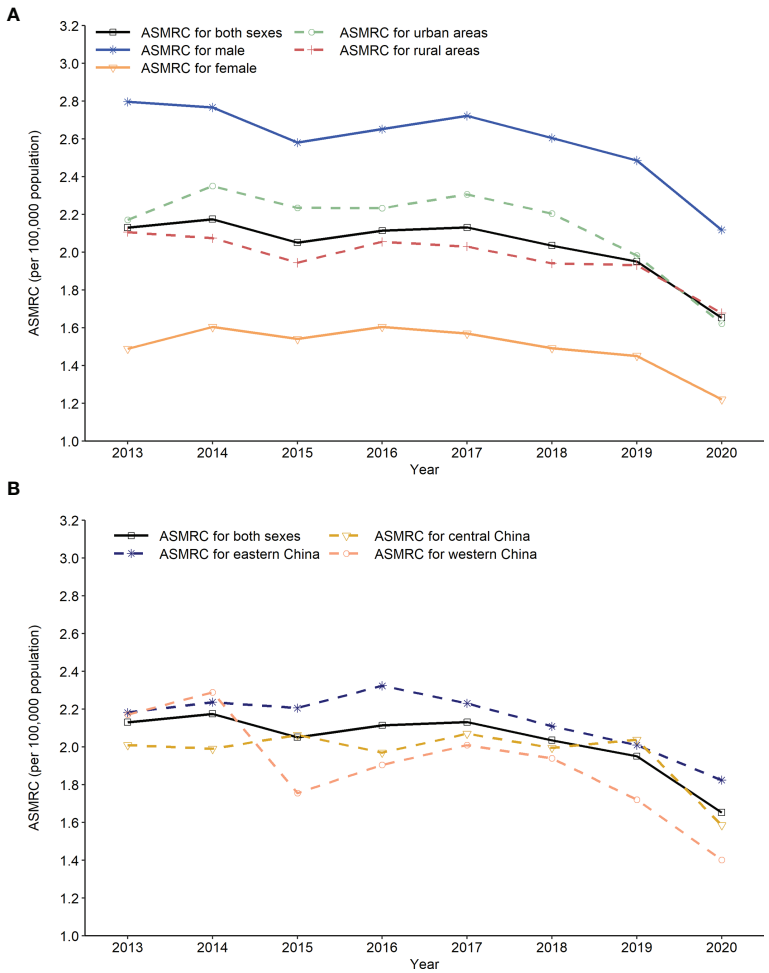


FIGURE 3 | Trends in age-standardized mortality rate of China (ASMRC) of lymphoma by sex, residence and region in China, 2013 to 2020. **(A)** Trends in ASMRC of lymphoma by gender and residence, **(B)** Trends in ASMRC of lymphoma by region.

study provided detailed information on the mortality rate of lymphoma at the national and provincial levels. These results may assist in establishing stratified strategies when policies for disease prevention and management are implemented.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

WL conceived and designed the study, analyzed the data, and drafted and revised the paper. JQ and JL prepared and analyzed the data. YS and LW drafted and revised the manuscript. MZ, JM, and JZ designed the study, interpreted the results, and drafted and revised the paper. All the authors have provided critical comments on the manuscript.

FUNDING

This study was supported by the Clinical Research Fund for Distinguished Young Scholars of Beijing Cancer Hospital (grant no. QNJJ202106). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

ACKNOWLEDGMENTS

The authors thank all of the persons who contributed to the data collection, analysis, vetting, and critical interpretation of the present study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.902643/full#supplementary-material

Supplementary Figure 1 | Mortality rate of lymphoma in males by provinces in 2020, China**. (A)** Crude mortality rate; **(B)** age-standardized mortality rate of China.

Supplementary Figure 2 | Mortality rate of lymphoma in females by provinces in 2020, China**. (A)** Crude mortality rate; **(B)** age-standardized mortality rate of China.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
3. Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, Flowers CR, et al. 2016 US Lymphoid Malignancy Statistics by World Health Organization Subtypes. *CA Cancer J Clin* (2016) 66:443–59. doi: 10.3322/caac.21357
4. Liu W, Liu J, Song Y, Zeng X, Wang X, Mi L, et al. Burden of Lymphoma in China, 2006-2016: An Analysis of the Global Burden of Disease Study 2016. *J Hematol Oncol* (2019) 12:115. doi: 10.1186/s13045-019-0785-7
5. Liu W, Liu J, Song Y, Zeng X, Wang X, Mi L, et al. Mortality of Lymphoma and Myeloma in China, 2004-2017: An Observational Study. *J Hematol Oncol* (2019) 12:22. doi: 10.1186/s13045-019-0706-9
6. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* (2019) 5:1749–68. doi: 10.1001/jamaoncol.2019.2996
7. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338
8. Segi M. *Cancer Mortality for Selected Sites in 24 Countries (1950–57).* Sendai, Japan: Department of Public Health, Tohoku University of Medicine (1960).
9. National Bureau of Statistics of China. *Communiqué on the Fourth National Economic Census (No. 7)* . Available at: http://www.stats.gov.cn/english/PressRelease/201911/t20191120_1710332.html.
10. Liu S, Wu X, Lopez AD, Wang L, Cai Y, Page A, et al. An Integrated National Mortality Surveillance System for Death Registration and Mortality Surveillance, China. *Bull World Health Organ.* (2016) 94(1):46–57. doi: 10.2471/BLT.15.153148
11. Cao M, Li H, Sun D, Chen W. Cancer Burden of Major Cancers in China: A Need for Sustainable Actions. *Cancer Commun (Lond).* (2020) 40:205–10. doi: 10.1002/cac2.12025, et al
12. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
13. DeSantis CE, Miller KD, Dale W, Mohile SG, Cohen HJ, Leach CR, et al. Cancer Statistics for Adults Aged 85 Years and Older, 2019. *CA Cancer J Clin* (2019) 69:452–67. doi: 10.3322/caac.21577
14. Hashim D, Carioli G, Malvezzi M, Bertuccio P, Waxman S, Negri E, et al. Cancer Mortality in the Oldest Old: A Global Overview. *Aging (Albany NY).* (2020) 12:16744–58. doi: 10.18632/aging.103503
15. Carioli G, Malvezzi M, Bertuccio P, Hashim D, Waxman S, Negri E, et al. Cancer Mortality in the Elderly in 11 Countries Worldwide, 1970-2015. *Ann Oncol* (2019) 30:1344–55. doi: 10.1093/annonc/mdz178
16. Smith A, Crouch S, Lax S, Li J, Painter D, Howell D, et al. Lymphoma Incidence, Survival and Prevalence 2004-2014: Sub-Type Analyses From the UK's Haematological Malignancy Research Network. *Br J Cancer* (2015) 112:1575–84. doi: 10.1038/bjc.2015.94
17. Blum KA, Keller FG, Castellino S, Phan A, Flowers CR. Incidence and Outcomes of Lymphoid Malignancies in Adolescent and Young Adult Patients in the United States. *Br J Haematol* (2018) 183:385–99. doi: 10.1111/bjh.15532
18. Pan R, Zhu M, Yu C, Lv J, Guo Y, Bian Z, et al. Cancer Incidence and Mortality: A Cohort Study in China, 2008-2013. *Int J Cancer.* (2017) 141:1315–23. doi: 10.1002/ijc.30825
19. Liu L. Rural-Urban Inequities in Deaths and Cancer Mortality Amid Rapid Economic and Environmental Changes in China. *Int J Public Health* (2019) 64:39–48. doi: 10.1007/s00038-018-1109-3
20. Chen W, Sun K, Zheng R, Zeng H, Zhang S, Xia C, et al. Cancer Incidence and Mortality in China, 2014. *Chin J Cancer Res* (2018) 30:1–12. doi: 10.21147/j.issn.1000-9604.2018.01.01
21. Huang C, Liu CJ, Pan XF, Liu X, Li NX. Correlates of Unequal Access to Preventive Care in China: A Multilevel Analysis of National Data From the 2011 China Health and Nutrition Survey. *BMC Health Serv Res* (2016) 16:177. doi: 10.1186/s12913-016-1426-2
22. Wang J, Chen L, Ye T, Zhang Z, Ma J. Financial Protection Effects of Modification of China's New Cooperative Medical Scheme on Rural Households With Chronic Diseases. *BMC Health Serv Res* (2014) 14:305. doi: 10.1186/1472-6963-14-305
23. Zhu D, Guo N, Wang J, Nicholas S, Chen L. Socioeconomic Inequalities of Outpatient and Inpatient Service Utilization in China: Personal and Regional Perspectives. *Int J Equity Health* (2017) 16:210. doi: 10.1186/s12939-017-0706-8
24. Zeng H, Zheng R, Guo Y, Zhang S, Zou X, Wang N, et al. Cancer Survival in China, 2003-2005: A Population-Based Study. *Int J Cancer.* (2015) 136:1921–30. doi: 10.1002/ijc.29227
25. Zhu J, Ma J. Chinese Society of Clinical Oncology (CSCO) Diagnosis and Treatment Guidelines for Malignant Lymphoma 2021 (English Version). *Chin J Cancer Res* (2021) 33(3):289–301. doi: 10.21147/j.issn.1000-9604.2021.03.01
26. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing Cancer Survival in China During 2003-15: A Pooled Analysis of 17 Population-Based Cancer Registries. *Lancet Glob Health* (2018) 6:e555–67. doi: 10.1016/S2214-109X(18)30127-X
27. Liu W, Ji X, Song Y, Wang X, Zheng W, Lin N, et al. Improving Survival of 3760 Patients With Lymphoma: Experience of an Academic Center Over Two Decades. *Cancer Med* (2020) 9:3765–74. doi: 10.1002/cam4.3037
28. Tsutsué S, Tobinai K, Yi J, Crawford B. Nationwide Claims Database Analysis of Treatment Patterns, Costs and Survival of Japanese Patients With Diffuse Large B-Cell Lymphoma. *PLoS One* (2020) 15:e0237509. doi: 10.1371/journal.pone.0237509
29. Shah BK, Bista A, Shafii B. Survival in Advanced Diffuse Large B-Cell Lymphoma in Pre- and Post-Rituximab Eras in the United States. *Anticancer Res* (2014) 34:5117–20. doi: 10.1245/s10434-014-3557-1
30. Khor S, Beca J, Krahn M, Hodgson D, Lee L, Crump M, et al. Real World Costs and Cost-Effectiveness of Rituximab for Diffuse Large B-Cell Lymphoma Patients: A Population-Based Analysis. *BMC Cancer* (2014) 14:586. doi: 10.1186/1471-2407-14-586
31. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global Surveillance of Trends in Cancer Survival 2000-14 (CONCORD-3): Analysis of Individual Records for 37 513 025 Patients Diagnosed With One of 18 Cancers From 322 Population-Based Registries in 71 Countries. *Lancet* (2018) 391:1023–75. doi: 10.1016/S0140-6736(17)33326-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

REVIEW
published: 07 June 2022
doi: 10.3389/fonc.2022.906125

Endoscopic Screening for Second Primary Tumors of the Esophagus Among Head and Neck Cancer Patients

Chen-Shuan Chung[1,2,3†], Li-Jen Liao[3,4,5†], Chia-Yun Wu[3,6], Wu-Chia Lo[3,4,7], Chen-Hsi Hsieh[3,8,9,10], Tzong-His Lee[1], Chao-Yu Liu[3,10,11], Deng-Yu Kuo[9] and Pei-Wei Shueng[3,9,10,12]*

[1] Division of Gastroenterology and Hepatology, Department of Internal Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan, [2] College of Medicine, Fu Jen Catholic University, New Taipei City, Taiwan, [3] Head and Neck Cancer Surveillance & Research Group, Far Eastern Memorial Hospital, New Taipei City, Taiwan, [4] Otolaryngology Head and Neck Surgery, Far Eastern Memorial Hospital, New Taipei City, Taiwan, [5] Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan, [6] Division of Medical Oncology, Department of Internal Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan, [7] Graduate Institute of Medicine, Yuan Ze University, Taoyuan, Taiwan, [8] Institute of Traditional Medicine, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, [9] Division of Radiation Oncology, Department of Radiology, Far Eastern Memorial Hospital, New Taipei City, Taiwan, [10] School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, [11] Division of Thoracic Surgery, Department of Surgery, Far Eastern Memorial Hospital, New Taipei City, Taiwan, [12] Medical Device Innovation and Translation Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Wei-Kuo Chang, Tri-Service General Hospital, Taiwan
Yu-Hsuan Kuo, Chi Mei Medical CenterTaiwan

***Correspondence:**
Pei-Wei Shueng
shuengsir@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *28 March 2022*
Accepted: *11 May 2022*
Published: *07 June 2022*

Citation:
Chung C-S, Liao L-J, Wu C-Y, Lo W-C, Hsieh C-H, Lee T-H, Liu C-Y, Kuo D-Y and Shueng P-W (2022) Endoscopic Screening for Second Primary Tumors of the Esophagus Among Head and Neck Cancer Patients. Front. Oncol. 12:906125. doi: 10.3389/fonc.2022.906125

Malignancies of the head and neck (HN) region and esophagus are among the most common cancers worldwide. Due to exposure to common carcinogens and the theory of field cancerization, HN cancer patients have a high risk of developing second primary tumors (SPTs). In our review of 28 studies with 51,454 HN cancer patients, the prevalence of SPTs was 12%. The HN area is the most common site of SPTs, followed by the lungs and esophagus, and 13% of HN cancer patients have been reported to have esophageal high-grade dysplasia or invasive carcinoma. The prognosis of HN cancer patients with concomitant esophageal SPTs is poor, and therefore identifying esophageal SPTs as early as possible is of paramount importance for risk stratification and to guide the treatment strategy. Image-enhanced endoscopy, especially using narrow-band imaging endoscopy and Lugol's chromoendoscopy, has been shown to improve the diagnostic performance in detecting esophageal neoplasms at an early stage. Moreover, the early detection and minimally invasive endoscopic treatment of early esophageal neoplasm has been shown to improve the prognosis. Well-designed prospective studies are warranted to establish appropriate treatment and surveillance programs for HN cancer patients with esophageal SPTs.

Keywords: head neck cancer, esophageal cancer, second primary tumor, cancer screening, image-enhanced endoscopy, narrow-band imaging, Lugol's chromoendoscopy

INTRODUCTION

Malignancies of the head and neck (HN) region and esophagus are among the most common cancers worldwide (1). In parallel with the advances in diagnostic modalities for cancer screening and surveillance, an increasing number of second primary tumors (SPTs) are being detected. SPTs may develop into any kind of malignancy, including malignancy of multicentric origins in the HN region, lungs and esophagus, particularly in HN cancer patients (2–5). This cancerization field known as the upper aerodigestive tract (UADT) is exposed to common carcinogens, particularly cigarette smoke, alcohol, and betel quid. The occurrence of SPTs in the UADT, either synchronously or metachronously, and single or multiple, in HN cancer patients is associated with worse survival despite appropriate management of the primary index HN tumor (2, 3, 6, 7). Of these SPTs, esophageal cancer is associated with a worse prognosis than other sites of the UADT (2, 3). Moreover, esophageal SPTs are easily overlooked as many are diagnosed at asymptomatic early stages (8–12). Therefore, the early identification of esophageal neoplasms and treatment of the primary index cancer and esophageal SPTs is of paramount importance to improve the overall outcomes of HN cancer patients. In this review, we describe the association between HN and esophageal cancers, and propose a screening strategy for esophageal SPTs among HN cancer patients.

DISEASE BURDEN OF HN CANCER AND ESOPHAGEAL CANCER

Head and neck cancers are the sixth and seventh most common cancers in Taiwan and worldwide, respectively (1, 13). Globally, HN cancer was the fifth most common cancer in men and the 12th most common cancer in women, accounting for an estimated 8,170 and 888,000 new cases in Taiwan and worldwide, respectively, in 2018 (1, 13). The incidence is higher in males, especially middle-aged males, with a male-to-female incidence ratio of 3:1, and most (about 70%) new cases occur in low- and middle-income countries (1). Regarding mortality from HN cancer, there were an estimated 3,027 and 453,000 deaths in Taiwan (the fifth leading cause of cancer deaths) and worldwide, respectively, in 2018 (1, 13). A Canadian study examined the 25-year survival outcomes of 1,657 patients, and reported 2, 5, 15 and 25-year HN cancer-specific survival rates of 74%, 63%, 53% and 49%, respectively (14). In addition, an Italian study of 801 cases reported a 5-year overall survival for HN cancer of 62%, including 55% for cancer of the oral cavity, 53% for the oropharynx, 41% for the hypopharynx, and 71% for the larynx (15). In Taiwan, the 5-year overall survival for HN cancer during the past decade ranged from 40~60%, and the standardized death growth rate in men was 7.7% (13).

Esophageal cancer is the eighth most common cancer (sixth in Taiwanese males) and the sixth most common cause of cancer deaths (ninth in Taiwan) worldwide (1, 13). Malignancy of the esophagus has two main histological subtypes, namely esophageal squamous cell carcinoma (ESCC), and esophageal adenocarcinoma (EAC). ESCC accounts for the majority (93.13% in Taiwan, 87% globally) of all esophageal cancer cases (1, 13). In 2012, there were an estimated 398,000 and 52,000 new cases of ESCC and EAC, respectively, worldwide (1). In Taiwan, 2,436 and 84 new cases of ESCC and EAC were reported in 2018 (13). The male-to-female incidence ratio is 2.7:1 for ESCC and 4.4:1 for EAC (1). Similar to HN cancer, about half (52.71%) of esophageal cancers develop in patients aged between 40~60 years in Taiwan (13). The overall prognosis of esophageal cancer is poor because most cases are diagnosed at a late stage with obstructive symptoms. Only 15.93% of esophageal cancer patients are diagnosed at stage 0/I, compared to 69.83% at stage III/IV in Taiwan (13). The overall 5-year survival rate for esophageal cancer is less than 10~20%, and lower than 5% in low- and middle-income countries (1, 13, 16). In Taiwan, the standardized death growth rate of esophageal cancer during the past decade was 15.5% (13).

The incidence rates of both HN and esophageal cancers are increasing and the prognosis is unsatisfactory, especially for esophageal cancer. Most cases occur in middle-aged males with a great impact on cancer-related morbidity and mortality. Consequently, early detection through screening programs for patients at high risk is crucial to improve their prognosis.

ASSOCIATION BETWEEN HN AND ESOPHAGEAL CANCERS

Common Risk Factors and the Epidemiology for HN Cancer and Esophageal SPTs

The risk factors for HN cancer include male sex, infectious agents [*human papillomaviruses (HPV)*, *Epstein–Barr virus*], exposure to carcinogens (tobacco or marijuana use, alcohol consumption, betel quid chewing), poor oral hygiene, history of esophageal cancer, drinking hot beverages such as maté, occupational exposure (metal smelting and textile production), and consumption of preserved foods with high nitrosamine content (1, 13, 17–19). In addition, genetic factors have also been associated with the development of HN cancer. Among non-*HPV*-related HN cancers, *TP53* and cyclin-dependent kinase inhibitor 2A (*CDKN2A*) are the most affected genes, while the genetic changes in *HPV*-related tumors are in the phosphoinositide 3-kinase (PI3K) pathway, particularly involving activating mutations and amplifications of the *PIK3CA* oncogene (1, 6). Alcohol-metabolizing enzyme gene polymorphisms have also been associated with a higher risk of HN cancer (19, 20). For ESCC, the risk factors are older age, male sex, low body mass index, lower socioeconomic status, exposure to carcinogens (alcohol consumption, cigarette smoking, and betel quid chewing), low fruit/vegetable consumption, high meat/high temperature beverage intake, family members with esophageal cancer, history of HN cancer, poor oral hygiene, genetic polymorphism of alcohol-dehydrogenase-1B (ADH1B)

and aldehyde dehydrogenase-2 (ALDH2), and motor disorders of the esophagus (e.g., achalasia) (7, 19, 21). For EAC, the most important risk factors are obesity, gastroesophageal reflux disease and Barrett's esophagus (7). Mutations of tumor suppressor genes, multiple allelic losses, hypermethylation of promoter genes, genetic overexpression, and changes in miRNA expression profile have also been reported in both EAC and ESCC (7).

There are many common risk factors for the development of HN cancer and ESCC. The squamous epithelium of both the HN region and esophagus are exposed to common environmental factors, particularly carcinogens. Consequently, with underlying genetic alterations such as polymorphisms in alcohol-metabolizing enzyme genes, those with accumulating exposure to carcinogens may develop both HN cancer and ESCC (**Figure 1**). Several epidemiology studies have demonstrated an increased risk of synchronous and metachronous SPTs among HN cancer patients. We used keywords including "head and neck" AND " esophageal cancer" AND " second tumor" AND "screening" for literature review on PubMed. Exclusion criteria were as followings: studies without data upon incidence of esophageal SPTs, review article, case reports and number of HN cancer patients less than one-hundred (**Figure 2**). In our review of 28 studies with 51,454 HN cancer patients, the estimated prevalence of SPTs was 12% (95% CI, 10-15% with a random effects model). The index primary cancer, sites of SPT, and screening modalities in these 51,454 HN cancer patients are shown in **Table 1** and **Figure 3** (3, 8, 11, 12, 17, 22–45, 47–51). One 10-year follow-up study of 6,258 HN cancer patients reported that 21.8% presented with SPTs, with the highest excess absolute risk (EAR) for SPTs of the lungs, followed by those located at the HN region and esophagus (52). Similar results were reported in a population-based cohort study of 64,673 HN cancer patients in the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) registry between 1979 and 2008, in which the standardized incidence ratio (SIR) of synchronous SPTs was 5.0, with the highest excess risk of a second cancer at the HN region (SIR, 41.4), followed by the esophagus (SIR, 21.8), and lungs (SIR, 7.4) (53). In addition, a meta-analysis reported an SIR for metachronous SPTs, which were defined as occurring six months after the primary index tumor, of 2.04 (95% CI, 1.61~2.59) (9). The highest risk for metachronous SPTs located at the HN region was for the oropharynx (SIR, 17.82; 95% CI, 6.79–46.77), followed by the hypopharynx (SIR, 9.17; 95% CI, 3.51–23.98) and larynx (SIR, 4.12; 95% CI, 2.87–5.90), while the highest risk for SPTs located outside the HN area was for the esophagus (SIR, 4.64; 95% CI, 3.12–6.89), followed by the salivary glands (SIR, 8.30; 95% CI, 2.37–29.09) and thyroid (SIR, 1.47; 95% CI, 1.22–1.76) (9). In a study that defined a metachronous SPT as occurring 2 months after the primary HN cancer, an increased risk for metachronous SPTs of the lungs (SIR, 4.32; 95% CI 2.15-8.68) was also noted (9). Another systematic review of 456,130 HN cancer patients from 61 articles with a minimum follow-up of 22 months reported a mean incidence of SPTs of 13.2% (95% CI, 11.56-14.84), including 5.3% for synchronous SPTs (95% CI, 4.24-6.36) and 9.4% for metachronous SPTs (95% CI, 7.9-10.9) (54). In addition, the most common site of SPTs was the HN area,

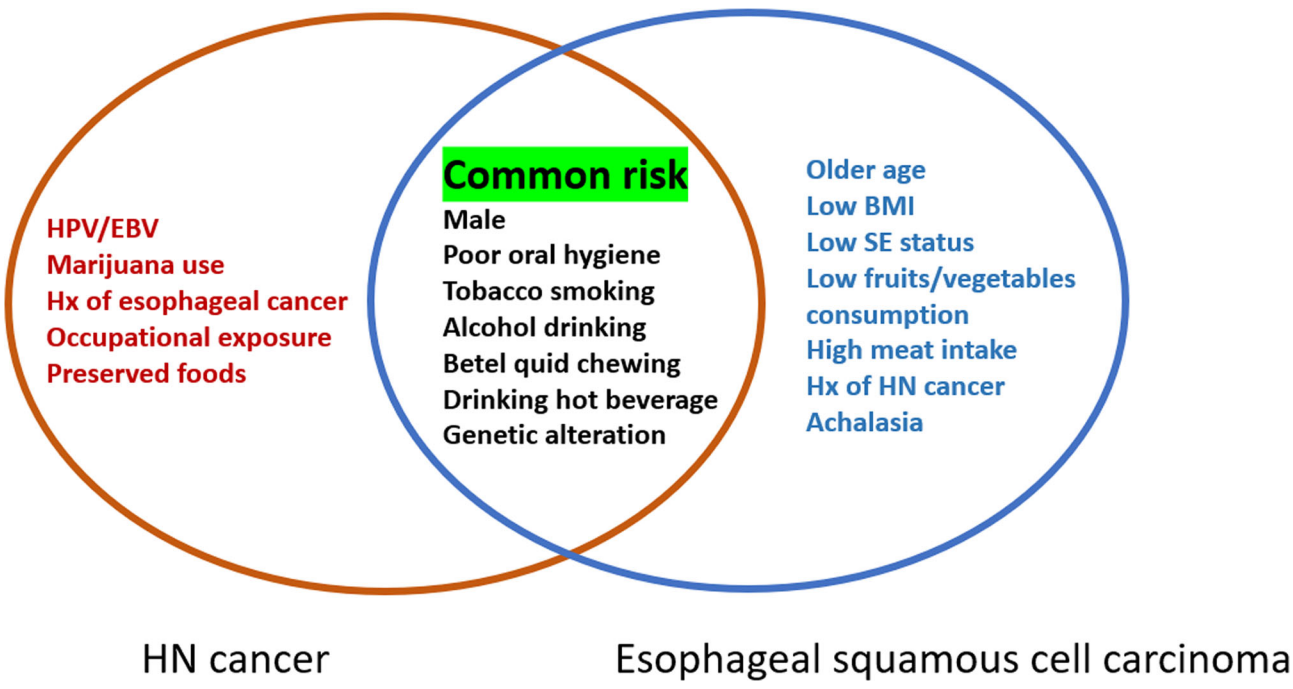


FIGURE 1 | Risk factors for head and neck cancer and esophageal squamous cell carcinoma. HPV, human papillomavirus; EBV, Epstein-Barr virus; Hx, history; BMI, body mass index; SE, socioeconomic.

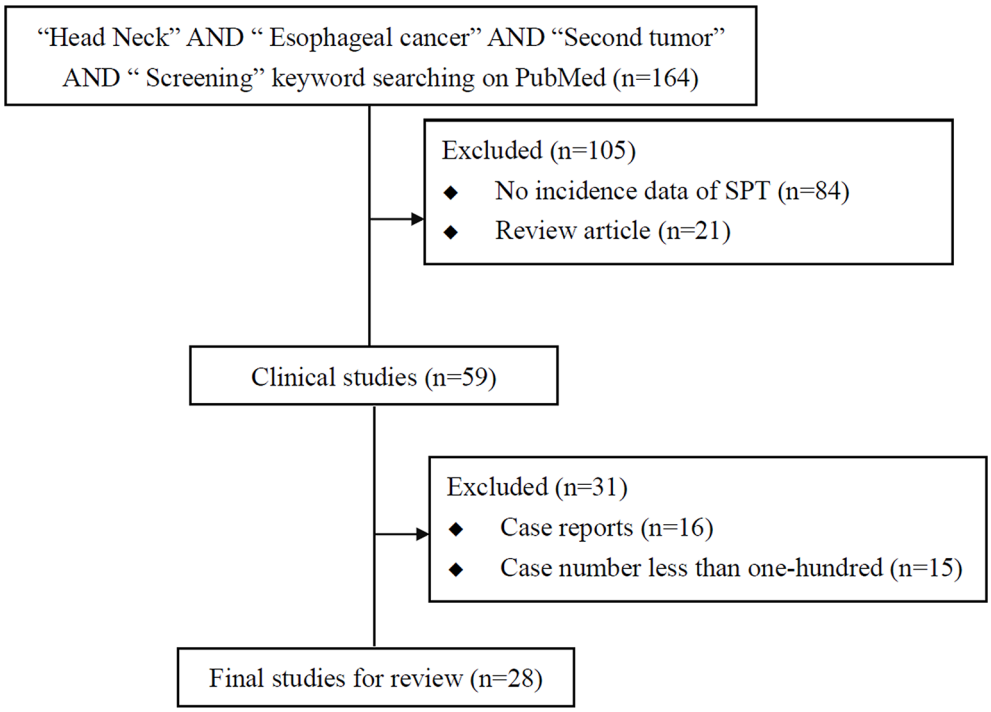


FIGURE 2 | Flowchart of literature review of studies on screening esophageal second primary tumor (SPT) in head and neck cancer patients.

TABLE 1 | Prevalence of SPT in HN cancer patients.

Author/Reference no.Year	No (%) of SPT/All/Index HN cancer	Esophagus,no (%)	Lung,no (%)	HN region,no (%)	Others,no (%)
Vrabec (22) **1979**	175 (11.5)/1,518/Oral cavity, pharynx or larynx	25 (14.3)	49 (28.0)	49 (28.0)	52 (29.7)
Wagenfeld (23) **1980**	48 (6.5)/740/Glottis	3 (6.3)	25 (52.1)	20 (41.7)	0 (0)
Tepperman (24) **1981**	101 (26.8)/377/Oral cavity	10 (9.9)	24 (23.8)	48 (47.5)	19 (18.8)
McDonald (25)	47 (20)/235/Larynx	0 (0)	22 (46.8)	9 (19.1)	16 (34.0)
Panosetti (17) **1989**	830 (9.1)/9,089/Oral cavity, pharynx, larynx	103 (12.4)	89 (10.7)	398 (47.9)	240 (28.9)
Larson (26) **1990**	207 (23.7)/875/Oral cavity, pharynx, larynx	13 (6.3)	54 (26.1)	129 (62.3)	11 (5.3)
Haughey (27) **1992**	528 (14.2)/3,706/Oral cavity, pharynx and larynx	17 (3.2)	106 (20.1)	246 (46.6)	159 (30.1)
Boysen (28) **1993**	84 (11.8)/714/Oral cavity, pharynx, larynx	10 (11.9)	19 (22.6)	29 (34.5)	26 (31.0)
Jovanovic (29) **1994**	74 (10.2)/727/lip and oral cavity	8 (10.8)	19 (25.7)	47 (63.5)	0 (0)
Dhooge (30) **1998**	15 (11.8)/127/Oral cavity, pharynx, larynx, cervical esophagus	4 (26.7)	6 (40.0)	5 (33.3)	0 (0)
Fujita (31) **1998**	34 (21.5)/158/Larynx	2 (5.9)	14 (41.2)	8 (23.5)	10 (29.4)
León (32) **1999**	302 (16.4)/1,845/Oral cavity, pharynx, and larynx	27 (8.9)	100 (33.1)	122 (40.4)	53 (17.5)
Nikolaou (33) **2000**	42 (8.2)/514/Larynx	12 (28.6)	13 (31.0)	5 (11.9)	12 (28.6)
Rafferty (34) **2001**	36 (8.5)/425/Oral cavity, pharynx, and larynx	3 (8.3)	6 (16.7)	27 (75.0)	0 (0)
Khuri (35) **2001**	172 (15.3)/1,127/Oral cavity, pharynx, larynx	6 (3.5)	57 (33.1)	50 (29.1)	59 (34.3)
Ećimović (36) **2002**	369 (16.2)/2,275/Larynx	15 (4.1)	155 (42.0)	81 (21.9)	118 (32.0)
Dikshit (37) **2005**	145 (16.6)/876/Larynx and hypopharynx	15 (10.3)	55 (37.9)	52 (35.9)	23 (15.9)
Lin (38) **2005**	117 (9.3)/1,257/Oral cavity and larynx	7 (5.9)	48 (41.0)	40 (34.2)	22 (18.8)
Strobel (39) **2009**	56 (9.5)/589/Oral cavity, pharynx, and larynx	5 (8.9)	26 (46.4)	15 (32.6)	10 (17.9)
Xu (40) **2013**	30 (7.4)/406/oropharynx	1 (3.3)	7 (23.3)	19 (63.3)	3 (10.0)
Liao (41) **2014**	359 (22.9)/1,570/Oral cavity	14 (3.9)	25 (7.0)	281 (78.3)	39 (10.9)
Liao (42) **2015**	77 (4.2)/1,822/Oral cavity	4 (5.2)	0 (0)	66 (85.7)	7 (9.1)
González-Botas (43) **2016**	87 (15.0)/579/Oral cavity, pharynx, and larynx	5 (5.7)	32 (36.8)	33 (37.9)	17 (19.5)
Min (44) **2019**	1,191 (7.8)/15,261/Oral cavity	92 (7.7)	250 (21.0)	168 (14.1)	681 (57.2)
Bertolini (45) **2021**	222 (18.9)/1,177/Oral cavity, pharynx, and larynx	9 (4.1)	67 (30.2)	70 (31.5)	76 (34.2)
Milliet (46) **2021**	75 (5.8)/1,291/Oropharynx	7 (9.3)	13 (17.3)	50 (66.7)	5 (6.7)
Bugter (47) **2021**	246 (15.6)/1,581/Oral cavity, pharynx, and larynx	23 (9.3)	82 (33.3)	141 (57.3)	0 (0)
Luo (48) **2022**	73 (12.3)/593/Hypopharynx	23 (31.5)	13 (17.8)	14 (19.2)	23 (31.5)
All reviewed studies	**5,742 (11.2)/51,454/HN region**	**463 (8.1)**	**1,376 (23.9)**	**2,222 (38.7)**	**1,681 (29.3)**

HN, head and neck; SPT, second primary tumor.
The bold values were the summary data of enrolled studies.

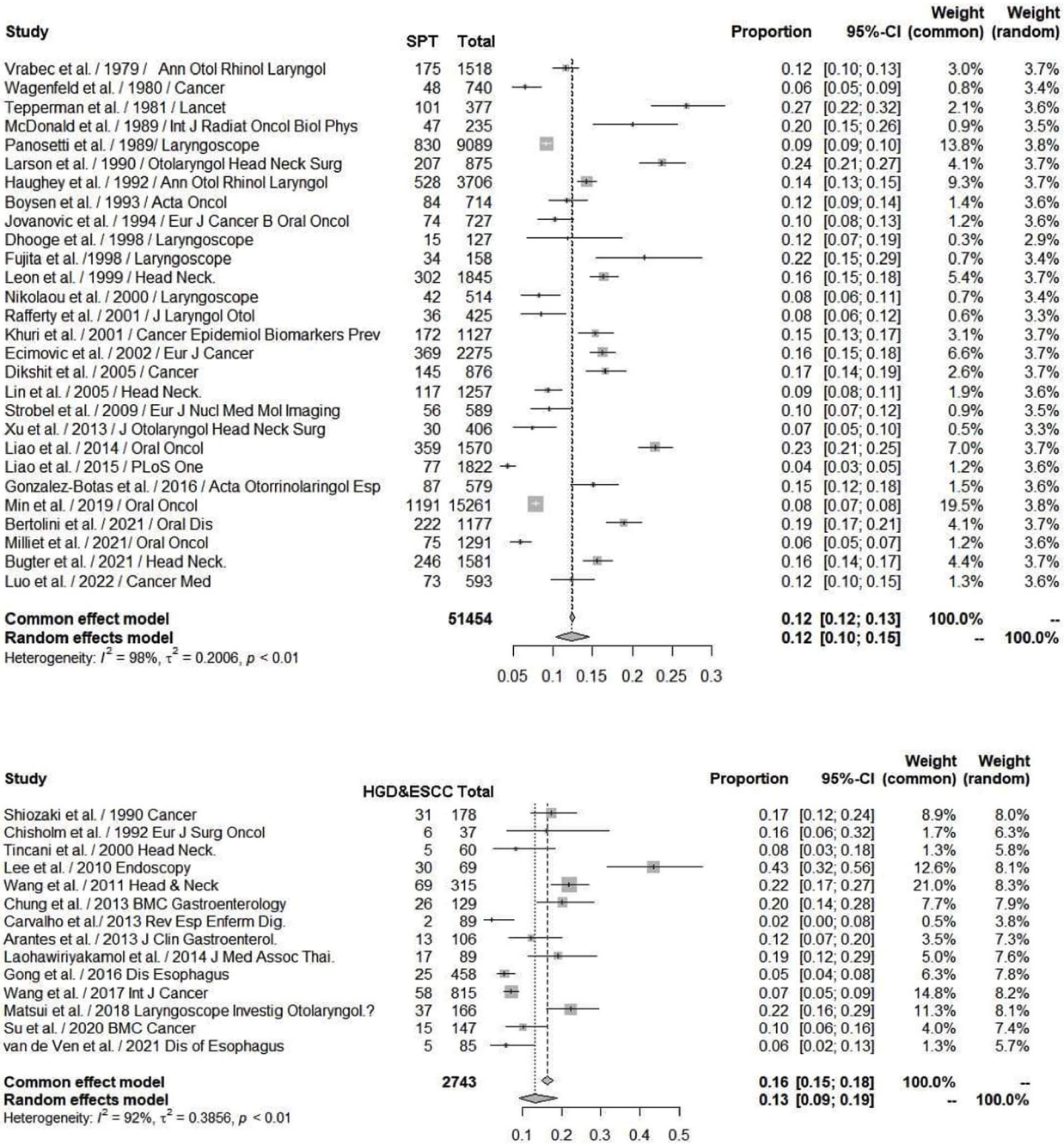


FIGURE 3 | *Upper*: Forest plots showing the reported proportion of SPTs among head and neck cancers with a random effect models due to significant heterogeneity, the overall SPT rate was 12% (95% CI, 10-15%). *Lower*: Forest plots showing a reported 13% incidence rate of HGD and ESCC (95% CI, 9-19%) by image-enhanced endoscopy screening among head and neck cancer patients. ESCC, esophageal squamous cell carcinoma; HGD, high-grade dysplasia; SPT, second primary tumor.

followed by the lungs and esophagus, which is similar to other studies (54). Metachronous SPTs are more prevalent than synchronous SPTs, and therefore, surveillance programs including investigations for SPTs are of paramount importance to improve the long-term care of HN cancer patients (17, 55).

Different Risk for Esophageal SPTs According to the Primary Site of HN Cancer

The risk factors for SPTs are different depending on the primary site of the index HN cancer. One study of 75,087 HN cancer

patients in the SEER database reported the highest risk for SPTs for primary hypopharyngeal cancer (SIR, 3.5; EAR, 307.1 per 10,000 person-years) and the lowest for laryngeal cancer (SIR, 1.9; EAR, 147.8 per 10,000 person-years) (56). Nasopharyngeal cancer (NPC) arises from a unique site with a large number of resident leukocytes, predominantly T-cells, together with other stromal cells. Therefore, the pathophysiology and tumor phenotype of NPC is quite different from other HN cancers, and the reported association between NPC and ESCC is lower than for other primary sites in the HN region. One large retrospective study of a cohort of 1,549 NPC patients following radiotherapy in Taiwan reported increased risks of developing SPTs in the HN region (SIR, 16.5; 95% CI, 10.0~26.8), stomach (SIR, 5.5; 95% CI, 2.2~11.4) and leukemia (SIR, 9; 95% CI, 1.9~26.3) (57). In a multicenter study of 8,947 NPC patients, 167 (1.9%) patients developed SPTs with increased risks of tongue cancer, non-Hodgkin's lymphoma, brain cancer, myeloid leukemia and non-melanoma skin cancer (58). Interestingly, the risk of developing SPTs has been shown to vary between different histological subtypes among NPC patients. A cross-sectional study of 1,175 NPC patients reported that SPTs, and especially those located in the HN region and UADT, were more prevalent in keratinizing NPC compared to non-keratinizing NPC (59). Another multicenter study of 3,166 NPC patients also reported significantly higher risks of cancer in the oral cavity, sarcoma, oropharynx, paranasal sinus, salivary gland, thyroid, skin and lungs (60).

Of note, a significantly lower risk of SPTs has been demonstrated among patients with oropharyngeal SCC in the *HPV* infection era (annual percentage change in EAR, -4.6%; $p = 0.03$), and that routine panendoscopy examinations are not even recommended in some studies (56, 61). A Canadian retrospective study of 406 oropharyngeal cancer patients reported a significantly lower incidence rate of SPTs in those who were p16-positive, which is indicative of *HPV*-related oropharyngeal cancer patients (0.7 per 100 patient-years vs. 8.5 in p16-negative patients, $p < 0.0001$) (40). In addition, the yield rate of field cancerization work-up (2.8% vs. 10.2%, $p = 0.02$) was lower in the *HPV*-positive than in the *HPV*-negative oropharyngeal cancer patients (40). Moreover, multivariate analysis from a multicenter study of 1,291 HN cancer patients showed that p16-negative tumor status ($p = 0.003$), tobacco/alcohol consumption ($p = 0.005$), and soft palate tumor site ($p = 0.009$) were significantly associated with a higher risk of metachronous SPTs (46). Furthermore, a higher proportion of metachronous SPTs arising outside the UADT was found in *HPV*-positive than in *HPV*-negative patients (46).

Second Primary Tumors of HN Region in Primary Esophageal Cancer Patients

Second primary neoplasms occur mutually in patients with UADT cancers. Patients with primary ESCC are also at risk of SPTs in the HN region. Analysis of data from a mean follow-up period of 76 months in a study of 285 ESCC patients showed 5-year cumulative occurrence rates of metachronous SPTs of the esophagus, HN region and stomach of 14.0%, 2.8% and 4.1%, respectively (62). Another study of 439 superficial esophageal cancer patients reported that 53 metachronous HN cancers developed in 40 (9.1%) patients after a median follow-up period of 46 months, and the cumulative incidence rates of metachronous HN cancers at 3, 5, and 7 years were 5.3%, 9.7%, and 17.2%, respectively (63). A systematic review of 6,483 ESCC patients from 12 studies in Japan revealed a pooled prevalence of HN SPTs of 6.7% (95% CI, 4.9~8.4%), including 48.2% synchronous and 51.8% metachronous SPTs, 85.3% at an early stage, and 60.3% located in the hypopharynx (18).

Prognosis of HN Cancer Patients With Esophageal SPTs

Esophageal SPTs not only occur synchronously or metachronously, but also have a negative impact on the prognosis of HN cancer patients (64). The 15-year survival rate of HN cancer patients with SPTs is lower than in those without SPTs (22% vs. 54%), and the prognosis is especially poor with a 5-year survival rate of only 6% in those with esophageal SPTs (vs. 25% in those with all SPTs) (2, 3, 26). Another study also demonstrated lower 5-year (68% vs. 76%) and 10-year (26% vs. 57%) overall survival rates in laryngeal cancer patients who developed SPTs ($p = 0.003$) (31). A nationwide analysis of 93,891 HN cancer patients from the Taiwan Cancer Registry reported that 9,996 (10.6%) patients presented with SPTs, and that those with SPTs had a significantly lower survival rate (univariate analysis: HR, 2.59; 95% CI, 2.53-2.65; multivariate analysis: HR, 2.34; 95% CI, 2.28-2.40) (65).

To summarize, the risk and distribution of SPTs differ significantly according to the subsite of the index primary HN cancer, with a lower risk in laryngeal and *HPV*-positive oropharyngeal cancer patients. About 11.2% of HN cancer patients develop either synchronous or metachronous SPTs at the HN region (38.7%), lung and bronchus (23.9%), and esophagus (8.1%) (**Table 1**). The occurrence of ESCC is especially associated with a poor prognosis, and thus identifying esophageal SPTs is crucial in screening and surveillance programs for HN cancer patients.

IMAGE-ENHANCED ENDOSCOPIC SCREENING AND RISK FACTORS FOR ESOPHAGEAL SPTS IN HN CANCER PATIENTS

Esophagogastroduodenoscopy is the most reliable diagnostic tool for esophageal neoplasms, especially using an image-enhanced endoscopy (IEE) system, which is composed of optical- and dye-based technology (49, 66, 67). Among several IEE techniques, narrow-band imaging (NBI) and chromoendoscopy with Lugol's solution are widely used for screening ESCC (49, 66–68). By using narrow-bandwidth filters to remove red light and narrow wavelengths of green (540 nm) and blue (415 nm) light, NBI can improve visualization of hemoglobin-rich vascular microstructures (**Figure 4**) (49). Because the color of gastrointestinal mucosa is primarily determined by

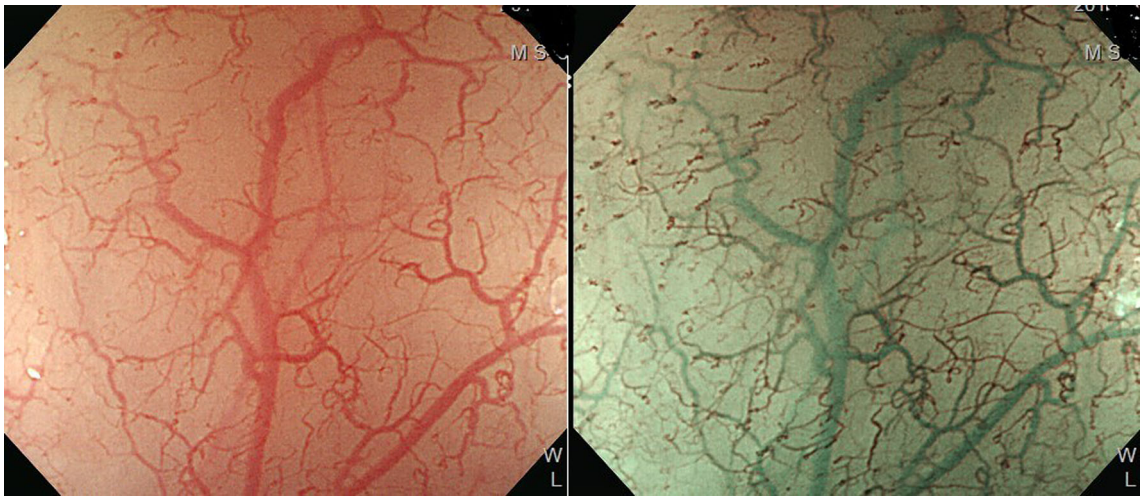


FIGURE 4 | Improved visualization of microvascular structure under narrow-band imaging endoscopy (*Left*: conventional white-light imaging. *Right*: narrow-band imaging.).

hemoglobin, and neovascularization occurs in neoplastic squamous epithelium of the esophagus, the light emitted from NBI is absorbed by neoplastic mucosa more than healthy mucosa. Therefore, early neoplasms, which usually have a flat morphology, can be differentiated from normal mucosa by dark brownish discoloration compared with the greenish color of healthy mucosa under NBI (**Figure 5**). In addition, when combining a magnifying endoscope with an NBI system, the microvascular pattern of neoplastic squamous cell epithelium can be well delineated (**Figure 6**) (49, 67, 69). These microvessels seen under magnifying NBI, so-called intra-epithelial papillary capillary loops, can also predict tumor invasion depth with accuracy of 90.5% (69). Among dye-based IEE, iodine-containing Lugol's solution is commonly used for ESCC screening. Normal glycogen-abundant squamous epithelium reacts with Lugol's solution, while dysplastic mucosa with diminished or absent glycogen remains unstained (67, 68, 70). By spraying Lugol's solution on esophageal mucosa, unstained areas are indicative of dysplastic or cancerous parts. Moreover, when unstained mucosa turns pink within a few minutes, high-grade dysplasia or squamous cell carcinoma can be diagnosed with a sensitivity of 91.9% and specificity of 94.0% (**Figure 7**) (68).

Both NBI and Lugol's chromoendoscopy (LCE) are effective real-time screening endoscopic techniques for the early detection of esophageal neoplasms. A meta-analysis of 4,918 esophageal and HN cancer patients from 16 prospective and randomized trials showed that NBI and LCE had better diagnostic performance than conventional white-light imaging, with pooled sensitivity, specificity and area under the receiver operating characteristic curve of 87% (95% CI, 83~90%) and 88% (95% CI, 85~91%) versus 53% (95% CI, 48~59%), 99% (95% CI, 98~99%) and 95% (95% CI, 94~96%) versus 63% (95% CI, 61~66%), and 97% and 82% versus 66%, respectively (66). Given that most esophageal SPTs detected in HN cancer patients are at asymptomatic premalignant or early cancer stages, these lesions might be overlooked by white-light imaging or even advanced cross-sectional and radionuclide imaging modalities. In a study of 147 HN cancer patients, suspicious esophageal SPTs were identified by position emission tomography/computed tomography (PET/CT) in 8 (5.4%) and by NBI endoscopy in 35 (23.8%) patients (71). In addition, the diagnostic sensitivity of NBI endoscopy (100.0%) was superior to whole body PET/CT (33.3%) in detecting esophageal SPTs (71). In a review of 14 studies with 2,743 HN cancer patients, IEE screening identified esophageal high grade dysplasia or invasive carcinoma in 13% (95% CI, 9-19% with a random effects model) of the patients (**Table 2**, **Figure 3**) (8, 10–12, 70–79). Most of the esophageal SPTs were at an early stage without tumor-related obstructive symptoms. Therefore, if these esophageal SPTs had not been identified, the patients may have had a poor prognosis from esophageal cancer.

There are many common risk factors for HN and esophageal cancers. Among environmental factors, alcohol is one of the most important carcinogens for esophageal cancer (1, 19, 21). The results from a meta-analysis of 8 cohort and 11 case-control studies showed that alcohol drinking was associated with significantly increased risk of UADT SPTs (RR, 2.97; 95% CI, 1.96~4.50), and that every increase of 10 g/day in alcohol intake resulted in a significantly increased RR of 1.09 (95% CI, 1.04-1.14) for UADT SPTs in a dose-response relationship (80). Alcohol metabolizing enzyme genes are disease modifiers which are responsible for the increased risk of cancer after alcohol consumption (81). Ethanol is metabolized to acetaldehyde by alcohol dehydrogenase (ADH), then converted to acetate by acetaldehyde dehydrogenase (ALDH). The intermediate metabolized product, acetaldehyde is not only associated with unpleasant disulfiram-like reactions such as facial flushing, nausea, vomiting, tachycardia and hypotension, but also increased oxidant stress, inflammation and reactions with deoxynucleosides, leading to the formation of deoxyribonucleic acid adducts and subsequently cancerization

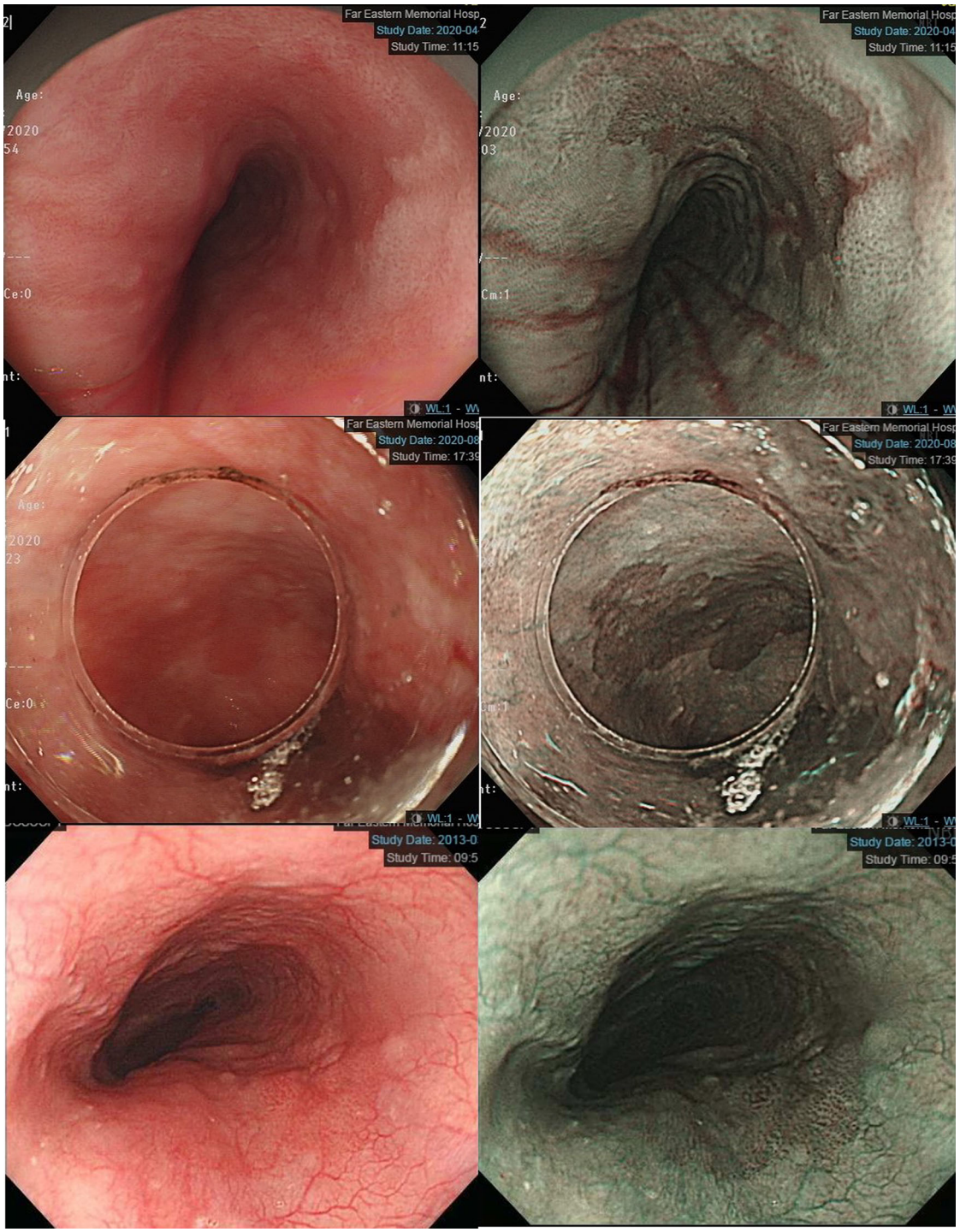


FIGURE 5 | *Left panels*: Early esophageal neoplasm with barely visible flat morphology under conventional white-light endoscopy. *Right panels*: Dark brownish color compared with the greenish color of healthy mucosa under narrow-band imaging endoscopy.

(19, 81, 82). The results from a case-control study of 120 HN cancer and 138 ESCC patients in Taiwan demonstrated that the minor alleles of ADHB (rs1229984) and ALDH2 (rs671) were associated with an increased risk of UADT cancers (OR, 3.53 and 2.59; 95% CI, 2.14~5.80 and 1.79~3.75), and also that they potentiated the carcinogenic effects of alcohol (OR, 53.44 and 70.08; 95% CI, 25.21~113.29 and 33.65~145.95) (19). In addition, the haplotypes GAGC and CCAATG on chromosome

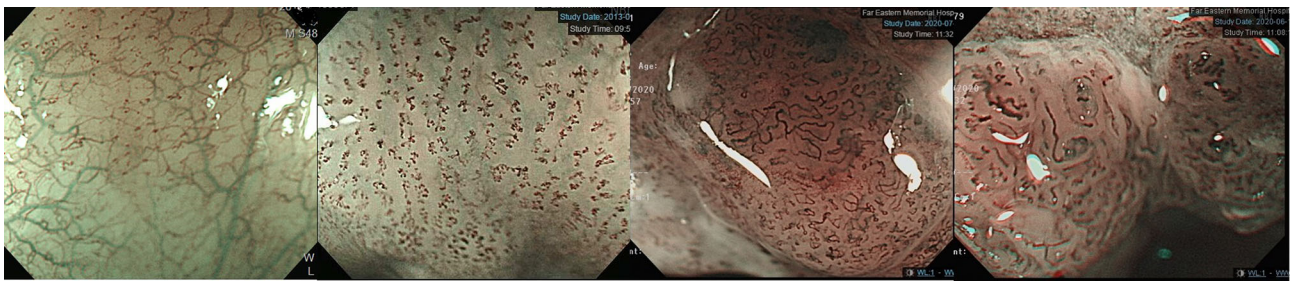


FIGURE 6 | JES classification of microvessel morphology of IPCL. *From left to right*: **JES type A-** Normal IPCL without irregularity. **JES type B1-** Abnormal microvessels with severe irregularity, meandering caliber or highly dilated proliferative abnormal vessels with a loop-like formation. **JES type B2-** Abnormal microvessels with severe irregularity, meandering calibers or highly dilated proliferative abnormal vessels without a loop-like formation. **JES type B3-** Highly dilated microvessels with three times as many calibers than usual type B2 vessels. IPCL, intraepithelial papillary capillary loop; JES, Japanese Esophageal Society.

4 and 12, respectively, have been associated with a higher risk of HN and esophageal cancers (19). Another case-control study with age- and gender-matched 164 HN cancer patients showed that polymorphisms in ADH1B (OR, 2.09; 95% CI, 1.15~3.18; $p < 0.05$) and ALDH2 (OR, 5.19; 95% CI, 2.44~11.00; $p < 0.001$) increased the risk of developing multiple SPTs (20). Thus, HN cancer patients who are alcohol drinkers have a higher risk of esophageal SPT, particularly those carrying risk genetic polymorphisms of alcohol-metabolizing enzymes.

Primary sites of HN cancer are associated with different risk of developing esophageal SPTs. Compared with oral cavity and nasopharyngeal cancers, primary malignancy of the hypopharynx, *HPV*-negative oropharynx, and larynx are more likely to have esophageal SPTs (8, 11, 12, 50, 53, 54, 71). Other demographic data, including older age, comorbidities, lower body mass index, advanced stages of primary HN cancer and alcohol flushing syndrome have also been associated with a higher risk of esophageal SPTs (8, 12, 47). A systematic review identified 51 genes that were significantly associated with an increased risk of SPTs among HN cancer patients (83). In addition, the presence of multiple Lugol-voiding lesions, which are indicative of dysplastic or cancerous lesions in the esophagus, has also been reported to be a significant risk factor for developing both synchronous and metachronous SPTs (62, 84). A 13-year follow-up study of 682 patients with esophageal dysplasia reported that 23.7%, 50% and 73.9% of patients with low-grade, moderate, and high-grade dysplasia (HGD) developed invasive carcinoma (85). The molecular changes in Lugol-voiding mucosa precede the cancerization process, and the hotspot *p53* mutation has been identified in 20% and 40% of non-dysplastic and dysplastic Lugol-voiding mucosa (84). Therefore, when multiple Lugol-unstained areas are noted after LCE screening, a shorter interval of IEE surveillance for metachronous esophageal SPTs is mandatory.

For HN cancer patients at risk of esophageal neoplasms, endoscopic screening and surveillance, especially using IEE techniques with NBI endoscopy and LCE, are crucial to identify esophageal SPTs. Before the development of obstructive symptoms from advanced esophageal neoplasms, the early detection of esophageal SPTs is one of the most important management strategies to improve the overall prognosis of HN cancer patients (**Figure 8**).

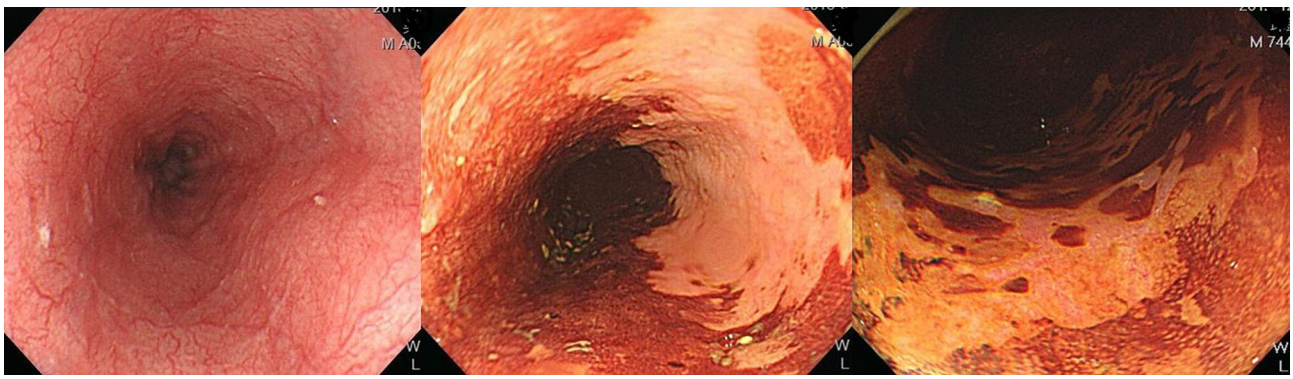


FIGURE 7 | Esophageal high-grade dysplastic lesion. *Left*: Normal appearance upon white-light endoscopy. *Middle*: Lugol-voiding unstained mucosa. *Right*: The color of Lugol-unstained mucosa turns pink in a few minutes.

TABLE 2 | Image-enhanced endoscopic screening of synchronous or metachronous esophageal neoplasm in HN cancer patients.

Author/Reference no.Year	Patient no./Study design/ Endoscopy techniques	Incidence (excluding LGD) (%)/Lesions	Treatment
Shiozaki (72) **1990**	178 oral cavity, pharynx, larynx/ Prospective/WLE, LCE	17.4/22 Dysplasia, 9 ESCC	CCRT, esophagectomy or laser
Chisholm (70) **1992**	37 oral cavity, pharynx, larynx/ Prospective/WLE, LCE	16.2/6 ESCC	Not mentioned
Tincani (73) **2000**	60 oral cavity, pharynx, larynx/ Prospective/WLE, LCE	8.3/5 ESCC	Esophagectomy
Lee (74) **2010**	69 oral cavity, pharynx, larynx/ Prospective/WLE, NBI, LCE	30.4/5 LGD, 8 HGD, 22 ESCC	CCRT or esophagectomy for advanced cancers, ER for superficial neoplasm, or no treatment
Wang (11) **2011**	315 oral cavity, pharynx, larynx/ Prospective/WLE, NBI, LCE	21.9/22 HGD, 47 ESCC	CCRT or esophagectomy for advanced cancers, ER for superficial neoplasm
Chung (8) **2013**	129 oral cavity, pharynx, larynx/ Prospective/WLE, NBI, LCE	20.2/11 LGD, 14 HGD, 12 ESCC	Extended RT field or esophagectomy for advanced cancers, ER or radiofrequency ablation for superficial neoplasm
Carvalho (75) **2013**	89 oral cavity, pharynx, larynx/ Prospective/WLE, LCE	2.2/2 HGD	ER
Arantes (76) **2013**	106 oral cavity, pharynx, larynx/ Prospective/WLE, FICE	12.3/3 HGD, 10 ESCC	CCRT and ER
Laohawiriyakamol (77) **2014**	89 oral cavity, pharynx, larynx/ Retrospective/WLE, LCE	12.4/6 Dysplasia, 11 ESCC	Not mentioned
Gong (78) **2016**	458 oral cavity, pharynx, larynx/ Prospective/WLE, NBI, LCE	5.2/3 LGD, 15 HGD, 10 ESCC	CCRT or esophagectomy for advanced cancers, ER for superficial neoplasm, or no treatment
Wang (12) **2017**	815 oral cavity, pharynx, larynx/ Prospective/WLE, NBI, LCE	7.1/66 LGD, 29 HGD, 29 ESCC	Not mentioned
Matsui (79) **2018**	166 oral cavity/retrospective/ WLE, FICE, LCE	22.3/37 ESCC	CCRT or esophagectomy for advanced cancers, ER for superficial neoplasm
Su (71) **2020**	147 oral cavity, pharynx, larynx/ Retrospective/WLE, NBI	10.2/5 HGD, 10 ESCC	Not mentioned
van de Ven (10) **2021**	85 oral cavity, pharynx, larynx/ Prospective/WLE, NBI, LCE	5.9/3 LGD, 4 HGD, 1 ESCC	Extended RT field for advanced cancers, ER for superficial neoplasm

CCRT, concurrent chemoradiotherapy; ER, endoscopic resection; ESCC, esophageal squamous cell carcinoma; FICE, Fuji Intelligent Color Enhancement; HGD, high grade dysplasia; LCE, Lugol's chromoendoscopy; LGD, low grade dysplasia; NBI, narrow band imaging; RT, radiotherapy; WLE, white-light endoscopy.

SCREENING AND TREATMENT STRATEGY OF ESOPHAGEAL SPTS FOR HN CANCER PATIENTS

After screening for esophageal SPTs, HN cancer patients who are free from synchronous esophageal SPTs have the best outcomes (16). Thus, before starting treatment of newly diagnosed HN cancers, risk stratification and identification of synchronous esophageal SPTs could modify the oncological treatment plan (8). When considering ESCC treatment, surgical esophagectomy was the traditional curative therapeutic option. However, in the early 20th century, with advances in minimally invasive endoscopic resection techniques, early esophageal neoplasms could be managed by endoscopic submucosal dissection (ESD) and radiofrequency ablation (RFA) (86, 87). Due to the low risk of nodal or distant metastasis of superficial esophageal neoplasms, ESD can be considered as the first-line therapy for HGD or ESCC limited to the epithelium and lamina propria without lymphovascular invasion, while RFA can be considered for flat-type esophageal HGD or ESCC confined above the lamina propria (86–88). The overall curative resection and recurrence rates of esophageal neoplasms for ESD have been reported to be 78~100% and 0~2.6%, respectively, with complete remission and recurrence rates of 50~100% and 0~50% for RFA (86, 88). Five-year overall, disease-specific and metastasis-free survival rates above 90% have been reported after ESD for early esophageal neoplasms (86, 89, 90). Compared with surgical intervention, ESD (relative hazard, 0.89; 95% CI, 0.51~1.56; p = 0.68) has comparable long-term outcomes for early esophageal neoplasms, with a better quality of life and lower rate of adverse events (86, 90, 91). However, stricture complications are one of the most important concerns after ESD for large size neoplasms or those which involve more than 75% of the circumference (86, 90, 91). Most post-ESD strictures can be managed by endoscopic balloon dilation or prophylactic steroid therapy. As a result, identifying early esophageal SPTs in HN cancer patients could be a triage for screening and surveillance programs, and could also provide a chance for minimally invasive endoscopic resection with curative intent of early esophageal SPTs.

When considering the treatment strategy, the curability of both primary and secondary neoplasms must be carefully evaluated and discussed with a multidisciplinary approach. In HN cancer patients, prior treatment of the primary cancer often affects the treatment of esophageal SPTs. Trismus, malnutrition with cancer cachexia, performance status, the location of the esophageal SPT, and patient preference are important factors which should be taken into account. The treatment for esophageal SPTs, including endoscopic resection, concurrent chemoradiotherapy (CCRT), surgical intervention or no treatment, varies between studies due

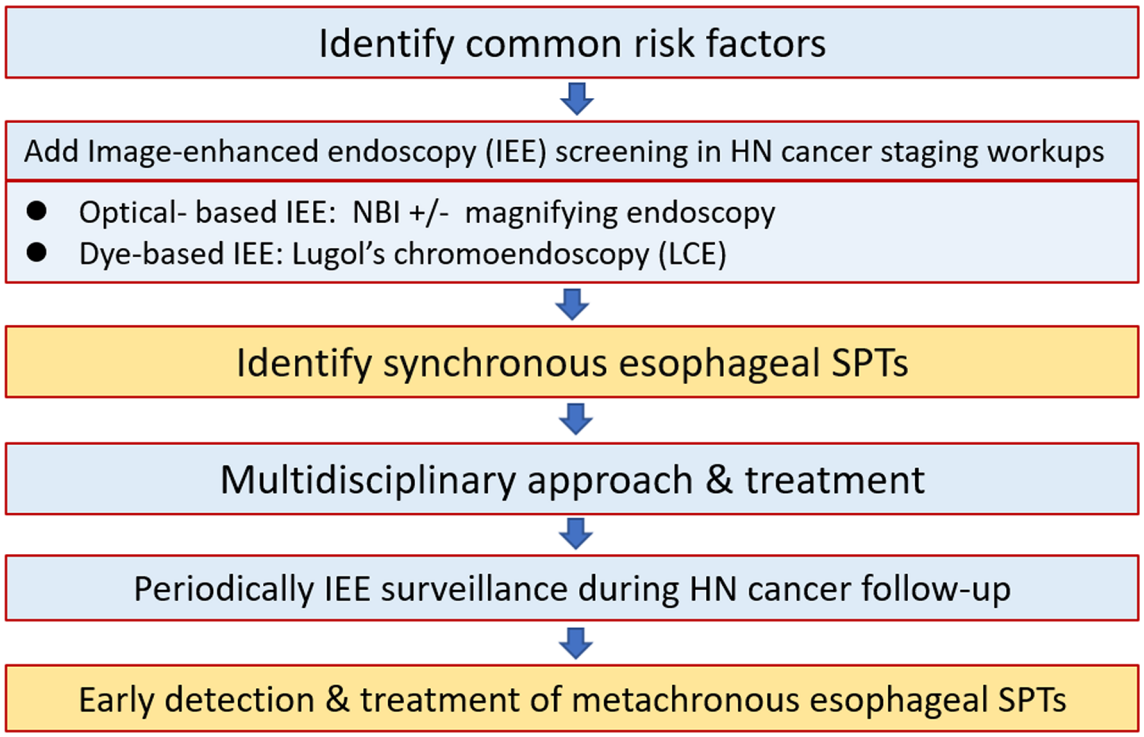


FIGURE 8 | Approach algorithm for head and neck cancer patients at risk of esophageal second primary tumors. HN, head and heck; IEE, image-enhanced endoscopy; LCE, Lugol's chromoendoscopy; NBI, narrow-band imaging; SPT, second primary tumor.

to the heterogeneous characteristics of HN cancer patients (**Table 2**). Cox proportional regression analysis of the SEER database which enrolled 3,038 HN cancer patients showed that those with SPTs of the HN region who underwent conservative surgery with radiation had the best 5-year overall survival rate (22.6%), those with lung SPTs who underwent radical surgery had the best 2-year overall survival rate (60.8%), and that there was no difference in the prognosis between treatment groups in those with esophageal SPTs (64). However, in a prospective study with long-term outcome analysis of 145 HN cancer patients, those with early esophageal SPTs who underwent aggressive treatment of both primary and secondary neoplasms had similar overall survival compared to HN cancer patients without esophageal SPTs ($p = 0.47$) (92). Definitive CCRT of esophageal cancer patients with synchronous HN SPTs can also safely be offered to improve overall survival, and those who receive CCRT have been shown to have better survival than those with radiotherapy alone (93).

Screening of esophageal SPTs by IEE should be performed in every newly diagnosed HN cancer patient, and regular IEE surveillance is also important to detect metachronous esophageal neoplasms. After identifying esophageal SPTs in HN cancer patients, management of neoplasms at the primary and secondary sites is quite complex and should be individualized according to the patient's condition. It depends on the stage and survival of the primary and secondary tumors, prior treatments, expertise in endoscopic resection techniques and CCRT, as well as the patient's performance and preference. Close cooperation between medical staff members including HN surgeon, gastroenterologist, endoscopist, oncologist and radio-oncologist are essential in a multidisciplinary approach.

SUMMARY

The development of synchronous or metachronous SPTs is more frequently being identified due to advances in diagnostic modalities, and it is an emerging issue in oncology medicine. SPTs are not uncommon among HN cancer patients, particularly those located in the HN region, lungs and esophagus. Patients with HN cancer and concomitant esophageal SPTs have the worst prognosis. Therefore, identifying esophageal SPTs in HN cancer patients is of paramount importance for risk stratification and to guide the treatment strategy. IEE, especially using NBI endoscopy and LCE, improves the diagnostic performance in detecting early esophageal neoplasms. Several studies have demonstrated a high diagnostic yield of IEE to identify esophageal SPTs at an early stage in HN cancer patients, particularly in patients at high risk, such those with primary sites of the hypopharynx and larynx, alcoholism with flushing syndrome, older age, and advanced stage primary HN cancer. In addition, with minimally invasive endoscopic resection and radiotherapy techniques, HN cancer patients with early esophageal neoplasms can be managed

without surgical interventions to allow for a better quality of life. However, there are currently no standardized surveillance protocols with regards to the interval and therapeutic options for primary HN cancers and esophageal SPTs. In terms of personalized medicine, the treatment strategy should be individualized and discussed by a multidisciplinary team involving gastroenterologists, endoscopists, oncologists, radiologists, and HN and chest surgeons. Most of the enrolled studies in this review were retrospective or case-control design and the results might be influenced by the bias upon independent literature review. More well-designed prospective studies are warranted to establish the most appropriate treatment and surveillance programs to improve overall outcomes for HN cancer patients with esophageal SPTs.

AUTHOR CONTRIBUTIONS

Conceptualization, C-SC, L-JL and P-WS; Methodology, C-SC, L-JL and P-WS; Software, C-SC and L-JL; Validation, C-SC, L-JL, C-YW, W-CL, C-HH, T-HL, C-YL and P-WS; Formal analysis, C-SC and L-JL; Investigation, C-SC, L-JL, C-YW, W-CL, C-HH, T-HL, C-YL and P-WS; Resources, C-SC, L-JL, C-YW, W-CL, C-HH, T-HL, C-YL and P-WS; Data curation, C-SC, L-JL and P-WS; Writing—original draft preparation, C-SC; Writing—review and editing, C-SC, L-JL, D-YK and P-WS; Visualization, C-SC, L-JL, C-YW, W-CL, C-HH, T-HL, C-YL and P-WS; Supervision, C-SC, L-JL and P-WS; Project administration, C-SC; Funding acquisition, C-SC and P-WS. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. CP WILD, Weiderpass E, Stewart BW. *World Cancer Report: Cancer Research for Cancer Prevention*. Lyon, France: International Agency for Research on Cancer (2020). Available at: http://publications.iarc.fr/586.Licence.
2. Chen MC, Chen PT, Chan CH, Yang CT, Chen CC, Huan CE, et al. Second Primary Esophageal or Lung Cancer in Patients With Head and Neck Carcinoma in Taiwan: Incidence and Risk in Relation to Primary Index Tumor Site. *J Cancer Res Clin Oncol* (2011) 137(1):115–23. doi: 10.1007/s00432-010-0865-0
3. León X, Ferlito A, Myer CM3rd, Saffiotti U, Shaha AR, Bradley PJ, et al. Second Primary Tumors in Head and Neck Cancer Patients. *Acta Otolaryngol* (2002) 122(7):765–78. doi: 10.1080/003655402/000028048
4. Slaughter DP, Southwick HW, Smejkal W. Field Cancerization in Oral Stratified Squamous Epithelium; Clinical Implications of Multicentric Origin. *Cancer* (1953) 6(5):963–8. doi: 10.1002/1097-0142(195309)6:5<963::AID-CNCR2820060515>3.0.CO;2-Q
5. Coyte A, Morrison DS, McLoone P. Second Primary Cancer Risk - the Impact of Applying Different Definitions of Multiple Primaries: Results From a Retrospective Population-Based Cancer Registry Study. *BMC Cancer* (2014) 14:272. doi: 10.1186/1471-2407-14-272
6. Chung CS, Lee YC, Wang CP, Ko JY, Wang WL, Wu MS, et al. Secondary Prevention of Esophageal Squamous Cell Carcinoma in Areas Where Smoking, Alcohol, and Betel Quid Chewing are Prevalent. *J Formos Med Assoc* (2010) 109(6):408–21. doi: 10.1016/S0929-6646(10)60072-1
7. Chung CS, Lee YC, Wu MS. Prevention Strategies for Esophageal Cancer: Perspectives of the East vs. West. *Best Pract Res Clin Gastroenterol* (2015) 29(6):869–83. doi: 10.1016/j.bpg.2015.09.010
8. Chung CS, Liao LJ, Lo WC, Chou YH, Chang YC, Lin YC, et al. Risk Factors for Second Primary Neoplasia of Esophagus in Newly Diagnosed Head and Neck Cancer Patients: A Case-Control Study. *BMC Gastroenterol* (2013) 13:154. doi: 10.1186/1471-230X-13-154
9. Hoxhaj I, Hysaj O, Vukovic V, Leoncini E, Amore R, Pastorino R, et al. Occurrence of Metachronous Second Primary Cancer in Head and Neck Cancer Survivors: A Systematic Review and Meta-Analysis of the Literature. *Eur J Cancer Care (Engl)* (2020) 29(5):e13255. doi: 10.1111/ecc.13255
10. van de Ven S, de Graaf W, Bugter O, Spaander MCW, Nikkessen S, de Jonge PJF, et al. Screening for Synchronous Esophageal Second Primary Tumors in Patients With Head and Neck Cancer. *Dis Esophagus* (2021) 11(10):doab037. doi: 10.1093/dote/doab037
11. Wang WL, Lee CT, Lee YC, Hwang TZ, Wang CC, Hwang JC, et al. Risk Factors for Developing Synchronous Esophageal Neoplasia in Patients With Head and Neck Cancer. *Head Neck* (2011) 33(1):77–81. doi: 10.1002/hed.21397
12. Wang YK, Chuang YS, Wu TS, Lee KW, Wu CW, Wang HC, et al. Endoscopic Screening for Synchronous Esophageal Neoplasia Among Patients With Incident Head and Neck Cancer: Prevalence, Risk Factors, and Outcomes. *Int J Cancer* (2017) 141(10):1987–96. doi: 10.1002/ijc.30911
13. Wang YW. *2018 Annual Report of Health Promotion Administration*. Health Promotion Administration, Ministry of Health and Welfare. (2018).
14. Tiwana MS, Wu J, Hay J, Wong F, Cheung W, Olson RA. 25 Year Survival Outcomes for Squamous Cell Carcinomas of the Head and Neck: Population-Based Outcomes From a Canadian Province. *Oral Oncol* (2014) 50(7):651–6. doi: 10.1016/j.oraloncology.2014.03.009
15. Leoncini E, Vukovic V, Cadoni G, Pastorino R, Arzani D, Bosetti C, et al. Clinical Features and Prognostic Factors in Patients With Head and Neck Cancer: Results From a Multicentric Study. *Cancer Epidemiol* (2015) 39(3):367–74. doi: 10.1016/j.canep.2015.02.004
16. Chung CS, Lo WC, Chen KC, Lin CL, Wen MH, Hsieh CH, et al. Clinical Benefits From Endoscopy Screening of Esophageal Second Primary Tumor for Head and Neck Cancer Patients: Analysis of a Hospital-Based Registry. *Oral Oncol* (2019) 96:27–33. doi: 10.1016/j.oraloncology.2019.06.038
17. Panosetti E, Luboinski B, Mamelle G, Richard JM. Multiple Synchronous and Metachronous Cancers of the Upper Aerodigestive Tract: A Nine-Year Study. *Laryngoscope* (1989) 99:1267–73. doi: 10.1288/00005537-198912000-00011
18. van de Ven S, Bugter O, Hardillo JA, Bruno MJ, Baatenburg de Jong RJ, Koch AD. Screening for Head and Neck Second Primary Tumors in Patients With Esophageal Squamous Cell Cancer: A Systematic Review and Meta-Analysis. *United European Gastroenterol J* (2019) 7(10):1304–11. doi: 10.1177/2050640619856459
19. Chung CS, Lee YC, Liou JM, Wang CP, Ko JY, Lee JM, et al. Tag Single Nucleotide Polymorphisms of Alcohol-Metabolizing Enzymes Modify the Risk of Upper Aerodigestive Tract Cancers: HapMap Database Analysis. *Dis Esophagus* (2014) 27(5):493–503. doi: 10.1111/j.1442-2050.2012.01437.x
20. Chien HT, Young CK, Chen TP, Liao CT, Wang HM, Cheng SD, et al. Alcohol-Metabolizing Enzymes' Gene Polymorphisms and Susceptibility to Multiple Head and Neck Cancers. *Cancer Prev Res (Phila)* (2019) 12(4):247–54. doi: 10.1158/1940-6207.CAPR-18-0449
21. Lee CH, Lee JM, Wu DC, Hsu HK, Kao EL, Huang HL, et al. Independent and Combined Effects of Alcohol Intake, Tobacco Smoking and Betel Quid Chewing on the Risk of Esophageal Cancer in Taiwan. *Int J Cancer* (2005) 113(3):475–82. doi: 10.1002/ijc.20619
22. Vrabec DP. Multiple Primary Malignancies of the Upper Aerodigestive System. *Ann Otol Rhinol Laryngol* (1979) 88:846–54. doi: 10.1177/000348947908800620
23. Wagenfeld DJ, Harwood AR, Bryce DP, van Nostrand AW, DeBoer G. Second Primary Respiratory Tract Malignancies in Glottic Carcinoma. *Cancer* (1980) 46(8):1883–6. doi: 10.1002/1097-0142(19801015)46:8<1883::AID-CNCR2820460830>3.0.CO;2-G
24. Tepperman BS, Fitzpatrick PJ. Second Respiratory and Upper Digestive Tract Cancers After Oral Cancer. *Lancet* (1981) 2(8246):547–9. doi: 10.1016/S0140-6736(81)90938-7
25. McDonald S, Haie C, Rubin P, Nelson D, Divers LD. Second Malignant Tumors in Patients With Laryngeal Carcinoma: Diagnosis, Treatment, and Prevention. *Int J Radiat Oncol Biol Phys* (1989) 17(3):457–65. doi: 10.1016/0360-3016(89)90095-3

26. Larson JT, Adams GL, Fattah HA. Survival Statistics for Multiple Primaries in Head and Neck Cancer. *Otolaryngol Head Neck Surg* (1990) 103(1):14–24. doi: 10.1177/019459989010300103
27. Haughey BH, Gates GA, Arfken CA, Harvey J. Meta-Analysis of Second Malignant Tumors in Head and Neck Cancer: The Case for an Endoscopic Screening Protocol. *Ann Otol Rhinol Laryngol* (1992) 102:105–12. doi: 10.1177/000348949210100201
28. Boysen M, Loven JO. Second Malignant Neoplasms in Patients With Head and Neck Squamous Cell Carcinomas. *Acta Oncol* (1993) 32(3):283–8. doi: 10.3109/02841869309093596
29. Jovanovic A, van der Tol IG, Kostense PJ, Schulten EA, de Vries N, Snow GB, et al. Second Respiratory and Upper Digestive Tract Cancer Following Oral Squamous Cell Carcinoma. *Eur J Cancer B Oral Oncol* (1994) 30B(4):225–9. doi: 10.1016/0964-1955(94)90001-9
30. Dhooge IJ, De Vos M, Van Cauwenberge PB. Multiple Primary Malignant Tumors in Patients With Head and Neck Cancer: Results of a Prospective Study and Future Perspectives. *Laryngoscope* (1998) 108(2):250–6. doi: 10.1097/00005537-199802000-00017
31. Fujita M, Rudoltz MS, Canady DJ, Patel P, Machtay M, Pittard MQ, et al. Second Malignant Neoplasia in Patients With T1 Glottic Cancer Treated With Radiation. *Laryngoscope* (1998) 108(12):1853–5. doi: 10.1097/00005537-199812000-00016
32. León X, Quer M, Diez S, Orús C, López-Pousa A, Burgués J. Second Neoplasm in Patients With Head and Neck Cancer. *Head Neck* (1999) 21(3):204–10. doi: 10.1002/(SICI)1097-0347(199905)21:3<204::AID-HED4>3.0.CO;2-7
33. Nikolaou AC, Markou CD, Petridis DG, Daniilidis IC. Second Primary Neoplasms in Patients With Laryngeal Carcinoma. *Laryngoscope* (2000) 110 (1):58–64. doi: 10.1097/00005537-200001000-00012
34. Rafferty MA, O'Dwyer TP. Secondary Primary Malignancies in Head and Neck Squamous Cell Carcinoma. *J Laryngol Otol* (2001) 115(12):988–91. doi: 10.1258/0022215011909567
35. Khuri FR, Kim ES, Lee JJ, Winn RJ, Benner SE, Lippman SM, et al. The Impact of Smoking Status, Disease Stage, and Index Tumor Site on Second Primary Tumor Incidence and Tumor Recurrence in the Head and Neck Retinoid Chemoprevention Trial. *Cancer Epidemiol Biomarkers Prev* (2001) 10(8):823–9. PMID: 11489748.
36. Ećimović P, Pompe-Kirn V. Second Primary Cancers in Laryngeal Cancer Patients in Slovenia, 1961-1996. *Eur J Cancer* (2002) 38(9):1254–60. doi: 10.1016/S0959-8049(02)00012-6
37. Dikshit RP, Boffetta P, Bouchardy C, Merletti F, Crosignani P, Cuchi T, et al. Risk Factors for the Development of Second Primary Tumors Among Men After Laryngeal and Hypopharyngeal Carcinoma. *Cancer* (2005) 103 (11):2326–33. doi: 10.1002/cncr.21051
38. Lin K, Patel SG, Chu PY, Matsuo JMS, Singh B, Wong RJ, et al. Second Primary Malignancy of the Aerodigestive Tract in Patients Treated for Cancer of the Oral Cavity and Larynx. *Head Neck* (2005) 27(12):1042–8. doi: 10.1002/hed.20272
39. Strobel K, Haerle SK, Stoeckli SJ, Schrank M, Soyka JD, Veit-Haibach P, et al. Head and Neck Squamous Cell Carcinoma (HNSCC)–detection of Synchronous Primaries With (18)F-FDG-PET/Ct. *Eur J Nucl Med Mol Imaging* (2009) 36(6):919–27. doi: 10.1007/s00259-009-1064-6
40. Xu CC, Biron VL, Puttagunta L, Seikaly H. HPV Status and Second Primary Tumours in Oropharyngeal Squamous Cell Carcinoma. *J Otolaryngol Head Neck Surg* (2013) 42(1):36. doi: 10.1186/1916-0216-42-36
41. Liao CT, Wallace CG, Lee LY, Hsueh C, Lin CY, Fan KH, et al. Clinical Evidence of Field Cancerization in Patients With Oral Cavity Cancer in a Betel Quid Chewing Area. *Oral Oncol* (2014) 50(8):721–31. doi: 10.1016/j.oraloncology.2014.04.010
42. Liao CT, Fan KH, Kang CJ, Lin CY, Chang JTC, Tsang NM, et al. Clinical Outcomes of Patients With Resected Oral Cavity Cancer and Simultaneous Second Primary Malignancies. *PLoS One* (2015) 10(9):e0136918. doi: 10.1371/journal.pone.0136918
43. González-Botas JH, Vázquez PV, Barro CV. Second Primary Tumours in Head and Neck Cancer. *Acta Otorrinolaringol Esp* (2016) 67(3):123–9. doi: 10.1016/j.otoeng.2016.04.005
44. Min SK, Choi SW, Lim J, Park JY, Jung KW, Won YJ. Second Primary Cancers in Patients With Oral Cavity Cancer Included in the Korea Central Cancer Registry. *Oral Oncol* (2019) 95:16–28. doi: 10.1016/j.oraloncology.2019.05.025
45. Bertolini F, Trudu L, Banchelli F, Schipilliti F, Napolitano M, Alberici MP, et al. Second Primary Tumors in Head and Neck Cancer Patients: The Importance of a "Tailored" Surveillance. *Oral Dis* (2021) 27(6):1412–20. doi: 10.1111/odi.13681
46. Milliet F, Bozec A, Schiappa R, Viotti J, Modesto A, Dassonville O, et al. Metachronous Second Primary Neoplasia in Oropharyngeal Cancer Patients: Impact of Tumor HPV Status. A GETTEC Multicentric Study. *Oral Oncol* (2021) 122:105503. doi: 10.1016/j.oraloncology.2021.105503
47. Bugter O, van Iwaarden DLP, van Leeuwen N, Nieboer D, Dronkers EAC, Hardillo JAU, et al. A Cause-Specific Cox Model for Second Primary Tumors in Patients With Head and Neck Cancer: A RONCDOC Study. *Head Neck* (2021) 43(6):1881–9. doi: 10.1002/hed.26666
48. Luo X, Huang X, Liu S, Wang X, Luo J, Xiao J, et al. Evaluation of the Prevalence of Metachronous Second Primary Malignancies in Hypopharyngeal Carcinoma and Their Effect on Outcomes. *Cancer Med* (2022) 11(4):1059–67. doi: 10.1002/cam4.4501
49. Asge Technology Committee, Song LMWK, Adler DG, Conway JD, Diehl DL, Farraye FA, et al. Narrow Band Imaging and Multiband Imaging. *Gastrointest Endosc* (2008) 67(4):581–9. doi: 10.1016/j.gie.2008.01.013
50. Boakye EA, Buchanan P, Hinyard L, Osazuwa-Peters N, Schootman M, Piccirillo JF. Incidence and Risk of Second Primary Malignant Neoplasm After a First Head and Neck Squamous Cell Carcinoma. *JAMA Otolaryngol Head Neck Surg* (2018) 144(8):727–37. doi: 10.1001/jamaoto.2018.0993
51. Mroueh R, Nevala A, Haapaniemi A, Pitkäniemi J, Salo T, Mäkitie AA. Risk of Second Primary Cancer in Oral Squamous Cell Carcinoma. *Head Neck* (2020) 42(8):1848–58. doi: 10.1002/hed.26107
52. Jégu J, Binder-Foucard F, Borel C, Velten M. Trends Over Three Decades of the Risk of Second Primary Cancer Among Patients With Head and Neck Cancer. *Oral Oncol* (2013) 49(1):9–14. doi: 10.1016/j.oraloncology.2012.06.018
53. Jain KS, Sikora AG, Baxi SS, Morris LGT. Synchronous Cancers in Patients With Head and Neck Cancer: Risks in the Era of Human Papillomavirus-Associated Oropharyngeal Cancer. *Cancer* (2013) 119(10):1832–7. doi: 10.1002/cncr.27988
54. Coca-Pelaz A, Rodrigo JP, Suárez C, Nixon IJ, Makitie A, Sanabria A, et al. The Risk of Second Primary Tumors in Head and Neck Cancer: A Systematic Review. *Head Neck* (2020) 42(3):456–66. doi: 10.1002/hed.26016
55. Yamashita T, Araki K, Tomifuji M, Tanaka Y, Harada E, Suzuki T, et al. Clinical Features and Treatment Outcomes of Japanese Head and Neck Cancer Patients With a Second Primary Cancer. *Asia Pac J Clin Oncol* (2017) 13(3):172–8. doi: 10.1111/ajco.12599
56. Morris LGT, Sikora AG, Patel SG, Hayes PB, Ganly I. Second Primary Cancers After an Index Head and Neck Cancer: Subsite-Specific Trends in the Era of Human Papillomavirus-Associated Oropharyngeal Cancer. *J Clin Oncol* (2011) 29(6):739–46. doi: 10.1200/JCO.2010.31.8311
57. Wang CC, Chen ML, Hsu KH, Lee SP, Chen TC, Chang YS, et al. Second Malignant Tumors in Patients With Nasopharyngeal Carcinoma and Their Association With Epstein-Barr Virus. *Int J Cancer* (2000) 87(2):228–31. doi: 10.1002/1097-0215(20000715)87:2<228::AID-IJC12>3.0.CO;2-T
58. Scélo G, Boffetta P, Corbex M, Chia KS, Hemminki K, Friis S, et al. Second Primary Cancers in Patients With Nasopharyngeal Carcinoma: A Pooled Analysis of 13 Cancer Registries. *Cancer Causes Control* (2007) 18(3):269–78. doi: 10.1007/s10552-006-0101-z
59. Ooft ML, van Ipenburg J, Braunius WW, Stegeman I, Wegner I, de Bree R, et al. A Nation-Wide Epidemiological Study on the Risk of Developing Second Malignancies in Patients With Different Histological Subtypes of Nasopharyngeal Carcinoma. *Oral Oncol* (2016) 56:40–6. doi: 10.1016/j.oraloncology.2016.02.009
60. Chow JCH, Tam AHP, Cheung KM, Lee VHF, Chiang CL, Tong M, et al. Second Primary Cancer After Intensity-Modulated Radiotherapy for Nasopharyngeal Carcinoma: A Territory-Wide Study by HKNPCSG. *Oral Oncol* (2020) 111:105012. doi: 10.1016/j.oraloncology.2020.105012
61. Valentin A, Goetz M, Hetzel J, Reinert S, Hoefert S. Routine Panendoscopy in Oral Squamous Cell Cancer Patients: Mandatory or Facultative? *Clin Oral Investig* (2021) 25(3):1245–54. doi: 10.1007/s00784-020-03429-8
62. Onochi K, Shiga H, Takahashi S, Watanabe N, Fukuda S, Ishioka M, et al. Risk Factors Linking Esophageal Squamous Cell Carcinoma With Head and Neck Cancer or Gastric Cancer. *J Clin Gastroenterol* (2019) 53(4):e164–70. doi: 10.1097/MCG.0000000000001019

63. Kato M, Ishihara R, Hamada K, Tonai Y, Yamasaki Y, Matsuura N, et al. Endoscopic Surveillance of Head and Neck Cancer in Patients With Esophageal Squamous Cell Carcinoma. *Endosc Int Open* (2016) 4(7):E752–755. doi: 10.1055/s-0042-106720
64. Wang X, Mauer EA, Christos P, Manzerova J, Wernicke AG, Parashar B. First Clinical Report on Comparative Treatment and Survival Outcomes in Second Cancers After Primary Head and Neck Cancer: A Population-Based Study. *Cureus* (2017) 9(5):e1284. doi: 10.7759/cureus.1284
65. Liao LJ, Chou HW, Wang CT, Chung CS, Lai MS. The Impact of Second Primary Malignancies on Head and Neck Cancer Survivors: A Nationwide Cohort Study. *PLoS One* (2013) 8(4):e62116. doi: 10.1371/journal.pone.0062116
66. Chung CS, Lo WC, Lee YC, Wu MS, Wang HP, Liao LJ. Image-Enhanced Endoscopy for Detection of Second Primary Neoplasm in Patients With Esophageal and Head and Neck Cancer: A Systematic Review and Meta-Analysis. *Head Neck* (2016) 38 Suppl 1:E2343–2349. doi: 10.1002/hed.24277
67. Chung CS, Wang HP. Screening for Precancerous Lesions of Upper Gastrointestinal Tract: From the Endoscopists' Viewpoint. *Gastroenterol Res Pract* (2013) 2013:681439. doi: 10.1155/2013/681439
68. Shimizu Y, Omori T, Yokoyama A, Yoshida T, Hirota J, Ono Y, et al. Endoscopic Diagnosis of Early Squamous Neoplasia of the Esophagus With Iodine Staining: High-Grade Intra-Epithelial Neoplasia Turns Pink Within a Few Minutes. *J Gastroenterol Hepatol* (2008) 23(4):546–50. doi: 10.1111/j.1440-1746.2007.04990.x
69. Oyama T, Inoue H, Arima M, Momma K, Omori T, Ishihara R, et al. Prediction of the Invasion Depth of Superficial Squamous Cell Carcinoma Based on Microvessel Morphology: Magnifying Endoscopic Classification of the Japan Esophageal Society. *Esophagus* (2017) 14(2):105–12. doi: 10.1007/s10388-016-0527-7
70. Chisholm EM, Williams SR, Leung JW, Chung SC, van Hasselt CA, Li AK. Lugol's Iodine Dye-Enhanced Endoscopy in Patients With Cancer of the Oesophagus and Head and Neck. *Eur J Surg Oncol* (1992) 18 (6):550–2.
71. Su HA, Hsiao SW, Hsu YC, Wang LY, Yen HH. Superiority of NBI Endoscopy to PET/CT Scan in Detecting Esophageal Cancer Among Head and Neck Cancer Patients: A Retrospective Cohort Analysis. *BMC Cancer* (2020) 20 (1):69. doi: 10.1186/s12885-020-6558-4
72. Shiozaki H, Tahara H, Kobayashi K, Yano H, Tamura S, Imamoto H , et al. Endoscopic Screening of Early Esophageal Cancer With the Lugol Dye Method in Patients With Head and Neck Cancers. *Cancer* (1990) 66 (10):2068–71. doi: 10.1002/1097-0142(19901115)66:10<2068::AID-CNCR2820661005>3.0.CO;2-W
73. Tincani AJ, Brandalise N, Altemani A, Scanavini RC, Valério JB, Lage HT, et al. Diagnosis of Superficial Esophageal Cancer and Dysplasia Using Endoscopic Screening With a 2% Lugol Dye Solution in Patients With Head and Neck Cancer. *Head Neck* (2000) 22(2):170–4. doi: 10.1002/(sici)1097-0347(200003)22:2<170::aid-hed9>3.0.co;2-7
74. Lee CT, Chang CY, Lee YC, Tai CM, Wang WL, Tseng HP, et al. Narrow-Band Imaging With Magnifying Endoscopy for the Screening of Esophageal Cancer in Patients With Primary Head and Neck Cancers. *Endoscopy* (2010) 42(8):613–9. doi: 10.1055/s-0030-1255514
75. Carvalho R, Areia M, Brito D, Saraiva S, Alves S, Cadime AT. Diagnostic Accuracy of Lugol Chromoendoscopy in the Oesophagus in Patients With Head and Neck Cancer. *Rev Esp Enferm Dig* (2013) 105(2):79–83. doi: 10.4321/s1130-01082013000200004
76. Arantes V, Albuquerque W, Salles JMP, Dias CAF, Alberti LR, Kahaleh M, et al. Effectiveness of Unsedated Transnasal Endoscopy With White-Light, Flexible Spectral Imaging Color Enhancement, and Lugol Staining for Esophageal Cancer Screening in High-Risk Patients. *J Clin Gastroenterol* (2013) 47(4):314–21. doi: 10.1097/MCG.0b013e3182617fc1
77. Laohawiriyakamol S, Sunpaweravong S, Leelamanit V, Pruegsanusak K, Sinkijcharoenchai W. Evaluating Synchronous Esophageal Cancer in Head and Neck Cancer Patients Using Lugol Dye Chromoendoscopy. *J Med Assoc Thai* (2014) 97(11):1164–70. PMID: 25675681.
78. Gong EJ, Kim DH, Ahn JY, Choi KS, Jung KW, Lee JH, et al. Routine Endoscopic Screening for Synchronous Esophageal Neoplasm in Patients With Head and Neck Squamous Cell Carcinoma: A Prospective Study. *Dis Esophagus* (2016) 29(7):752–29. doi: 10.1111/dote.12404
79. Matsui T, Okada T, Kawada K, Okuda M, Ogo T, Nakajima Y, et al. Detection of Second Primary Malignancies of the Esophagus and Hypophraynx in Oral Squamous Cell Carcinoma Patients. *Laryngoscope Investig Otolaryngol* (2018) 3(4):263–7. doi: 10.1002/lio2.179
80. Druesne-Pecollo N, Keita Y, Touvier M, Chan DSM, Norat T, Hercberg S, et al. Alcohol Drinking and Second Primary Cancer Risk in Patients With Upper Aerodigestive Tract Cancers: A Systematic Review and Meta-Analysis of Observational Studies. *Cancer Epidemiol Biomarkers Prev* (2014) 23 (2):324–31. doi: 10.1158/1055-9965.EPI-13-0779
81. Vaca CE, Fang JL, Schweda EK. Studies of the Reaction of Acetaldehyde With Deoxynucleosides. *Chem Biol Interact* (1995) 98(1):51–67. doi: 10.1016/0009-2797(95)03632-V
82. Harada H, Shinohara S, Takebayashi S, Kikuchi M, Fujiwara K, Michida T, et al. Facial Flushing After Alcohol Intake as a Predictor for a High Risk of Synchronous or Metachronous Cancer of the Upper Gastrointestinal Tract. *Jpn J Clin Oncol* (2017) 47(12):1123–8. doi: 10.1093/jjco/hyx150
83. Hoxhaj I, Vukovic V, Boccia S, Pastorino R. Single Nucleotide Polymorphisms and the Risk of Developing a Second Primary Cancer Among Head and Neck Cancer Patients: A Systematic Literature Review and Meta-Analysis. *BMC Cancer* (2021) 21(1):660. doi: 10.1186/s12885-021-08335-0
84. Kaneko K, Katagiri A, Konishi K, Kurahashi T, Ito H, Kumekawa Y, Yamamoto T, et al. Study of P53 Gene Alteration as a Biomarker to Evaluate the Malignant Risk of Lugol-Unstained Lesion With non-Dysplasia in the Oesophagus. *Br J Cancer* (2007) 96(3):492–8. doi: 10.1038/sj.bjc.6603582
85. Wang GQ, Abnet CC, Shen Q, Lewin KL, Sun SD, Roth MJ, et al. Histological Precursors of Oesophageal Squamous Cell Carcinoma: Results From a 13 Year Prospective Follow Up Study in a High Risk Population. *Gut* (2005) 54(2):187–92. doi: 10.1136/gut.2004.046631
86. Aadam AA, Abe S. Endoscopic Submucosal Dissection for Superficial Esophageal Cancer. *Dis Esophagus* (2018) 31(7). doi: 10.1093/dote/doy021
87. Wang WL, Chang IW, Chen CC, Chang CY, Mo LR, Lin JT, et al. Radiofrequency Ablation Versus Endoscopic Submucosal Dissection in Treating Large Early Esophageal Squamous Cell Neoplasia. *Medicine (Baltimore)* (2015) 94(49):e2240. doi: 10.1097/MD.0000`000000002240
88. Lei S, Shrestha SM, Shi R. Radiofrequency Ablation for Early Superficial Flat Esophageal Squamous Cell Neoplasia: A Comprehensive Review. *Gastroenterol Res Pract* (2020) 2020:4152453. doi: 10.1155/2020/4152453
89. Nagami Y, Ominami M, Shiba M, Minamino H, Fukunaga S, Kameda N, et al. The Five-Year Survival Rate After Endoscopic Submucosal Dissection for Superficial Esophageal Squamous Cell Neoplasia. *Dig Liver Dis* (2017) 49 (4):427–33. doi: 10.1016/j.dld.2016.12.009
90. Oda I, Shimizu Y, Yoshio T, Katada C, Yokoyama T, Yano T, et al. Long-Term Outcome of Endoscopic Resection for Intramucosal Esophageal Squamous Cell Cancer: A Secondary Analysis of the Japan Esophageal Cohort Study. *Endoscopy* (2020) 52(11):967–75. doi: 10.1055/a-1185-9329
91. Das A, Singh V, Fleischer DE, Sharma VK. A Comparison of Endoscopic Treatment and Surgery in Early Esophageal Cancer: An Analysis of Surveillance Epidemiology and End Results Data. *Am J Gastroenterol* (2008) 103(6):1340–5. doi: 10.1111/j.1572-0241.2008.01889.x
92. Chung CS, Lo WC, Wen MH, Hsieh CH, Lin YC, Liao LJ. Long Term Outcome of Routine Image-Enhanced Endoscopy in Newly Diagnosed Head and Neck Cancer: A Prospective Study of 145 Patients. *Sci Rep* (2016) 6:29573. doi: 10.1038/srep29573
93. Sarma G, Nath J, Medhi PP, Momin FJ, Kalita AK, Bhattacharyya M. Definitive Radiotherapy in Patients of Synchronous Esophageal and Head and Neck Cancer-Treatment Outcome and Toxicity. *Cancer Treat Res Commun* (2020) 25:100248. doi: 10.1016/j.ctarc.2020.100248

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fonc.2022.895148

Development and Validation of Prognostic Model for Lung Adenocarcinoma Patients Based on m6A Methylation Related Transcriptomics

Huijun Li[1†], Song-Bai Liu[2†], Junjie Shen[1,3], Lu Bai[1,3], Xinyan Zhang[4], Jianping Cao[5], Nengjun Yi[6], Ke Lu[7] and Zaixiang Tang[1,3*]*

[1] Department of Biostatistics, School of Public Health, Medical College of Soochow University, Suzhou, China, [2] Department of Medical Biotechnology, Suzhou Key Laboratory of Medical Biotechnology, Suzhou Vocational Health College, Suzhou, China, [3] Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Medical College of Soochow University, Suzhou, China, [4] School of Data Science and Analytics, Kennesaw State University, Kennesaw, GA, United States, [5] School of Radiation Medicine and Protection and Collaborative Innovation Center of Radiation Medicine of Jiangsu Higher Education Institutions, Soochow University, Suzhou, China, [6] Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, United States, [7] Department of Orthopedics, Affiliated Kunshan Hospital of Jiangsu University, Suzhou, China

Existing studies suggest that m^{6}A methylation is closely related to the prognosis of cancer. We developed three prognostic models based on m^{6}A-related transcriptomics in lung adenocarcinoma patients and performed external validations. The TCGA-LUAD cohort served as the derivation cohort and six GEO data sets as external validation cohorts. The first model (mRNA model) was developed based on m^{6}A-related mRNA. LASSO and stepwise regression were used to screen genes and the prognostic model was developed from multivariate Cox regression model. The second model (lncRNA model) was constructed based on m^{6}A related lncRNAs. The four steps of random survival forest, LASSO, best subset selection and stepwise regression were used to screen genes and develop a Cox regression prognostic model. The third model combined the risk scores of the first two models with clinical variable. Variables were screened by stepwise regression. The mRNA model included 11 predictors. The internal validation C index was 0.736. The lncRNA model has 15 predictors. The internal validation C index was 0.707. The third model combined the risk scores of the first two models with tumor stage. The internal validation C index was 0.794. In validation sets, all C-indexes of models were about 0.6, and three models had good calibration accuracy. Freely online calculator on the web at https://lhj0520.shinyapps.io/LUAD_prediction_model/.

Keywords: prognostic model, lung adenocarcinoma, m^{6}A, immunotherapy, drug prediction.

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Zhixiang Jian, Guangdong Provincial People's Hospital, China
Shen Yin, Memorial Sloan Kettering Cancer Center, United States

***Correspondence:**
Zaixiang Tang
tangzx@suda.edu.cn
Ke Lu
sgu8434@sina.com

†These authors share first authorship

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *13 March 2022*
Accepted: *17 May 2022*
Published: *16 June 2022*

Citation:
Li H, Liu S-B, Shen J, Bai L, Zhang X, Cao J, Yi N, Lu K and Tang Z (2022) Development and Validation of Prognostic Model for Lung Adenocarcinoma Patients Based on m6A Methylation Related Transcriptomics. Front. Oncol. 12:895148. doi: 10.3389/fonc.2022.895148

INTRODUCTION

Lung cancer ranks as the major cause of cancer death, accounting for almost a quarter of cancer deaths (1). Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer, accounting for more than 40% of lung cancer incidence (2).N^6-methyladenosine (m^6A), the most abundant form of posttranscriptional RNA modification in eukaryotes, plays an important role in a variety of biological processes by regulating the translation, processing, splicing, stabilization, and degradation of target RNA (3). The abundance and effects of m^6A methylation modification on RNA are maintained by its methyltransferases ('writers'), binding proteins ('readers'), and demethylases ('erasers') (4).

Existing studies suggest that m^6A methylation is closely related to the prognosis of cancer. An increasing number of m^6A-related genes have been developed as molecular markers of cancer prognosis. In lung adenocarcinoma, several biomarkers have also been developed. Some of the biomarkers are based on single gene model, such as *YTHDC2* (5), *NPM1* (6) and *LCAT3* (7). Some others are multigene-based, including Wang (5 genes) (8), Sun (10 genes) (9), and Zhu (6 genes) (10). Such molecular biomarkers have been shown to enhance the accuracy of overall survival (OS) prediction in LUAD.

However, the predictive power of these markers is often limited. First, most models were constructed based only on mRNAs or lncRNAs. Second, most of the models lack some key parameters, prognostic index or baseline survival function, which make it difficult for others to validate or use them. Further on, even if complete parameters related to model validation are provided (unfortunately, none is found in prediction model related to lung adenocarcinoma at present), few convenient online interaction tools are available.

Based on the above fact, we attempted to develop models to fill in the gaps in prognostic model of lung adenocarcinoma using m^6A-related transcriptomics to predict OS. First, we developed a mRNA prognostic model and a lncRNA prognostic model for lung adenocarcinoma on TCGA cohort and evaluated the two models on several GEO data sets. And then we used the two models and some clinical variables as alternative predictors to construct a multi-omics clinical prediction model. All prediction models developed have two to six independent external validation sets. To further facilitate the practical application of the constructed prediction model in clinical practice, we developed a free online calculator: https://lhj0520.shinyapps.io/LUAD_prediction_model/.

METHODS

Data Acquisition and Processing

For model derivation, we downloaded RNA-seq data (counts values) of 585 LUAD patients (version: 07-20-2019) and corresponding clinical information (version: 08-07-2019) in GDC TCGA from the UCSC Xena public data hub (http://xena.ucsc.edu/). A total of 486 samples with primary tumors and overall survival greater than 30 days were retained. The expression data from the TCGA data portal were quantile normalized and log2-transformed (11). In addition, the somatic mutation data of LUAD patients were also downloaded as a mutation annotation format (MAF) file from GDC Data Portal (https://portal.gdc.cancer.gov/).

For model validation study, 6 datasets from GEO (https://www.ncbi.nlm.nih.gov/geo/) database were considered, including GSE29016 (GPL6947, n=38), GSE29013 (GPL570, n=30) GSE3141 (GPL570, n=58), GSE30219 (GPL570, n=85), GSE37745 (GPL570, n=106), and GSE50081 (GPL570, n=127). We downloaded the series matrix files and their platform annotation information. All the microarray data were quantile normalized and the Robust Multichip Average (RMA) method was used for background adjusted (12).

Annotation of LncRNA Expression

The lncRNAs were extracted according to file downloaded from GENCODE project (https://www.gencodegenes.org/, release 37).

Selection of m^6A Methylation Regulators and m^6A-Related mRNAs

We obtained m^6A methylation regulators from the literature (13). For m^6A-related genes in LUAD, genes annotated as 'protein coding' were retained from the m6AVar database (http://rmvar.renlab.org/) (14), which is a comprehensive database of m^6A-associated variants.

Selection of m^6A-Related LncRNAs

Spearman rank correlation analysis was conducted between m^6A methylation regulatory factors and lncRNAs. Rank correlation coefficient $|R_s|>0.3$ and $P<0.05$ were used as the selection criteria.

Development and Validation of Model Based on mRNAs

Using the mRNA dataset of TCGA LUAD patients as a derivation cohort, we developed a prognostic model to predict OS. As the first step of variable selection, the least absolute shrinkage and selection operator (LASSO) method (15) of R package 'glmnet' was used to reduce the dimension of genes. The optimal value of λ was selected by tenfold cross-validation, and corresponding variables with nonzero coefficients were retained. Next, the "stepAIC" function with "both" in the R package "MASS" was applied to perform stepwise Cox regression (16) for the retained genes, and the optimal gene combination was obtained according to the lowest Akaike information criterion (AIC) value.

Based on the obtained Cox model, the risk score, i.e., prognostic index (PI), could be calculated directly using the "predict" function in R package "rms" with the parameter "type=lp" (17). The calculation formula is as follows:

$$Risk \quad Score \quad (PI) = \left(\sum_{i=1}^{n} \beta_i * Exp_i\right) - \bar{x}, \tag{1}$$

where n refers to the total number of genes in the model; β_i refers to the coefficient of each gene; and Exp_i refers to the expression level of each gene; $\bar{x}$ refers to the mean of PI.

There are two fundamental aspects, discrimination and calibration, to evaluate the performance of the model. Discrimination refers to the ability of a model to differentiate between high-risk patients and low-risk patients (18). It is represented by Harrell's c-index of concordance (C-index) (19). Internal validation adopted bootstrapping (1000 resamples). The C-index was calculated by the "validate" function in the R package "rms" (17). Time-dependent ROC curves at 1-, 3- and 5-year were created by the "survivalROC" R package (20). Through the "cindex" function of the "pec" R package (21), the dynamic time-dependent C-index curve of each dataset was plotted. Calibration refers to the agreement between the predicted and observed survival probabilities (18). The calibration plot was applied to assess the calibration of our model at 1, 3 and 5 years respectively by the "rms" R package (17).

In addition, we estimated the baseline survival function, $S_0(t)$ which is an essential indicator for prediction model (22) and presented it by Kaplan–Meier curves. For the Cox proportional hazards model, the survival probability at different time points are calculated by the following formula (23):

$$S(t|X) = S_0(t)^{exp(PI)}, \quad (2)$$

where S(t|X) denotes the predicted survival at time t for a patient with predictors X; $S_0(t)$ denotes the baseline survival function; and PI denotes the linear predictors. The baseline survival is estimated as , $S_0(t) = exp[-H_0(t)]$ where $H_0(t)$ is the baseline cumulative hazard (22). It can be computed by the "basehaz" function in the "survival" R package (24).

The baseline survival function is crucial, which loads the information needed to evaluate the calibration of survival probabilities in the derivation dataset and more than that calibration in validation datasets (22). Therefore, if we want to validate the Cox model, it is necessary to know the baseline survival function and regression coefficient of the model.

The "surv_cutpoint" function in the R package "survminer" was used to determine the appropriate cutoff value of PI based on the maximum rank statistics (25), and patients in each data set were divided into two risk groups. The predicted survival curve of each person could be calculated by the baseline survival probability. Then, the calibration accuracy of the model can also be evaluated by comparing the average predicted survival probability curve with the observed survival probability curve in the two risk groups (18).

The mRNA model has four GEO external validation sets. Three single data sets included: GSE37745 (n=106), GSE29016 (n=38), and GSE50081 (n=127). Another dataset was pooled by five datasets (GSE3141, GSE29013, GSE30219, GSE37745 and GSE50081). The combined dataset was adjusted for batch effect through the "ComBat" function of the "sva" R package (26). We referred to this combined dataset as the "GSE5total" dataset.

Development and Validation of Model Based on LncRNAs

For lncRNA model, we used four steps to obtain appropriate lncRNAs. First, the random survival forest (RSF) (27), a machine learning method for regression, was used to conduct preliminary feature screening for m^6A-related lncRNAs through "rfsrc" function of "randomForestSRC" R package (28). This algorithm was used to rank prognostic lncRNAs (ntree =1000), and we selected the top 100 lncRNAs for the next step of selection. Second, we applied LASSO to shrink variables. Then, the prognostic factors retained by the LASSO algorithm were analyzed by best subset selection. To realize this method in the Cox proportional hazards model, we used the R package "BeSS" (29). Finally, stepwise Cox regression was used to select the optimal model from the factors obtained in the previous step.

The performance evaluation and PI calculation methods of lncRNA model were the same as mRNA model.

Two datasets, GSE30219 (n=85) and GSE50081 (n=127), were used to validate the lncRNA model. For expanding the sample size of the validation set, we combined the above two data sets into one data set and named it "GSE2total" to validate.

Development and Validation of Comprehensive Prediction Model

To further expand the clinical prediction capacity of m^6A-related model, we decided to develop a more comprehensive clinical prediction model (we called it the "comprehensive prediction model") by combining two risk scores obtained from the above models with clinical variables.

We used multiple imputation by chained equations of the R package "mice" to impute the missing values of clinical variables (5 times) (30). The number of iterations in each imputation was five by default. The variables used in the multiple imputation model included the two risk scores(mRNA risk score and lncRNA risk score), three clinical factors that were common in the derivation and validation datasets (age, sex and tumor stage) and the outcome (the Nelson–Aalen estimator of the baseline cumulative hazard and the outcome indicator) (31, 32). For 5 imputed data sets, we put each imputed set below each other into a stacked data with a weight of 1/5 per patient (5 means number of imputation) (33).

The predictive factors in the multivariate Cox regression model were screened by stepwise regression. The performance evaluation and PI calculation methods of this model were still the same as mRNA model. Two datasets from GEO database, GSE37745 (n=106) and GSE50081 (n=127), were used to validate.

Somatic Mutation Analysis

The "maftools "R package was used to analyze TCGA somatic mutation data (34).

Immunotherapy Exploration of the Model

Immune checkpoints, negative regulators of immune activation, can downregulate the immune state of the body and limit antitumor responses (35, 36). Tumor Immune Dysfunction and Rejection (TIDE) is a computational framework developed to assess the potential of tumor immune escape from gene-expressed cancer samples and to measure the responsiveness of immune checkpoint inhibitors (37, 38). TIDE scores were calculated for each of 486 LUAD patients by the TIDE website (http://tide.dfci.harvard.edu/).

Drug Prediction

By using the "calcPhenotype" function of the R package "oncoPredict" (39) and the database resources of Genomics of Drug Sensitivity in Cancer (GDSC) V2 as development data, six commonly used chemotherapy drugs (paclitaxel, fluorouracil, cisplatin, vinorelbine, gemcitabine, and docetaxel) were used for analysis, and the half-maximum inhibitory concentration (IC50) of each drug was estimated in every sample.

Statistical Analysis

All statistical analyses were performed using R (version 4.1.0). A bivariate normal distribution test was performed on the data requiring correlation analysis. The Shapiro-Wilk test and Bartlett's test of homogeneity of variances were performed on the data requiring comparison between groups. Student's t test was used if the continuous variable was normally distributed, and the Wilcoxon rank sum test was used if the continuous variable was not normally distributed. $P < 0.05$ was considered statistically significant. Median follow-up time was calculated by reverse Kaplan-Meier method (40). The survival curves were analyzed using log-rank test.

RESULTS

Patient Cohorts

The design and workflow of the models constructed in this study are shown in **Figure 1**. The patient characteristics are summarized in **Table 1**. For the derivation cohort, a total of 486 patients had 175 deaths and an event rate of 36%, with a median overall survival of 2.4 years (95%CI: 2.2-2.8).

In the comprehensive prediction model, the number of events per variable in derivation model was 35 (175/5), indicating a reasonable number of events compared to the number of candidate predictors. This quantity meets the EPV principle required by the sample size of the prediction model, that is, there should be at least ten events per variable (23). We observed only a slight percentage of missing values for age and tumor stage

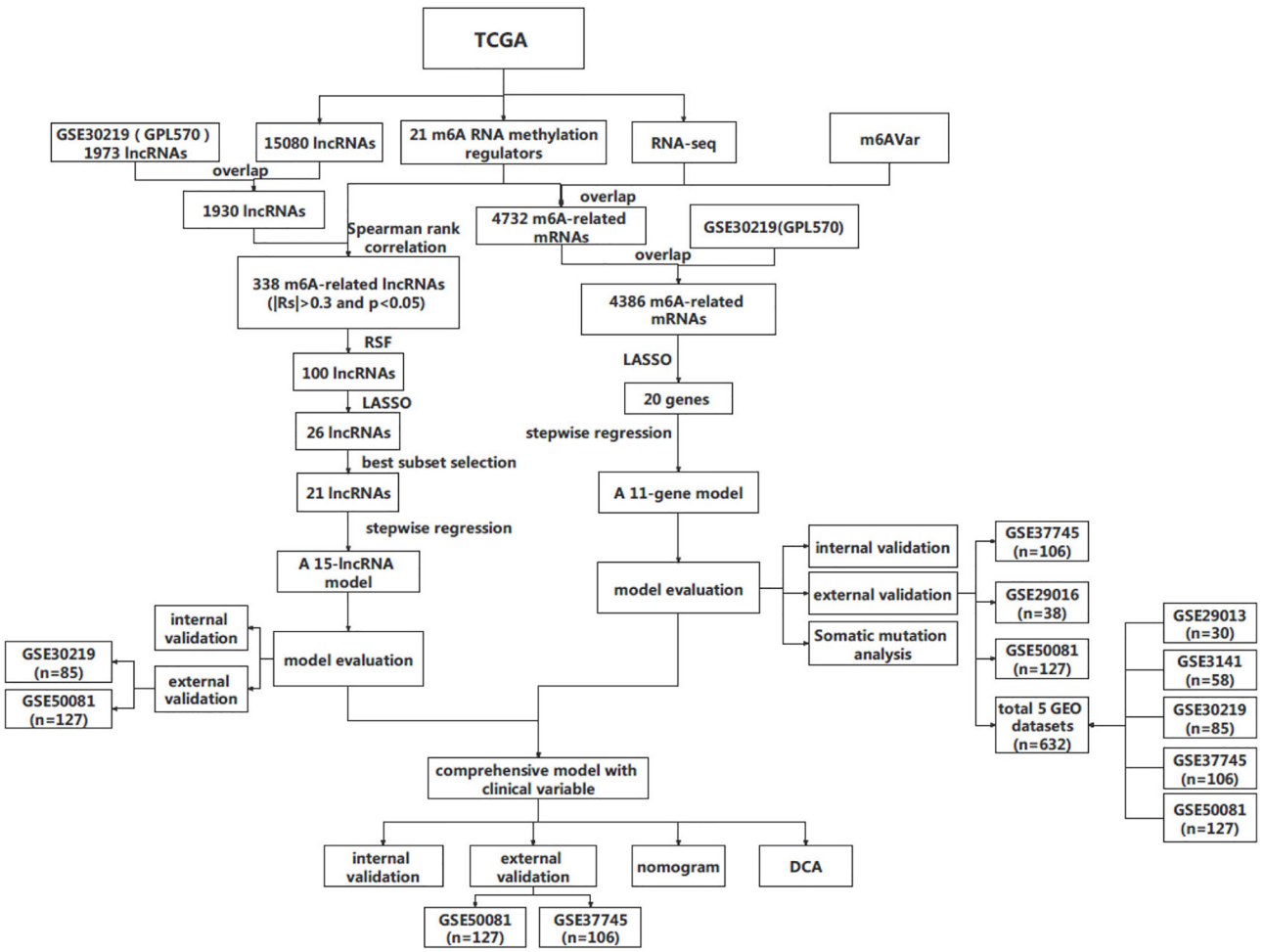


FIGURE 1 | The workflow of this study. RSF, random survival forest; DCA, decision curve analysis.

TABLE 1 | Patient characteristics.

Characteristic	Derivation Cohort TCGA (n=486)	Validation Cohorts					
		GSE29016 (n=38)	GSE30219 (n=85)	GSE37745 (n=106)	GSE50081 (n=127)	GSE5 total (n=406)	GSE2 total (n=212)
Age, year (IQR)	66.0 (59.0,72.0)	69.0 (59.0,73.0)	60.0 (55.0,69.0)	64.0 (55.0,70.0)	69.9 (62.8,75.7)	–	–
Missing values, n (%)	10 (2.0)	–	–	–	–	–	–
Gender (%)							
Female	261 (54)	20 (53)	19 (22.4)	60 (56.6)	62 (48.8)	–	–
Male	225 (46)	18 (47)	66 (77.6)	46 (43.4)	65 (51.2)	–	–
Tumor stage (%)							
Stage I	261 (53.7)	29 (76)	–	70 (66.0)	92 (72.4)	–	–
Stage II	114 (23.4)	6 (16)	–	19 (17.9)	35 (27.6)	–	–
Stage III	79 (16.2)	2 (5.3)	–	13 (12.3)	–	–	–
Stage IV	25 (5.1)	–	–	4 (3.8)	–	–	–
Missing values, n (%)	7 (0.01)	1 (2.6)	–	–	–	–	–
Follow-up time, years (95%CI)	2.4 (2.2, 2.8)	11.8 (11.4, 13.4)	9.7 (8.3,11.2)	10.5 (9.2,13.0)	5.5 (5.2,6.0)	6.2 (5.8, 6.8)	6.2 (5.8,6.8)
Death events (%)	175 (36)	28 (73.7)	45 (52.9)	77 (72.6)	51 (40.1)	213 (52.5)	96 (45.3)

in the TCGA cohort, 2.1% and 1.4%, respectively (**Figure S1A**). **Figure S1B** shows that the missing values of the data variables correspond to random missing values (41). All 486 patients who met the requirements for the development data were included in the model after imputation.

Figure 2 shows the survival curves (**Figures 2A–C**) and baseline survival probability curves (**Figures 2D–F**) of each data set in the three models.

Development and Validation of the mRNA Model

The 21 m^6A regulatory factors extracted from the literature are listed in **Table S1**. Common genes obtained from the three data sets m6AVar, TCGA and GSE30219 and 21 regulatory factors were included; finally, we obtained 4386 mRNAs related to m^6A (**Figure 3A**).

These genes were screened by LASSO (**Figure 3B**) and stepwise regression successively, and a prediction model containing 11 mRNAs associated with OS was obtained (CASC3, USP4, CTCFL, SETDB2, MARCH4, KIRREL3, GRIK2, EIF2AK3, SNTG2, LINGO2 and ZNF708). **Figure 3C** shows the coefficients of the model visually. Based on the genes and coefficients in the development data set, PI was constructed as follows:

$$PI = -0.46605 \times CASC3 - 0.64556 \times USP4 + 0.11549 \times \text{CTCFL} - 0.34872 \times \text{SETDB2} + 0.09105 \times \text{MARCH4} - 0.16502 \times \text{KIRREL3} + 0.12956 \times \text{GRIK2} - 0.60740 \times \text{EIF2AK3} - 0.15933 \times \text{SNTG2} + 0.06450 \times \text{LINGO2} - 0.24452 \times \text{ZNF708} + 23.20828$$

The distribution of PI in the derivation and validation data sets were shown in **Figure S2A**. The base survival probability of the mRNA model from 1 to 10 years was given in **Table S2**. By substituting the calculated PI and the basic survival probability at different time points into formula (2), the prognostic survival probability of individual at corresponding time points can be obtained

In internal validation, the apparent C-index of the model was 0.751(95%CI:0.711-0.791), and the optimism-corrected C-index with 1000 bootstrap resamples was 0.736. The 1-year, 3-year and 5-year AUCs of the model were 0.768, 0.788, and 0.756, respectively (**Figure 4A**). The calibration plot shows that the model has good agreement between predicted and observed survival probabilities at 1, 3 and 5 years (**Figure 4B**). In addition, patients were divided into two risk groups based on the optimal cutoff value of PI (**Figure 4C**). In **Figure 4D**, the observed Kaplan–Meier survival curves (the solid line) were close to the average predicted survival curves (the dotted line) in the two risk groups, which also proved that our prediction model had good calibration accuracy. **Figure S3** shows the Kaplan–Meier survival curves (**Figure S3A**) and risk factor association diagrams (**Figure S3B**) for the two risk groups.

In the external validation cohorts, C indexes of the model were acceptable, which were 0.598(95%CI:0.511-0.685) (GSE50081), 0.608(95%CI:0.510-0.707)(GSE29016), 0.634(95% CI:0.571-0.697)(GSE37745) and 0.608(95%CI:0.567-0.649) (GSE5total). In addition, **Figure S4A** shows C-indexes of the model over 1-10 years in all datasets. According to the time-dependent ROC curves (**Figure 5**), the area under the curves of the model in the four validation sets of 1, 3 and 5 years were all above 0.6, which also indicated that its discriminative ability is satisfactory. The calibration diagrams from the four validation sets show the good calibration accuracy of the model in external validation (**Figure 6**). Patients in the validation sets were divided into two risk groups based on the maximum rank statistics (**Figures S5A–D**), and the average predicted survival curves (the dotted line) and observed survival curves (the solid line) of the two groups were compared to further verify the calibration accuracy of the model (**Figures S5E–H**). The long-term prediction ability of the model in the GSE50081 (**Figure S5E**) and GSE29016 (**Figure S5F**) datasets was not as good as that in the other two datasets (**Figures S5G–H**). However, within 5 years, the calibration accuracy of the model is acceptable. Subsequently, the Kaplan–Meier survival curves of the two risk

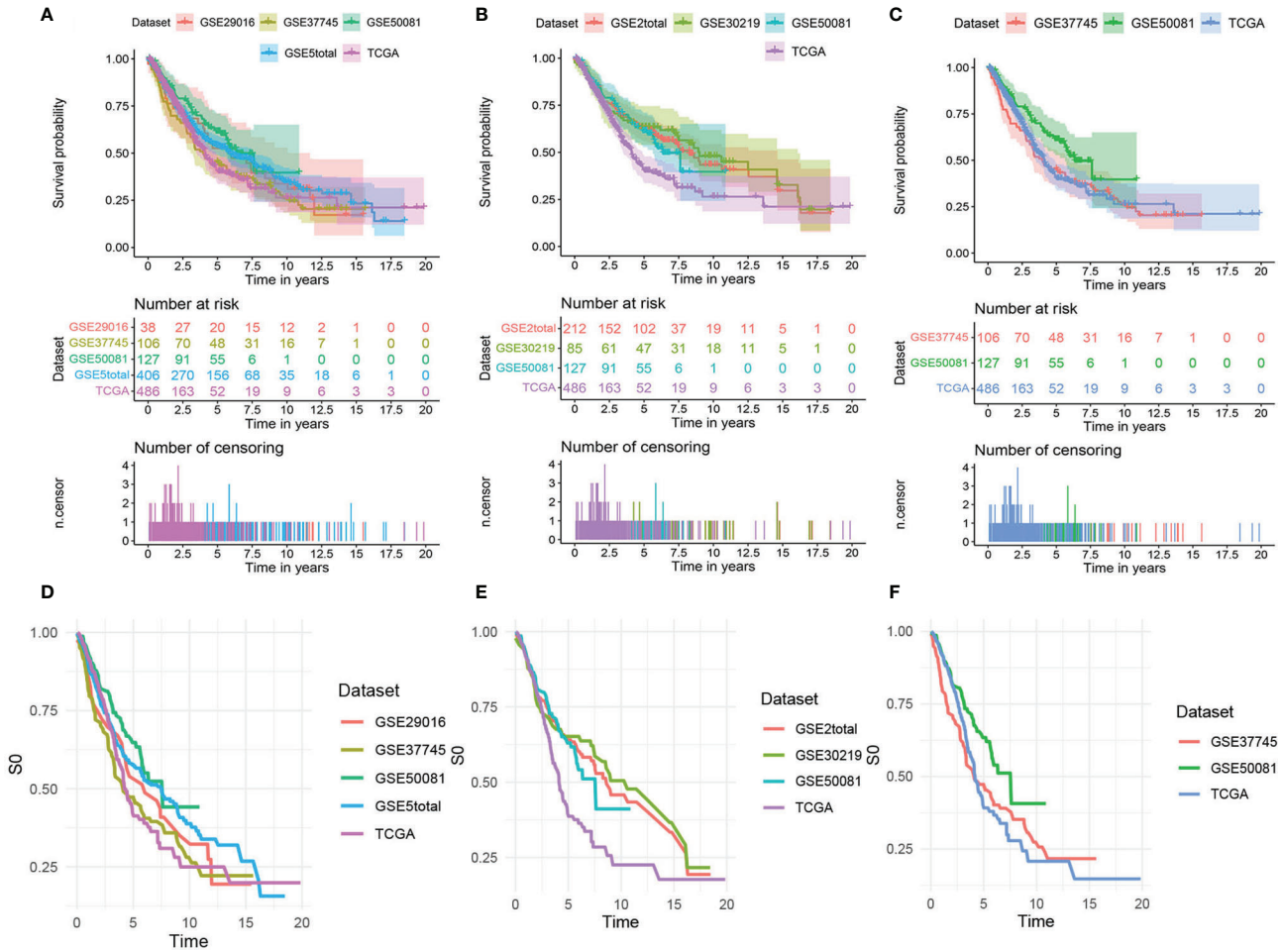


FIGURE 2 | The survival curves and baseline survival probability curves of each data set in the three models. The survival curves of each data set in **(A)** the mRNA model, **(B)** the lncRNA model, and **(C)** the comprehensive clinical model. The baseline survival probability curves of each data set in **(D)** the mRNA model, **(E)** the lncRNA model, and **(F)** the comprehensive clinical model.

groups and the risk factor association diagrams of the model in each validation set are shown in **Figures S6, S7** respectively.

Development and Validation of the lncRNA Model

First, 1930 common lncRNAs of TCGA and GSE30219 data sets were obtained (**Figure 7A**). Then, genes were screened by the importance score of random survival forest (**Figure 7B**), and the top 100 genes were reserved for the next step. Twenty-six genes were obtained by LASSO screening of 100 reserved genes (**Figure 7C**). Next, we selected the best subset selection method for further screening of genes and obtained 21 genes (**Figure 7D**). Finally, 15 lncRNAs of the prediction model associated with OS were obtained by stepwise regression (*SNHG12, RPARP-AS1, CRNDE, LMO7DN, AC008467.1, LINC00639, AC107464.1, AL445931.1, FLG-AS1, C5orf66, AC026250.1, AC245595.1, LINC01933, LINC01137, RUSC1-AS1*). Furthermore, the co-expression networks of 21 m^6A and 1930 lncRNAs were visualized by a Sankey diagram, as shown in **Figure 8A**. In addition, the heatmap of the correlation between 21 m^6A genes and 15 lncRNAs in the model is shown in **Figure 8B**.

Based on the genes and coefficients in the development data set, PI was constructed as follows:

$$\begin{aligned} PI = &-0.17135 \times LM07DN - 0.33117 \times SNHG12 + \\ &0.14349 \times \text{C5orf66} + 0.41125 \times (\text{RUSC1} - \text{AS1}) + 0.16394 \times \\ &\text{AC245595.1} + 0.27029 \times \text{LINC01137} + 0.10490 \times \\ &\text{AL445931.1} + 0.11064 \times (\text{FLG} - \text{AS1}) - 0.10828 \times \\ &\text{AC107464.1} + 0.15101 \times \text{AC026250.1} - 0.17919 \times \\ &\text{CRNDE} - 0.15018 \times \text{AC008467.1} + 0.22517 \times \\ &\text{LINCO1933} - 0.11297 \times \text{LINC00639} - 0.25657 \times \\ &(\text{PRARP} - \text{AS1}) - 0.07307 \end{aligned}$$

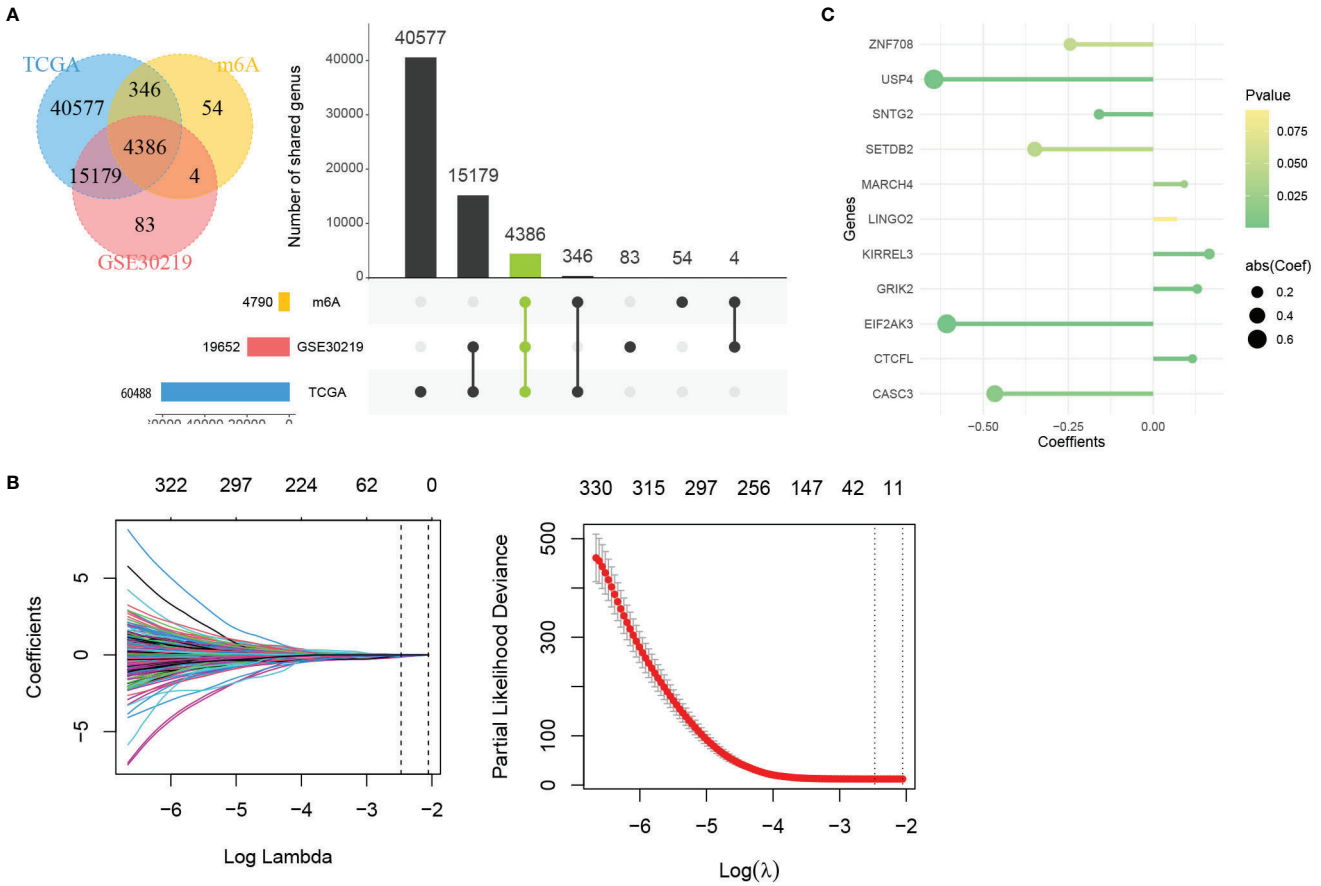


FIGURE 3 | Identification of genes in mRNA model. **(A)** Venn plot of 4386 mRNAs related to m^{6}A. **(B)** LASSO shrinking path diagram. **(C)** The coefficients of 11 genes in the model.

Figure 8C shows the coefficients of the model visually. The distribution of PI in the development data set and validation set is shown in **Figure S2B**. The base survival probability of the lncRNA model from 1 to 10 years is given in **Table S2**

In internal validation, the apparent C-index was 0.730(95% CI:0.688-0.772), and the optimism-corrected C-index with 1000 bootstrap replications was 0.707. The AUCs of the model at 1, 3 and 5 years were 0.754, 0.796, and 0.751, respectively (**Figure 9A**). The calibration plot shows that the model has good agreement between predicted and observed survival probabilities at 1, 3 and 5 years (**Figure 9B**). Furthermore, patients were divided into two risk groups based on the optimal truncation value of PI (**Figure 9C**). It was further found that the observed Kaplan–Meier survival curves in the two risk groups were close to the average predicted survival curves (**Figure 9D**), which also proved that our prediction model had good calibration accuracy. **Figure S8** shows the Kaplan–Meier survival curves (**Figure S8A**) and risk factor association diagrams (**Figure S8B**) for the two risk groups.

In the external validation cohorts, three C indexes of the model were 0.596(95%CI:0.506-0.685)(GSE50081), 0.602(95% CI:0.525-0.682)(GSE30219) and 0.596(95%CI:0.534-0.658) (GSE2total). In addition, **Figure S4B** shows C-indexes of the model over 1-10 years in four datasets. Although C-indexes of the model in the validation set is lower than derivation set, they remained at 0.6 during the decade. According to the time-dependent ROC curves (**Figures 10A–C**), the area under the curves of the model in the three validation sets of 1, 3 and 5 years were all above 0.6. **Figures 10D–F** shows the calibration accuracy of the model in three external verification sets. Patients in the validation sets were divided into two risk groups based on the maximum rank statistics (**Figures S9A–C**), and the average predicted survival curves and observed survival curves in the two groups were compared to further validate the calibration accuracy of the model (**Figures S9D–F**). Unfortunately, the external validation calibration accuracy of the lncRNA model was not as ideal as that of mRNA model, but the prediction results within three years were close to the observations and did not deviate too far from reality within five years. Subsequently, the Kaplan–Meier survival curves of the two risk groups and the risk factor association diagrams of the model in each validation set are shown in **Figures S10, S11** respectively.

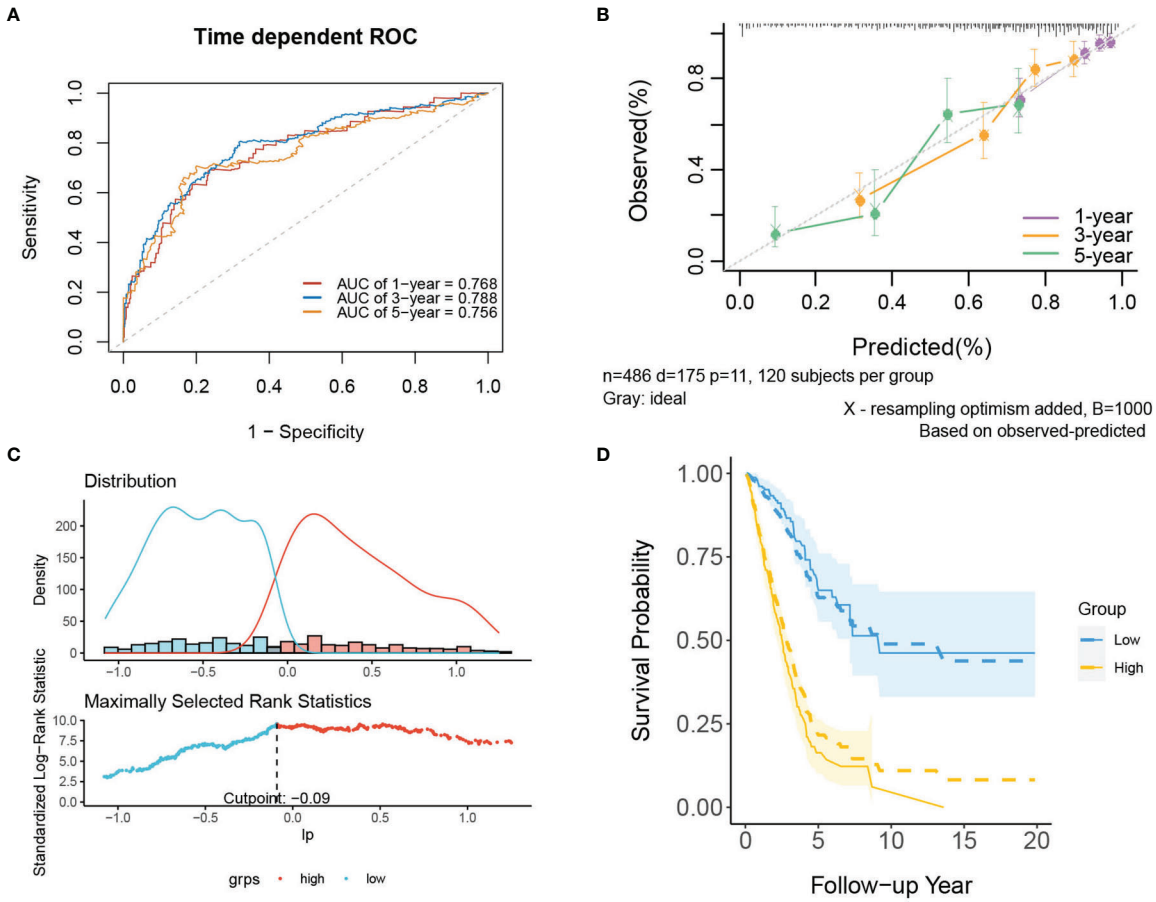


FIGURE 4 | The performance of the mRNA model in the derivation dataset. **(A)** 1-,3-,5-year ROC curves and **(B)** calibration plot of the mRNA model. **(C)** The optimal cutoff value of PI. **(D)** Predicted versus observed survival probability in per risk group. Solid line: observed Kaplan-Meier curve; dotted line: average predicted survival curve; shaded area: 95% confidence interval of observed survival probability.

Development and Validation of the Comprehensive Prediction Model

The prognostic indexes of the two gene models were used as candidate predictors, and the comprehensive prediction model was constructed by stepwise regression combined with three clinical variables (age, sex and tumor stage) to predict OS. The final model included three predictors: mRNA risk score, lncRNA risk score, and tumor stage. Based on the coefficients and predictors obtained from all imputed datasets, the final PI is structured as:

$$PI = -0.3295 + 0.6015 \times \text{mRNA Risk Score} + 0.4540 \times \text{lncRNA Risk Score} + \text{tumor stage}$$

in which:

Tumor stage: stage I=0, stage II= 0.6567, stage III= 0.7510, stage IV= 0.9675

The distribution of PI in the development data set and validation set is shown in **Figure S2C**. The base survival probability of the comprehensive prediction model from 1 to 10 years is also given in **Table S2**.

In internal validation, the apparent C-index was 0.795(95% CI:0.780-0.810) the optimism-corrected C-index with 1000 bootstrap replications was 0.794. The 1-year, 3-year and 5-year AUCs of this model were 0.824, 0.847, and 0.809, respectively (**Figure 11A**). The calibration plot shows that the model has good agreement between predicted and observed survival probabilities at 1, 3 and 5 years (**Figure 11B**). Again, patients were divided into two risk groups based on the optimal truncation value of PI (**Figure 11C**). The observed Kaplan–Meier survival curves for the two risk groups almost overlap with the average predicted survival curves shown in **Figure 11D**, further confirming that the prediction model has good calibration accuracy in the derivation set.

There are two data sets used as external validation sets for this model. In the external validation cohorts, the two C indexes of the model were 0.649(95%CI:0.564-0.733)(GSE50081) and 0.606 (95%CI:0.536-0.677) (GSE37745). **Figure S4C** shows the C-index of the model over 1-10 years in the three datasets. **Figures 12A–B** shows the ROC curve of the model in the two validation sets, and **Figures 12C–D** shows the calibration plots.

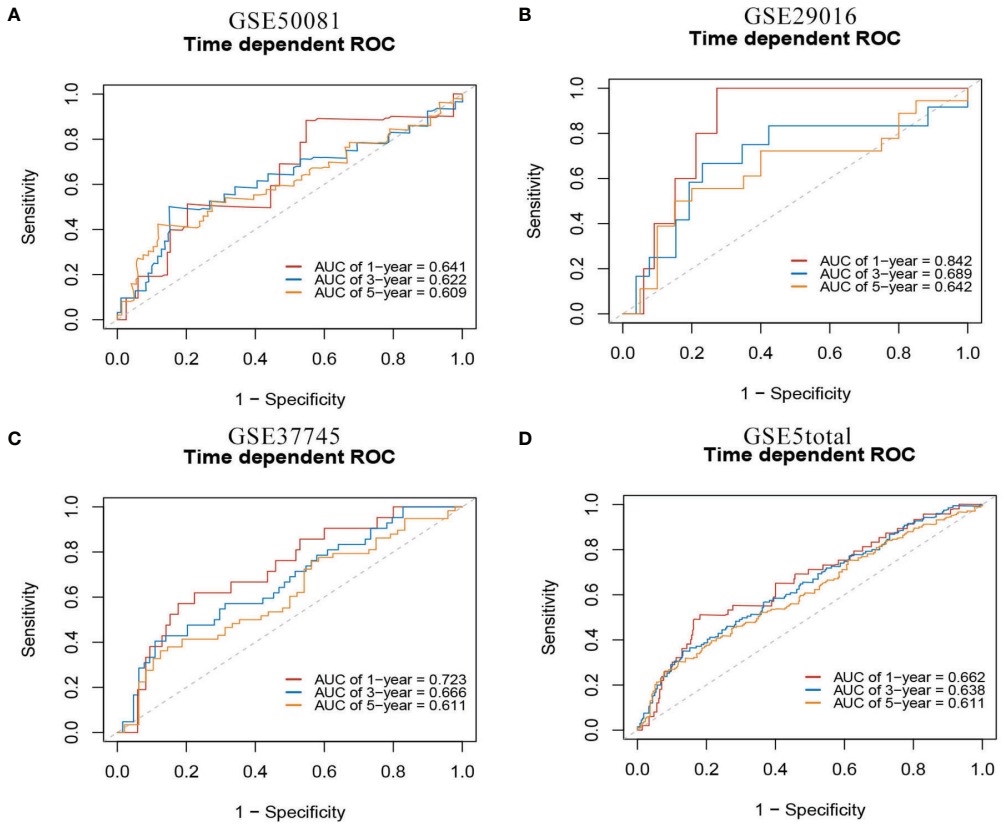


FIGURE 5 | 1-,3-,5-year ROC curves of mRNA model in external validation data sets. **(A)** GSE50081 dataset. **(B)** GSE29016 dataset. **(C)** GSE37745 dataset. **(D)** GSE5total dataset.

Again, we divided samples into two risk groups (**Figures S12A, B**) and then compared the observed survival curves in the two risk groups with the average predicted survival curves (**Figures S12C, D**). In GSE50081, the model still has the risk of underestimating the survival probability (**Figure S12C**). However, in GSE37745, the predicted average survival probability curves were quite close to the actual curve, showing very good consistency (**Figure S12D).**

From this model, we created a nomogram to predict the prognostic survival probability of patients with lung adenocarcinoma at 1, 3 and 5 years (**Figure S13**). Subsequently, we used decision curve analysis (DCA) to compare and demonstrate the net benefits of the clinical utility of the three models at 1, 3 and 5 years (**Figure S14**). With increasing time, the net benefits of the three models continued to increase, and the net benefits of the mRNA model and lncRNA model at the three time points showed little difference. As a matter of course, the net benefit of the comprehensive model is always the greatest.

Online Calculators for Models

To facilitate the clinical application of the model, the three model calculations mentioned in this paper can be completed by this website: https://lhj0520.shinyapps.io/LUAD_prediction_model/. Enter or select the value of the variable and the time you want to predict in the gray box on the left side of the page and then click the "forecast" button at the bottom to obtain the corresponding point estimate or survival curve on the right side (**Figure S15**).

Drug Prediction and TIDE Immunotherapy Prediction Analyses

Chemotherapy plays a critical role in curing or controlling lung adenocarcinoma. The IC50 estimates of 6 common chemotherapeutic drugs were calculated from the GDSC database. The difference of IC50 between the high and low risk groups in the mRNA model was compared. The results (**Figure 13A**) showed that the IC50 values of all 6 drugs were significantly different between the high-risk group and the low-risk group, and patients in the low-risk group were more sensitive than the high-risk group.

Immunotherapy using immune checkpoint inhibitors has brought hope to LUAD patients. The response of 486 patients in the TCGA dataset to immune checkpoint inhibitors was calculated based on the gene expression matrix through the TIDE website. As shown in **Figure 13B**, for the mRNA model, the risk score of patients in the nonresponse group (n=259) was higher than that in the response group (n=227), and the difference was statistically significant (Wilcoxon test, p=0.002). Further

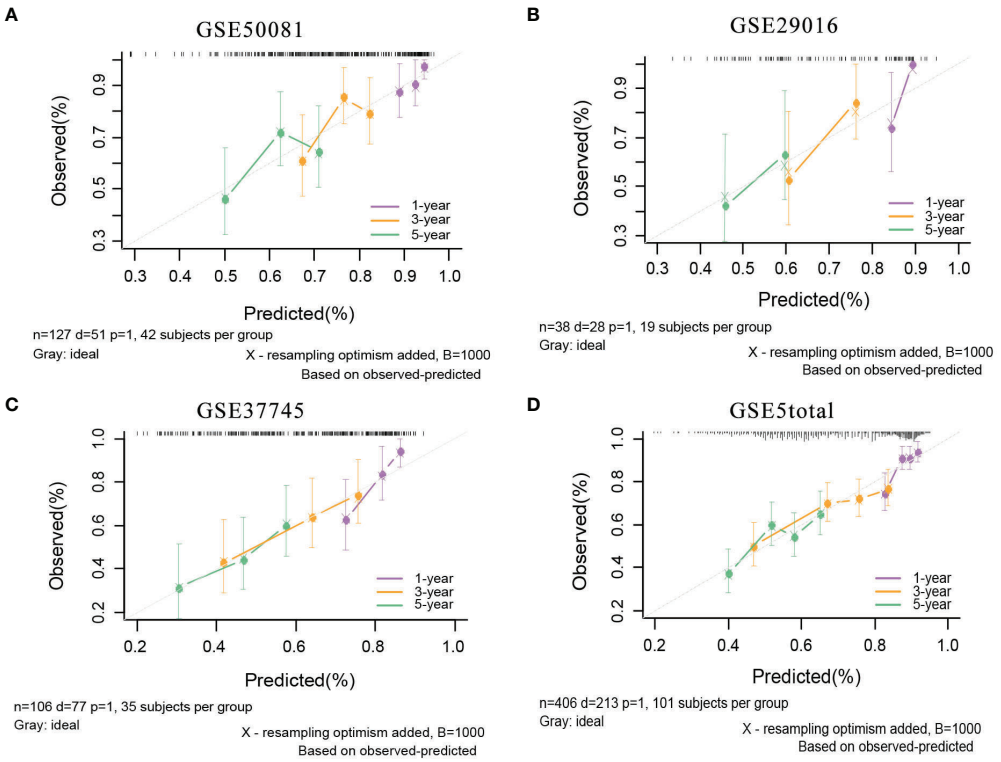


FIGURE 6 | The calibration plots of mRNA model in external validation data sets. **(A)** GSE50081 dataset. **(B)** GSE29016 dataset. **(C)** GSE37745 dataset. **(D)** GSE5total dataset.

analysis (**Figure 13C**) showed that patients in the low-risk group (127/245) were more sensitive to immunotherapy than those in the high-risk group (100/241). In contrast, in the lncRNA model, the difference was not statistically significant (Wilcoxon test, p=0.095), so it could not be considered that there was a difference in risk scores between the two groups (**Figure 13D**).

Study of Somatic Variation in the mRNA Model

We obtained single nucleotide mutations data for 476 LUAD patients (ten samples were not available) from the GDC Data Portal. **Figure 14A** is a summary of the mutation data. More detailed mutation information is shown in **Figure 14B**. Different colors represent different types of mutations. In addition, we compared the mutations in genes in the mRNA model between the two risk groups (**Figure 14C**). *GRIK2* was found to be the mutated gene with the most common frequency in both groups, which mutated more in the high-risk group (**Figure 14D**). More intriguingly, we calculated co-occurrence and mutually exclusive mutations between 11 genes and found only two group co-occurrence mutations, including *GRIK2*(**Figure S16A**). Subsequently, we plotted the mutation frequency of genes into gene word clouds, as shown in **Figure S16B**. Further, we calculated the tumor mutation burden (TMB) in 476 samples (**Figure S16C**). We compared the TMB of the responder and non-responder groups in TIDE. The TMB of the responder group was higher than that of the non-responder group (Wilcoxon test, p=0.028, **Figure S16D**), indicating that patients with higher TMB may have a better effect on immunotherapy.

DISCUSSION

Commonly used predictive models for lung adenocarcinoma based on m^{6}A methylated relevant genes have been developed, but these models are not yet complete in terms of application. This study constructed clinical prediction models at three different levels based on m^{6}A-related mRNAs, lncRNAs and clinical information data, and collected multiple external validation sets for validation. We reported this study according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Statement (TRIPOD). The complete checklist is shown in **Table S3**.

The first model was developed based on m^{6}A-related mRNA and contained 11 genes in total (**Table S4**). Compared with other models, our model contains more genes. However, in several independent external validation sets, the model shows relatively stable and good discrimination and calibration. At present, studies have shown that *USP4*, *EIF2AK3* and *CTCFL* genes are related to the prognosis of lung adenocarcinoma (42–44).

The 11 genes are all obtained from m6Avar database (now updated to "RMVar"). Variants of these genes were hypothesized

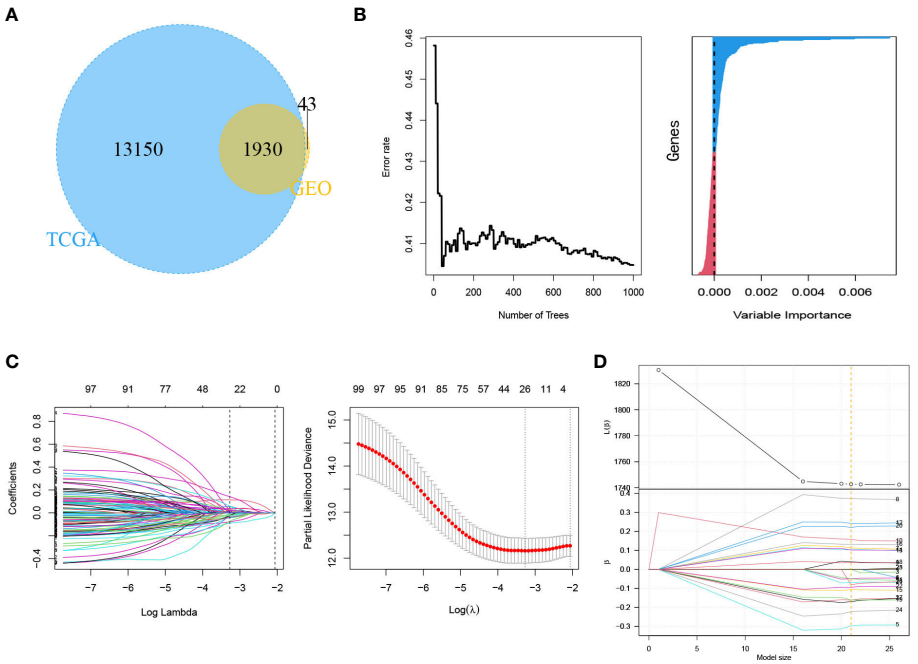


FIGURE 7 | Identification of genes in lncRNA model. **(A)** Venn plot of 1930 lncRNAs related to m^{6}A. **(B)** Random survival forest analysis. **(C)** LASSO shrinking path diagram. **(D)** The coefficient profile plot of the coefficient and loss paths for best subset selection.

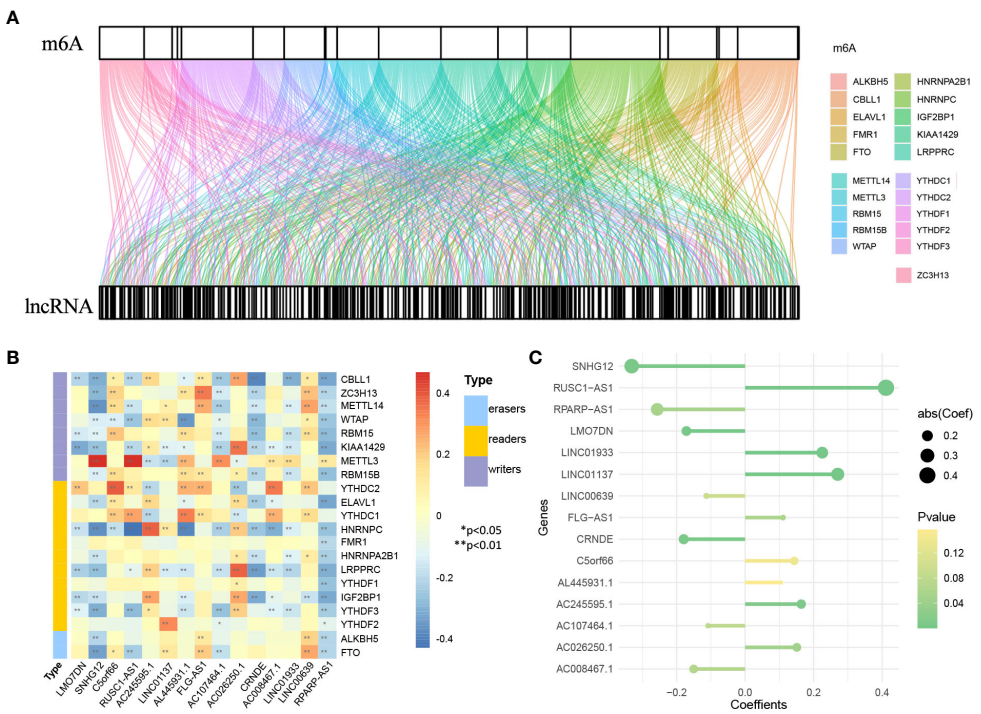


FIGURE 8 | Identification of genes in lncRNA model. **(A)** Sankey diagram of 21 m^{6}A regulators and 1930 m^{6}A-related lncRNAs. **(B)** The heatmap for the correlation between 21 m^{6}A genes and 15 prognostic m^{6}A-related lncRNAs. **(C)** The coefficients of 15 lncRNAs in the model.

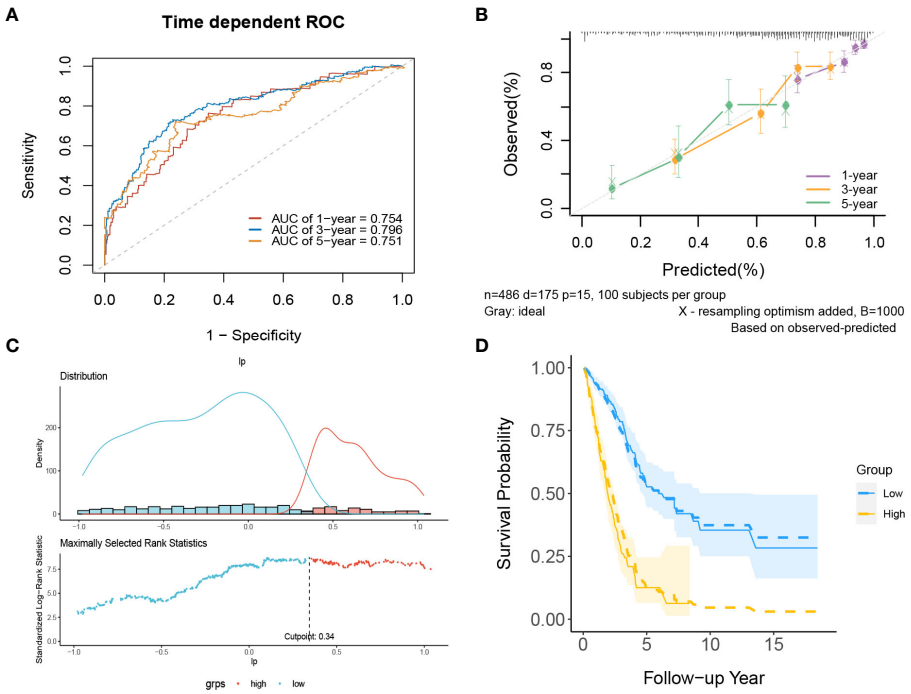


FIGURE 9 | The performance of the lncRNA model in the derivation dataset. **(A)** 1-,3-,5-year ROC curves and **(B)** calibration plot of the lncRNA model. **(C)** The optimal cutoff value of PI. **(D)** Predicted versus observed survival probability in each risk group. Solid line: observed Kaplan-Meier curve; dotted line: average predicted survival curve; shaded area: 95% confidence interval of observed survival probability.

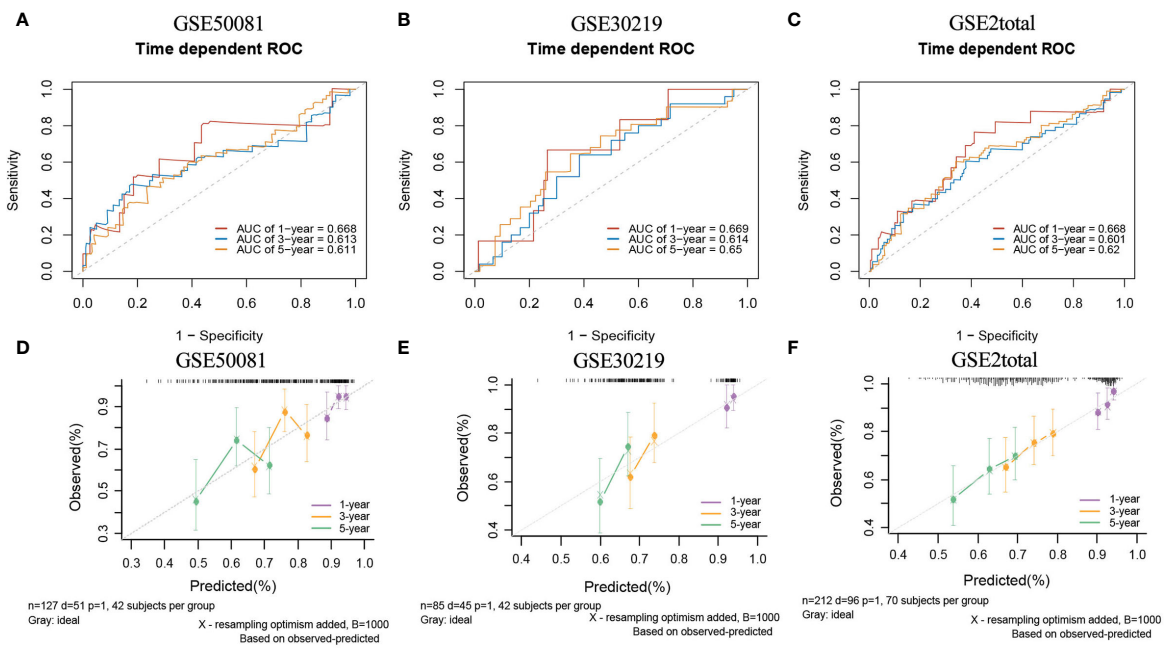


FIGURE 10 | The ROC curves and calibration plots of lncRNA model in external validation data sets. ROC curves at 1-,3-,5-year: **(A)** GSE50081 dataset, **(B)** GSE30219 dataset, and **(C)** GSE2total dataset. The calibration plots at 1-,3-,5-year: **(D)** GSE50081 dataset, **(E)** GSE30219 dataset, and **(F)** GSE2total dataset.

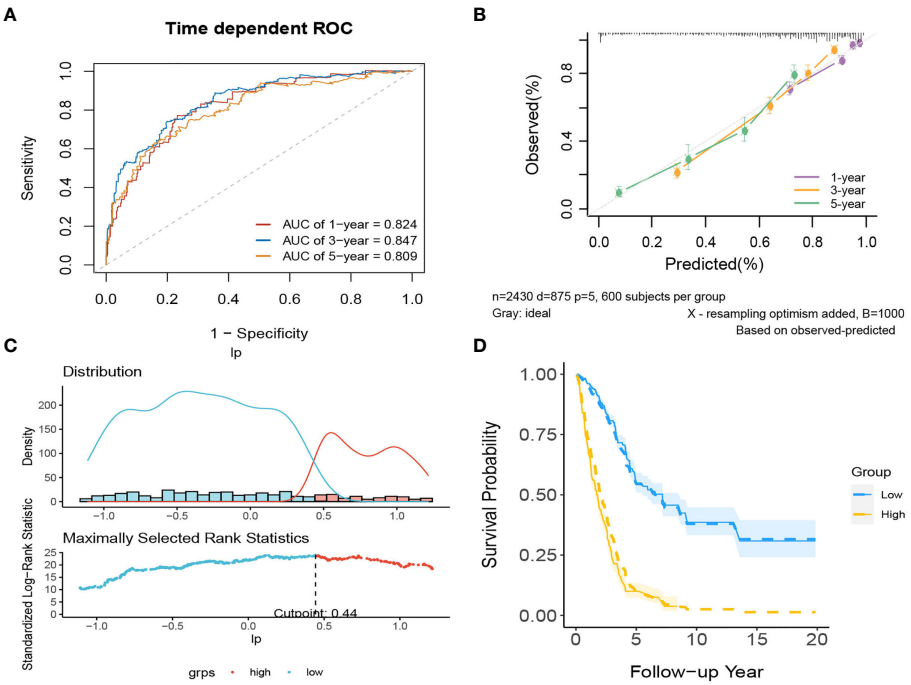


FIGURE 11 | The performance of the comprehensive model in the derivation dataset. **(A)** 1-,3-,5-year ROC curves and **(B)** calibration plot of the comprehensive model. **(C)** The optimal cutoff value of PI. **(D)** Predicted versus observed survival probability per risk group.

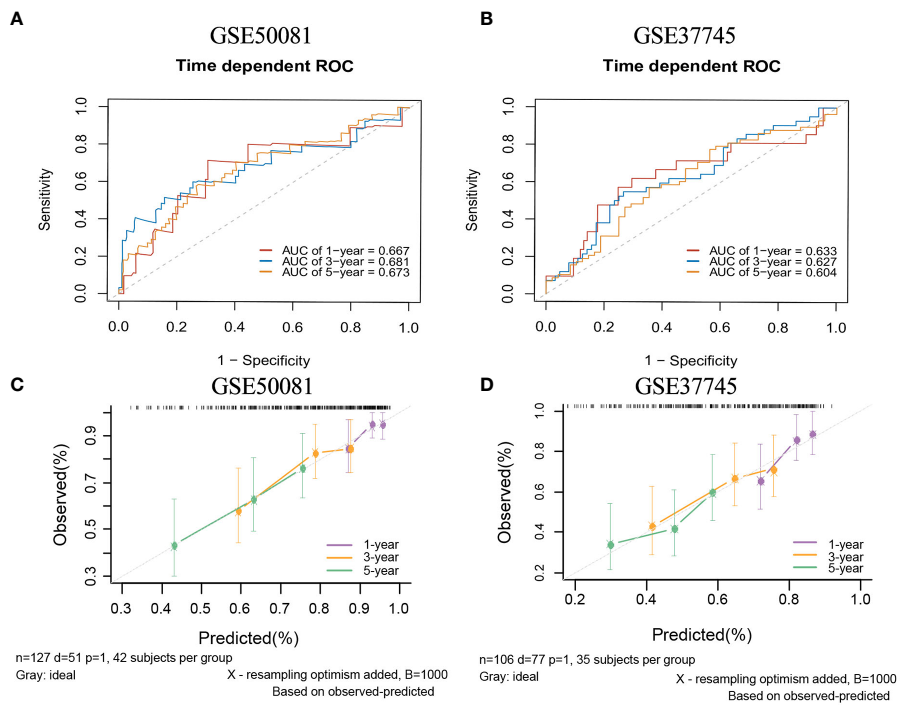


FIGURE 12 | The ROC curves and calibration plots of comprehensive model in external validation data sets.1-,3-,5-year ROC curves: **(A)** GSE50081 dataset and **(B)** GSE37745 dataset. The calibration plots: **(C)** GSE50081 dataset and **(D)** GSE37745 dataset.

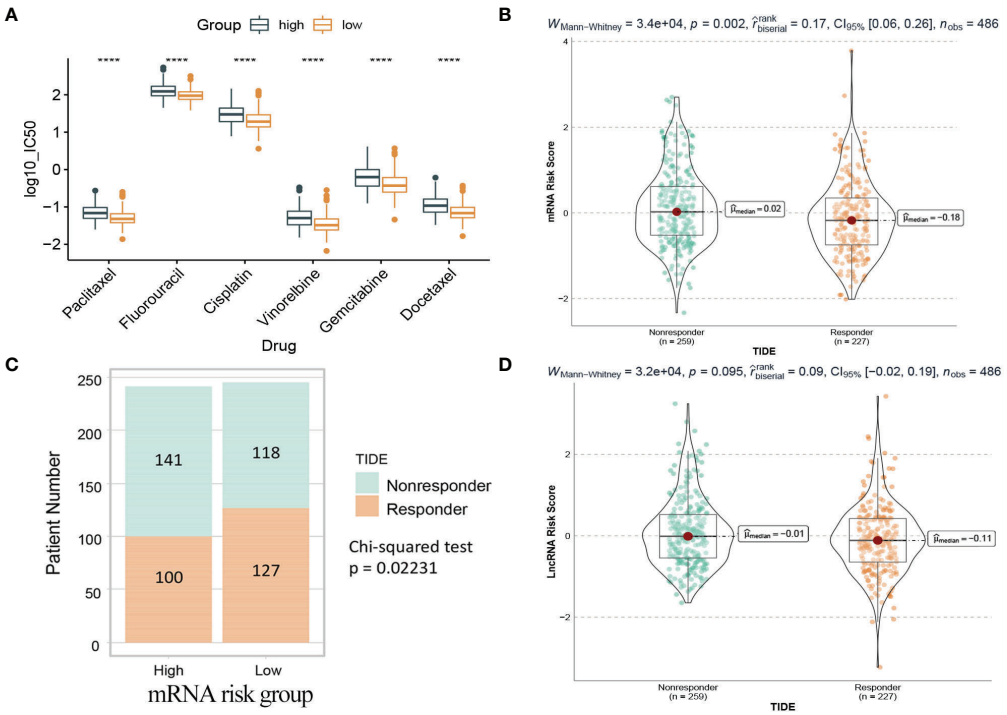


FIGURE 13 | Drug prediction and TIDE immunotherapy prediction analyses. **(A)** Box plot of IC50 values of six chemotherapy drugs between the two risk groups in the mRNA model. **(B)** The mRNA risk score between TIDE predicted responders and non-responders. **(C)** Distribution of TIDE responders and non-responders in the mRNA risk groups. **(D)** The lncRNA risk score between TIDE predicted responders and non-responders. Responder: the patient who responds to the immune checkpoint inhibitors. Nonresponder: the patient who does not respond to the immune checkpoint inhibitors. ****: <0.0001.

to affect RNA modifications (e.g., m^{6}A) and thus disease (14). The m^{6}A-associated variants of 11 genes came from three different confidence levels of sources and two aspects of modification function (**Figure S17**). Four of the mutations lead to lost m^{6}A sites (*USP4, CTCFL, GRIK2, SNTG2*) and ten of the mutations lead to gain m^{6}A sites (*ZNF708, LINGO2, EIF2AK3, KIRREL3, MARCH4, SETDB2, USP4, CASC3*). For m^{6}A sites with high confidence level were derived from miCLIP or PA-m6A-seq experiments (3, 45, 46) and the three m^{6}A-associated variants (*SETDB2, MARCH4, EIF2AK3*) were retained because of locating nearby the m^{6}A sites or disrupting *DRACH* motif around the m^{6}A sites (47–49). For m^{6}A sites having a medium confidence level which were predicted from the previously published MeRIP-seq data (50–52), the four m^{6}A-associated variants (*KIRREL3, EIF2AK3, ZNF708, LINGO2*) were derived from the intersection between the variants and the m^{6}A sites generated from MeRIP-Seq experiments. For m^{6}A sites with a low confidence level predicted by transcriptome-wide prediction, the seven m^{6}A-associated variants (*CTCFL, GRIK2, SNTG2, CASC3, KIRREL3* and *USP4* have two variants) were predicted by the Random Forest prediction model (14). In addition, disease-related data from GWAS and ClinVar databases were collected to determine that the variants of 11 genes were pathogenic mutations leading to dysregulation of m^{6}A modification in lung adenocarcinoma (14). Furthermore, we calculated the correlation coefficients between 11 genes and 21 m^{6}A regulatory factors (**Figure S18**). It turns out that there are varying degrees of correlation between each predictor and regulator.

For mRNA risk score, we also explored their relationship with common chemotherapy drugs and immunotherapy. The study found that patients in the low-risk group were less resistant to commonly used chemotherapy drugs than those in the high-risk group. Furthermore, 11 mRNAs and risk score were calculated for their association with each chemotherapy drug (**Figure S19**). Risk scores were positively correlated with IC50 of all drugs (i.e., patients with higher scores had higher resistance to chemotherapy drugs), indicating that patients with higher scores were insensitive to chemotherapy. Five of the 11 mRNAs (*CTCFL, MARCH4, KIRREL3, GRIK2, LINGO2*) were also positively correlated with IC50 of all drugs. By analyzing the relationship between TIDE score and mRNA risk score, we found that patients with low TIDE scores were more likely to respond to immune checkpoint inhibitors. This may help predict the efficacy of immunotherapy for LUAD. In addition, it is currently believed that a higher value of tumor mutation load represents the higher immunogenicity of the tumor, which is more conducive to immunotherapy drugs, and our analysis also confirmed this view again.

The second model was constructed based on m^{6}A related lncRNAs. There are 15 predictors in total (**Table S5**). The variables screening process of lncRNA model is relatively

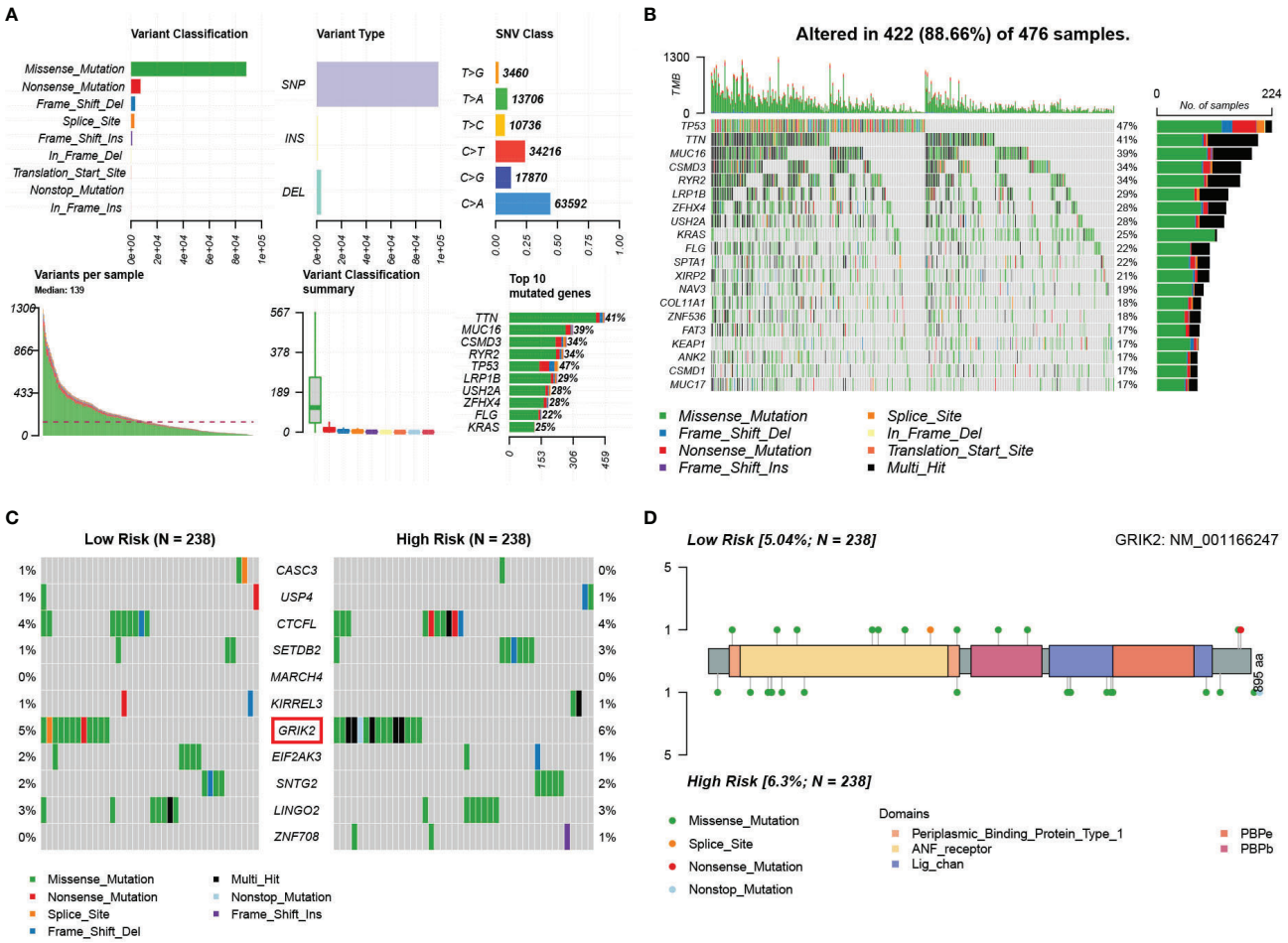


FIGURE 14 | Landscape of somatic mutations in lung adenocarcinoma patients in TCGA. **(A)** the summary of the mutation data. **(B)** The waterfall plot of the mutation distribution of the top 20 most frequently mutated genes. **(C)** The waterfall plot of the mutation distribution of 11 predictors between two risk groups in the mRNA model. **(D)** The lollipop pot of the differential distribution of variants of GRIK2 between two risk groups in mRNA model.

complex, and repeated exploration is to find a prediction model with relatively good discrimination. There are not enough studies on lncRNA in lung adenocarcinoma, but four at present: *SNHG12, RPARP-AS1, CRNDE, LMO7DN*. *SNHG12* has been experimentally predicted as a potential biomarker for the diagnosis, treatment and prognosis of LUAD (53). *RPARP-AS1* and *CRNDE* were included as two predictors in another literature (54). *LMO7DN* has also been suggested as a predictor of lung adenocarcinoma associated with ferroptosis (55).

The third model combined the risk scores of the first two models with clinical variable. There are 3 predictors in total: mRNA risk score, lncRNA risk score, tumor stage. We considered combination of prognostic indices of the two transcriptomic predictive models with clinical variables as a new approach to prognosis prediction and achieved good results.

This study has several advantages. First, all models are based on public cohort data from reliable sources that predict a long survival interval of up to 10 years. Each model was externally validated by multiple independent data sets and stable validation results were obtained. In addition, considering the usability of the model, a model-related web calculator has been developed for anyone to use.

There are several limitations to our study. First, when constructing the comprehensive model, we narrowed the candidate predictors in the development model to three (age, sex, and tumor stage), taking into account the fragmentary clinical variables in validation sets. But it also simplifies the final model somewhat. Secondly, the three models derived in this study are somewhat complicated. In order to reduce the difficulty of practical prediction caused by complex and diverse models, we developed a web calculator containing all models. Thirdly, the performance of our model in external verification will take into account the difference between verification set and derivation set. If the difference is too large, our model may not achieve good performance.

In conclusion, we developed and externally validated three models to predict survival probability of lung adenocarcinoma

based on m^{6}A-related transcriptomics. This may provide clues to new strategies or therapeutic targets for lung adenocarcinoma.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://xena.ucsc.edu/. https://www.ncbi.nlm.nih.gov/geo/.

AUTHOR CONTRIBUTIONS

HL, S-BL, and JS contributed to conceptualization and project administration. HL, S-BL, JC, and NY downloaded and analyzed the data. HL, SB-L wrote the manuscript, with assistance from JS, LB, XZ, KL, and ZT. All authors reviewed the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (81773541), funded from the Priority Academic Program Development of Jiangsu Higher Education Institutions at Soochow University, the State Key Laboratory of Radiation Medicine and Protection (GZK1201919) to ZT, National Natural Science Foundation of China (81872552, U1967220) to JC. Science and Technology Innovation team project of the Suzhou Vocational health college (SZWZYTD201804), Qing-Lan Project of Jiangsu Province in China (2021) to S-BL. National Natural Science Foundation of China (82172441), Suzhou Key Clinical Diagnosis and Treatment Technology Project (LCZX201925) to KL.

ACKNOWLEDGMENTS

We acknowledge the contributions of TCGA and GEO databases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.895148/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
2. Zhu M, Cui Y, Mo Q, Zhang J, Zhao T, Xu Y, et al. Characterization of M(6)A RNA Methylation Regulators Predicts Survival and Immunotherapy in Lung Adenocarcinoma. *Front Immunol* (2021) 12:782551. doi: 10.3389/fimmu.2021.782551
3. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the Human and Mouse M6a RNA Methylomes Revealed by M6a-Seq. *Nature* (2012) 485(7397):201–6. doi: 10.1038/nature11112
4. Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic Transcriptomic M(6)A Decoration: Writers, Erasers, Readers and Functions in RNA Metabolism. *Cell Res* (2018) 28(6):616–24. doi: 10.1038/s41422-018-0040-8
5. Ma L, Chen T, Zhang X, Miao Y, Tian X, Yu K, et al. The M(6)A Reader YTHDC2 Inhibits Lung Adenocarcinoma Tumorigenesis by Suppressing SLC7A11-Dependent Antioxidant Function. *Redox Biol* (2021) 38:101801. doi: 10.1016/j.redox.2020.101801
6. Liu XS, Zhou LM, Yuan LL, Gao Y, Kui XY, Liu XY, et al. NPM1 Is a Prognostic Biomarker Involved in Immune Infiltration of Lung Adenocarcinoma and Associated With M6a Modification and Glycolysis. *Front Immunol* (2021) 12:724741. doi: 10.3389/fimmu.2021.724741
7. Qian X, Yang J, Qiu Q, Li X, Jiang C, Li J, et al. LCAT3, a Novel M6a-Regulated Long non-Coding RNA, Plays an Oncogenic Role in Lung Cancer *via* Binding With FUBP1 to Activate C-MYC. *J Hematol Oncol* (2021) 14 (1):112. doi: 10.1186/s13045-021-01123-0
8. Wang H, Zhao X, Lu Z. M(6)A RNA Methylation Regulators Act as Potential Prognostic Biomarkers in Lung Adenocarcinoma. *Front Genet* (2021) 12:622233. doi: 10.3389/fgene.2021.622233
9. Sun J, Ping Y, Huang J, Zeng B, Ji P, Li D. N6-Methyladenosine-Regulated mRNAs: Potential Prognostic Biomarkers for Patients With Lung Adenocarcinoma. *Front Cell Dev Biol* (2021) 9:705962. doi: 10.3389/fcell.2021.705962
10. Zhu J, Wang M, Hu D. Deciphering N(6)-Methyladenosine-Related Genes Signature to Predict Survival in Lung Adenocarcinoma. *BioMed Res Int* (2020) 2020:2514230. doi: 10.1155/2020/2514230
11. Liu X, Li N, Liu S, Wang J, Zhang N, Zheng X, et al. Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review. *Front Bioeng Biotechnol* (2019) 7:358. doi: 10.3389/fbioe.2019.00358
12. Welsh EA, Eschrich SA, Berglund AE, Fenstermacher DA. Iterative Rank-Order Normalization of Gene Expression Microarray Data. *BMC Bioinf* (2013) 14:153. doi: 10.1186/1471-2105-14-153
13. Guo B, Zhang H, Wang J, Wu R, Zhang J, Zhang Q, et al. Identification of the Signature Associated With M(6)A RNA Methylation Regulators and M(6)A-Related Genes and Construction of the Risk Score for Prognostication in Early-Stage Lung Adenocarcinoma. *Front Genet* (2021) 12:656114. doi: 10.3389/fgene.2021.656114
14. Zheng Y, Nie P, Peng D, He Z, Liu M, Xie Y, et al. M6avar: A Database of Functional Variants Involved in M6a Modification. *Nucleic Acids Res* (2018) 46(D1):D139–D45. doi: 10.1093/nar/gkx895
15. Goeman JJ. L1 Penalized Estimation in the Cox Proportional Hazards Model. *Comp Stud* (2010) 52(1):70–84. doi: 10.1002/bimj.200900028
16. Venables WN, Ripley BD. Modern Applied Statistics With S. In: *Statistic and Computing*. Springer (2002).
17. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Cham: Springer (2010). XXV, 582 p. doi: 10.1007/978-3-319-19425-7
18. Wang J, Yuan T, Ding H, Xu J, Keusters WR, Ling X, et al. Development and External Validation of a Prognostic Model for Survival of People Living With HIV/AIDS Initiating Antiretroviral Therapy. *Lancet Reg Health West Pac* (2021) 16:100269. doi: 10.1016/j.lanwpc.2021.100269
19. Harrell FJr., Califf RM. Evaluating the Yield of Medical Tests. *JAMA* (1982) 247(18):2543. doi: 10.1001/jama.1982.03320430047030
20. Heagerty PJ, Lumley T, Pepe MSJB. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* (2000) 56 (2):337–44. doi: 10.1111/j.0006-341X.2000.00337.x
21. Gerds TA, Kattan MW, Schumacher M, Yu C. Estimating a Time-Dependent Concordance Index for Survival Prediction Models With Covariate Dependent Censoring. *Stat Med* (2013) 32(13):2173–84. doi: 10.1002/sim.5681
22. Royston P, Altman DG. External Validation of a Cox Prognostic Model: Principles and Methods. *BMC Med Res Methodol* (2013) 13:15. doi: 10.1186/1471-2288-13-33
23. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. *Springer* (2019) 66(2):661–2. doi: 10.1007/978-3-030-16399-0

24. Therneau T, Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model (Statistics for Biology and Health). *Springer* (2000). doi: 10.1007/978-1-4757-3294-8
25. Hothorn T. Maxstat: Maximally Selected Rank Statistics in R. *Biometri J.* (2015) 44. doi: 10.1002/1521-4036(200203)44:23.0.CO;2-Z
26. Leek JT, Johnson WE, Parker HS, Jaffe ®, Storey JDJB. The SVA Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
27. Ishwaran H. Random Survival Forest. *JAAS* (2008) 2:25–31. doi: 10.1214/08-AOAS169
28. Ishwaran H, Kogalur UB. Randomforestsrc: Random Forests for Survival, Regression and Classification (RF-SRC). (2021). Available at: https://cran.r-project.org/web/packages/randomForestSRC/citation.html.
29. Wen C, Zhang A, Quan S, Wang X. BeSS: Best Subset Selection for Sparse Generalized Linear Model and Cox Model. *J Stat Softw* (2020) 94:1–24. doi: 10.18637/jss.v094.i04
30. Murad H, Dankner R, Berlin A, Olmer L, Freedman LS. Imputing Missing Time-Dependent Covariate Values for the Discrete Time Cox Model. *Stat Methods Med Res* (2019) 29(8):962280219881168. doi: 10.1177/0962280219881168
31. Moons K, Donders R, Stijnen T, Harrell FE. Using the Outcome for Imputation of Missing Predictor Values was Preferred. *J Clin Epideminol* (2006) 59(10):1092–101. doi: 10.1016/j.jclinepi.2006.01.009
32. White IR, Royston P. Imputing Missing Covariate Values for the Cox Model. *Stat Med* (2010) 28(15):1982–98. doi: 10.1002/sim.3618
33. Wood AM, White IR, Royston P. How Should Variable Selection be Performed With Multiply Imputed Data? *Stat Med* (2010) 27(17):3227–46. doi: 10.1002/sim.3177
34. Mayakonda A, Koeffler HP. Maftools: Efficient Analysis, Visualization and Summarization of MAF Files From Large-Scale Cohort Based Cancer Studies. *BioRxiv* (2016) 2016:052662. doi: 10.1101/052662
35. Ribas A, Wolchok JD. Cancer Immunotherapy Using Checkpoint Blockade. *Science* (2018) 359(6382):1350–5. doi: 10.1126/science.aar4060
36. Chen X, Zhu Z, Li X, Yao X, Luo L. The Ferroptosis-Related Noncoding RNA Signature as a Novel Prognostic Biomarker in the Tumor Microenvironment, Immunotherapy, and Drug Screening of Gastric Adenocarcinoma. *Front Oncol* (2021) 11:778557. doi: 10.3389/fonc.2021.778557
37. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-Scale Public Data Reuse to Model Immunotherapy Response and Resistance. (2020) 12:1–8. doi: 10.1186/s13073-020-0721-z
38. Jiang P, Gu S, Deng P, Fu J, Sahu A, Hu X, et al. Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat Med* (2018) 2018:1550–8. doi: 10.1158/2326-6074.CRICIMTEATIAACR18-B077
39. Maeser D, Gruener RF, Huang RS. Predict: An R Package for Predicting *In Vivo* or Cancer Patient Drug Response and Biomarkers From Cell Line Screening Data. *JBiB Onco* (2021) 11:bbab260. doi: 10.1093/bib/bbab260
40. Schemper M, Smith TL. A Note on Quantifying Follow-Up in Studies of Failure Time. *Control Clin Trials* (1996) 17(4):343–6. doi: 10.1016/0197-2456(96)00075-X
41. Templ M, Alfons A, Filzmoser P. Exploring Incomplete Data Using Visualization Techniques. *Adv Data Anal Classificat* (2011) 6(1):29–47. doi: 10.1007/s11634-011-0102-y
42. Das T, Lee EY, You HJ, Kim EE, Song EJ. USP15 and USP4 Facilitate Lung Cancer Cell Proliferation by Regulating the Alternative Splicing of SRSF1. *Cell Death Discovery* (2022) 8(1):24. doi: 10.1038/s41420-022-00820-0
43. Li N, Qiu L, Zeng C, Fang Z, Chen S, Song X, et al. Bioinformatic Analysis of Differentially Expressed Genes and Pathways in Idiopathic Pulmonary Fibrosis. *Ann Transl Med* (2021) 9(18):1459. doi: 10.1183/13993003.congress-2021.OA4331
44. Tran AN, Dussaq AM, Kennell TJr., Willey CD, Hjelmeland AB. HPAanalyze: An R Package That Facilitates the Retrieval and Analysis of the Human Protein Atlas Data. *BMC Bioinf* (2019) 20(1):463. doi: 10.1186/s12859-019-3059-z
45. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons. *Cell* (2012) 149(7):1635–46.
46. Schwartz S, Mumbach M, Jovanovic M, Wang T, Maciag K, Bushkin GG, et al. Perturbation of M6a Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5′ Sites. *Cell Rep* (2014) 8(1):284–96. doi: 10.1016/j.celrep.2014.05.048
47. Zhang Z, Chen LQ, Zhao YL, Yang CG, Luo GZ. Single-Base Mapping of M6a by an Antibody-Independent Method. *Sci Adv* (2019) 2019:eaax0250. doi: 10.1101/575555
48. Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, et al. A Majority of M6a Residues are in the Last Exons, Allowing the Potential for 3' UTR Regulation. *Genes Dev* (2015) 29(19):2037–53. doi: 10.1101/gad.269415.115
49. Sendinc E, Valle-Garcia D, Dhall A, Chen H, Henriques T, Navarrete-Perea J, et al. PCIF1 Catalyzes M6am mRNA Methylation to Regulate Gene Expression. *Mol Cell* (2019) 75(3):620–30.e9. doi: 10.1016/j.molcel.2019.05.030
50. Liu J, Li K, Cai J, Zhang M, Zhang X, Xiong X, et al. Landscape and Regulation of M(6)A and M(6)Am Methylome Across Human and Mouse Tissues. *Mol Cell* (2020) 77(2):426–40 e6. doi: 10.1016/j.molcel.2019.09.032
51. Hasan I, Mohsin K, Gokhale NS, Mcintyre A, Geon-Woo K, Young JJ, et al. N6 -Methyladenosine Modification of Hepatitis B Virus RNA Differentially Regulates the Viral Life Cycle. *Proc Nat Acad Sci* (2018) 115:201808319. doi: 10.1073/pnas.1808319115
52. Li Z, Qian P, Shao W, Shi H, He XC, Madelaine G, et al. Suppression of m6A reader Ythdf2 promotes hematopoietic stem cell expansion. *Cell Res* (2018) 28(9):14. doi: 10.1038/s41422-018-0072-0
53. Zhang H, Wang Y, Lu J. Identification of Lung-Adenocarcinoma-Related Long non-Coding RNAs by Random Walking on a Competing Endogenous RNA Network. *Ann Transl Med* (2019) 7(14):339. doi: 10.21037/atm.2019.06.69
54. Zheng J, Zhao Z, Wan J, Guo M, Wang Y, Yang Z, et al. N-6 Methylation-Related lncRNA is Potential Signature in Lung Adenocarcinoma and Influences Tumor Microenvironment. *J Clin Lab Anal* (2021) 35(11):e23951. doi: 10.1002/jcla.23951
55. Lu L, Liu LP, Zhao QQ, Gui R, Zhao QY. Identification of a Ferroptosis-Related LncRNA Signature as a Novel Prognosis Model for Lung Adenocarcinoma. *Front Oncol* (2021) 11:675545. doi: 10.3389/fonc.2021.675545

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fonc.2022.906778

Epidemiological Characteristics of Primary Liver Cancer in Mainland China From 2003 to 2020: A Representative Multicenter Study

Jiansheng Lin[1,2†], Hongwei Zhang[1†], Hongping Yu[3†], Xinyu Bi[4†], Weilu Zhang[5†], Jianhua Yin[1], Pei Zhao[3], Xiumei Liang[6], Chunfeng Qu[7], Minjie Wang[8], Ming Hu[1], Kun Liu[5], Yuting Wang[7], Zihan Zhou[9], Junqi Wang[1], Xiaojie Tan[1], Wenbin Liu[1], Zhongjun Shao[5], Jianqiang Cai[4*], Weizhong Tang[10*] and Guangwen Cao[1*]*

[1] Department of Epidemiology, Second Military Medical University, Shanghai, China, [2] Department of Epidemiology, School of Medicine, Jinan University, Guangzhou, China, [3] Department of Research, Guangxi Medical University Cancer Hospital, Nanning, China, [4] Department of Hepatobiliary Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [5] Department of Epidemiology, School of Public Health, Fourth Military Medical University, Xi'an, China, [6] Office for Disease Process Management, Guangxi Medical University Cancer Hospital, Nanning, China, [7] State Key Lab of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [8] Department of Clinical Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [9] Guangxi Office for Cancer Prevention and Control, Guangxi Medical University Cancer Hospital, Nanning, China, [10] Department of Gastrointestinal Surgery and Guangxi Clinical Research Center for Colorectal Cancer, Guangxi Medical University Cancer Hospital, Nanning, China

OPEN ACCESS

Edited by:
Dana Kristjansson,
Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Ming Yue,
Nanjing Medical University, China
Zhaohui Tang,
Shanghai Jiao Tong University, China

***Correspondence:**
Zhongjun Shao
13759981783@163.com
Jianqiang Cai
caijiangqiang188@sina.com
Weizhong Tang
tangweizhong@gxmu.edu.cn
Guangwen Cao
gcao@smmu.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: 29 March 2022
Accepted: 11 May 2022
Published: 21 June 2022

Citation:
Lin J, Zhang H, Yu H, Bi X, Zhang W, Yin J, Zhao P, Liang X, Qu C, Wang M, Hu M, Liu K, Wang Y, Zhou Z, Wang J, Tan X, Liu W, Shao Z, Cai J, Tang W and Cao G (2022) Epidemiological Characteristics of Primary Liver Cancer in Mainland China From 2003 to 2020: A Representative Multicenter Study. Front. Oncol. 12:906778. doi: 10.3389/fonc.2022.906778

Background: The contribution of hepatitis B virus (HBV) and hepatitis C virus (HCV) to primary liver cancer (PLC) and their association with cancer aggressiveness remains uncertain in China, a country with half of global PLC. We aimed to characterize this using data from four representative medical centers.

Methods: In total, 15,801 PLC patients were enrolled from the centers distributed in Easter5n, Southern, Northern, and Western China from 2003 to 2020. Of those, 7585 with curative surgery were involved in survival analysis. A nomogram was constructed using preoperative parameters to predict postoperative survival.

Results: Hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma, and combined hepatocellular cholangiocarcinoma accounted for 93.0%, 4.3%, and 1.6% in PLC, respectively. The seropositivities of HBV and HCV were 84.4% and 3.2% in HCC, respectively. The seropositivity of anti-HCV antibody was significantly higher in HBV-negative than in HBV-positive HCC patients (13.2% vs. 1.1%). Compared to HCV-positive HCC (HCV-HCC), HBV-positive HCC (HBV-HCC) was associated with 12-year earlier onset, higher proportions of males, high α-fetoprotein, large tumor size, advanced Barcelona Clinic Liver Cancer (BCLC) stage, and vascular tumor thrombus. The proportions of HCC and HBV seropositivity increased, whereas that of anti-HCV decreased, from 2003 to 2020. Postoperative five-year survival rate was 73.5%, 64.1%, 34.9%, and 19.7% in HCC at BCLC stage 0, A, B, and C, respectively. The

multivariate Cox regression analysis showed that HBV seropositivity, incomplete tumor capsule, vascular tumor thrombus, tumor diameter (≥3 cm), advanced BCLC stage (B+C), α-fetoprotein (≥20ng/ml), and direct bilirubin (>8μmol/L) contributed independently to shorter overall survival (OS); whereas post-operative radiofrequency ablation and second resection independently improved OS in HCC. HCV-HCC had a more favorable prognosis than did HBV-HCC (Log-rank test, $P<0.001$). A nomogram composed of age, gender, and the preoperative independent risk factors was accurate in predicting postoperative survival in HCC (C-index: 0.735; 95% confidence interval: 0.727–0.743).

Conclusion: HBV contributes to 84.4% of HCC in China, and actively promotes hepatocarcinogenesis and HCC progression. A favorable postoperative survival obtained in patients at the early BCLC stage highlights the importance of screening for early HCC in high-risk populations. Our preoperative prognosis prediction model is important in clinical decision-making.

Keywords: primary liver cancer, hepatocellular carcinoma, hepatitis B virus, hepatitis C virus, prognosis

INTRODUCTION

Cancer may surpass cardiovascular disease as the leading cause of immature death in 57 countries including China (1). Primary liver cancer (PLC) is the sixth most commonly diagnosed cancer and the third leading cause of cancer death worldwide in 2020, with 905,677 new cases and 830,180 deaths in 2020 (2). PLC remains the second cause of cancer death and the first leading cause of immature cancer death in mainland China (2–4). The major histotypes of PLC are hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (ICC), and combined hepatocellular cholangiocarcinoma (CHC). The major causes of HCC are chronic infection with hepatitis B virus (HBV) and/or hepatitis C virus (HCV), alcohol consumption, aflatoxin B1 exposure, non-alcoholic fatty liver disease (NAFLD), and diabetes (5, 6). The risk factors of ICC include primary sclerosing cholangitis, hepatolithiasis, infection with *Opisthorchis viverrini* and *Clonorchis sinensis*, and chronic infection with HBV or HCV (7). However, the proportions of etiological agents and histotypes of PLC differ greatly among different studies. It was reported that 56% of global PLC were attributable to HBV and 20% to HCV (5); however, the corresponding proportions were estimated to be 33% and 21% in another global study (8). HCV infection is the leading cause of HCC in most European and American countries, while chronic HBV infection is the leading etiologic factor of HCC in Asian and African countries where HCC is endemic (5). HCC comprised 75%-85% while ICC 10%-15% of global PLC (2); however, HCC and ICC comprised 94.6% and 3.7% of PLC in eastern China (9). These data reflect apparent heterogeneities in the major etiological agents and the major histotypes of PLC worldwide.

In China, the contribution of major etiological agents to PLC and the proportion of major histotypes were only from very limited resources (9, 10). Currently, there are no representative data describing the major etiological agents and the proportion of major histotypes of PLC in China, a country with half of global PLC. A large, highly representative study population is indispensable to address these issues. The effect of HBV or HCV infection on the prognosis of HCC remains obscure. It has been shown that HBV-related HCC (HBV-HCC) has a better prognosis than HCV-related HCC (HCV-HCC) (11–13). However, this result is not repeated in other populations (14). This issue should be addressed by the propensity score (PS) matching method. In this study, we firstly evaluated the seropositivities of HBV and HCV in large-scale PLC patients from four representative medical centers in mainland China, and then investigated the difference in clinical characteristics of HBV-HCC and HCV-HCC. Finally, we developed a nomogram composed of preoperative clinical parameters to predict postoperative prognosis in HCC.

MATERIALS AND METHODS

Participants Enrollment

In total, 15,816 consecutive patients with PLC were enrolled from four medical centers located in Northern (Beijing), Eastern (Shanghai), Western (Wuwei, Gansu), and Southern (Nanning, Guangxi) parts of mainland China from January 1, 2003, to June 30, 2020. Of those, 13,978 cases were pathologically diagnosed. Post-operative cohort studies were established based on personal willingness for the analysis of risk factors related to survival. The follow-up was carried out after curative surgery according to our existing protocol (15). During the follow-up, the information on the survival situation, the exact date of death, and treatment(s) received after surgery were collected. If patients had imaging evidence of tumor recurrence, second resection or radiofrequency ablation (RFA) was suggested (16). Postoperative transcatheter arterial chemoembolization (TACE) was recommended for patients with microvascular invasion (MVI) as previously described (17).

Data Collection

The data of demographic and clinical characteristics were extracted from medical records, including birth date, age of

onset, gender, nationality, place of birth, pathological findings (including pathological types, capsule integrity of tumor, nodule number, and vascular invasion), and laboratory examinations (serum AFP, parameters of HBV and HCV, routine blood assay, and liver function test). Barcelona Clinic Liver Cancer (BCLC) stage was identified as previously described (18). Han Chinese accounted for 91.2% of the study participants. The study protocol conformed to the ethical guidelines of the 2000 Declaration of Helsinki and was approved by the ethics committee of each involved medical center.

Statistical Analysis

Data were independently checked by two researchers carefully. Fifteen duplicated cases were removed from the analysis. Those seropositive for HBsAg and/or HBV DNA were defined as positive HBV infection. The seropositive cutoff values for HBsAg and HBV DNA were >0.05 IU/mL and >500 copies/mL, respectively. The cutoff values of AFP, total bilirubin, direct bilirubin, and albumin were in accordance with the criteria adopted in clinic examination. For categorical variables, χ^2 test or Fisher's exact test was conducted for comparison between groups. Continuous variables with skewed distribution were compared by Mann–Whitney *U*-test. The trends of change in proportion of variables, such as HBV, HCC, and BCLC, were tested by using the Cochran-Armitage method. A Cox proportional hazard model was conducted to calculate the azard ratio (HR) and 95% confidence interval (CI) for each variable. Significant variables in the univariate Cox analysis and professionally meaningful variables were introduced into the multivariate Cox model, and the backward stepwise Wald method was applied to determine the factors that independently contributed to postoperative survival. The 1:2 propensity score (PS) matching was performed for survival comparison between HCV-HCC and HBV-HCC patients. For the comparison of the prognosis between HBV-HCC and HCV-HCC patients, Kaplan–Meier method was conducted to estimate overall survival (OS), and log-rank test was applied to compare the difference of OS between the two groups. For the prediction of HCC prognoses, the cohort was randomly grouped into training and validation sub-cohorts. A Cox regression model composed of statistically significant preoperative variables was established in the training, following the Akaike information criterion (19). A nomogram with these preoperative variables was formulated by the *rms* package in R (20). The fitness of the nomogram was evaluated by the concordance index (C-index) and calibration plots with 1000 bootstraps. The prediction power of the nomogram was verified in the validation cohort. Time-dependent receiver operating characteristics (ROC) curve was also applied to evaluate the accuracy of the nomogram in the training cohort and validation cohort. According to the medium of risk score calculated from the established nomogram, the subjects were divided into low-risk and high-risk groups. The difference in OS between the two groups was compared in the training cohort and validation cohort, respectively. All statistical analyses were two-sided and performed using SPSS V21.0 for Windows (http://www-01.ibm.com/software/uk/analytics/spss/,RRID:SCR_002865) and R software (version 4.0.2, https://www.r-project.org/). α=0.05 was considered astatistically significant level.

RESULTS

Characteristics of Patients With PLC

In total, 15,816 PLC cases admitted to the four medical centers from January 1, 2003, to June 30, 2020 were enrolled in this study. The chart flow of study patients is shown in **Supplementary Figure 1**. Enrolled patients were from almost all provinces of mainland China. Interestingly, patients enrolled in Beijing were frequently from Heilongjiang province, while patients in Shanghai were frequently from nearby provinces (**Supplementary Figure 2**). The medium age of 15,801 patients at the diagnosis of PLC was 54 (inter-quartile range [IQR], 46–62 years). A male-to-female ratio was 5.29:1. Among all PLC patients, 80.1% were seropositive for HBV and 3.4% were seropositive for anti-HCV antibody. The seropositivity for HBV was more frequent in the east (87.6%) and south (87.7%) than in the north (57.2%) and west (61.6%), which was quite in contrast to the seropositivity for anti-HCV antibody (**Table 1**). Male PLC patients were 2 years younger than female patients at diagnosis (54 [46-62] vs. 56 [48-63] years, P<0.001). Compared to female PLC patients, males had higher proportions of advanced BCLC stage, liver cirrhosis, abnormal total bilirubin, and direct bilirubin. The seropositivity for HBV was more frequent in male than in female PLC patients (81.7% vs. 71.0%, P<0.001), which is in contrast to the seropositivity for anti-HCV antibody (3.2% vs. 4.4%) (**Supplementary Table 1**). Of 15,801 PLC patients, 13,978 had pathological data. HCC, ICC, and CHC accounted for 93.0%, 4.3%, and 1.6%, respectively. HCC, ICC, and CHC accounted for 94.3%, 3.3%, and 1.5% in the males and 85.7%, 10.5%, and 2.0% in the female PLC patients, respectively. Patients with HCC were 3 years younger than patients with ICC at diagnosis (53 [46-61] vs. 57 [49-64], P<0.001). The rate of seropositivity for HBV was 84.4%, 38.6%, and 77.1% while that for anti-HCV antibody was 3.2%, 1.8%, and 1.5% in patients with HCC, patients with ICC, and those with CHC, respectively.

Trends in the Seropositivities of HBV and HCV, the Proportions of HCC and Early-Stage Tumors as Well as Postoperative Survival in PLC Patients From 2003 to 2020

From 2003 to 2020, the seropositivity of HBV as well as the proportions of HCC and early BCLC stage (0&A) showed increasing trends in PLC patients (**Figure 1A**). However, the seropositivity of HBV in HCC patients declined from 86.8% before 2010 to 84.32% in 2011-2015 (**Figure 1B**). The seropositivity for anti-HCV antibody decreased significantly either in all PLC or in HCC (**Figures 1A, B**). Interestingly, the postoperative survival including 1-year survival and 3-year survival rates showed increasing trends in HCC patients from 2003 to 2020; the 5-year survival rate had a similar tendency from 2003 to 2015 (**Figure 1C**, P_{trend} < 0.001).

TABLE 1 | Demographic and clinical characteristics of PLC patients from the four medical centers.

Variable	Total (n = 15801)	Shanghai (n = 8515)	Beijing (n = 2561)	Nanning (n = 2813)	Wuwei (n = 1912)	P
Age (yr)						
Medium (IQR)	54 (46–62)	54 (46–61)	56 (48–63)	51 (43–60)	58 (49-66)	<0.001
<40	1,562 (9.9)	780 (9.2)	206 (8.0)	443 (15.7)	133 (7.0)	<0.001
40–59	9,098 (57.6)	5,113 (60.0)	1,424 (55.6)	1,664 (59.2)	897 (46.9)	
≥60	5,141 (32.5)	2,622 (30.8)	931 (36.4)	706 (25.1)	882 (46.1)	
Gender						
Female	2,519 (15.9)	1,191 (14.0)	468 (18.3)	367 (13.0)	493 (25.8)	<0.001
Male	13,282 (84.1)	7,324 (86.0)	2,093 (81.7)	2,446 (87.0)	1,419 (74.2)	
HBV						
Negative	2,999 (19.9)	1,034 (12.4)	1,095 (42.8)	345 (12.3)	525 (38.4)	<0.001
Positive	12,064 (80.1)	7,293 (87.6)	1,461 (57.2)	2,467 (87.7)	843 (61.6)	
HCV						
Negative	11,027 (96.6)	7,860 (98.3)	1,941 (91.1)	739 (98.1)	487 (91.5)	<0.001
Positive	386 (3.4)	137 (1.7)	190 (8.9)	14 (1.9)	45 (8.5)	
Cirrhosis						
No	7,357 (53.8)	4,684 (55.2)	500 (59.2)	1,088 (39.5)	1,085 (68.4)	<0.001
Yes	6,318 (46.2)	3,803 (44.8)	345 (40.8)	1,668 (60.5)	502 (31.6)	
AFP (ng/ml)						
<20	4,940 (39.3)	3,187 (37.9)	820 (57.3)	933 (34.2)	–	<0.001
≥20	7,618 (60.7)	5,211 (62.1)	611 (42.7)	1,796 (65.2)	–	
Albumin (g/L)						
≥40	6,661 (69.6)	5,577 (68.2)	164 (48.7)	920 (34.5)	–	<0.001
<40	4,518 (40.4)	2,595 (31.8)	173 (51.3)	1,750 (65.5)	–	
Total bilirubin (μmol/L)						
≤23	10,077 (88.9)	7,530 (90.5)	315 (92.1)	2,232 (83.6)	–	<0.001
>23	1,257 (11.1)	792 (9.5)	27 (7.9)	438 (16.4)	–	
Direct bilirubin (μmol/L)						
≤8	8,976 (81.5)	7,034 (84.5)	–	1,942 (72.2)	–	<0.001
>8	2,035 (18.5)	1,288 (15.5)	–	747 (27.8)	–	
Ascites						
No	11,205 (96.7)	8,370 (98.4)	474 (95.0)	2,361 (91.6)	–	<0.001
Yes	381 (3.3)	140 (1.6)	25 (5.0)	216 (8.4)	–	
BCLC stage						
0	472 (4.0)	378 (4.5)	69 (8.9)	25 (1.0)	–	<0.001
A	4,906 (41.3)	3,028 (35.9)	633 (81.7)	1,245 (48.5)	–	
B	4,330 (36.7)	3,699 (43.8)	57 (7.3)	574 (22.4)	–	
C	2,076 (17.6)	1,336 (15.8)	16 (2.1)	724 (28.2)	–	
Pathological type						
HCC	13,003 (93.0)	8,056 (94.6)	2,110 (82.4)	2,759 (98.3)	78 (83.0)	<0.001
ICC	607 (4.3)	314 (3.7)	282 (11.0)	0	11 (11.7)	
CHC	222 (1.6)	145 (1.7)	59(2.3)	16 (0.6)	2 (2.1)	
Others	146 (1.0)	0	110 (4.3)	33 (1.2)	3 (3.2)	

Data are shown in n (%).

Clinical Characteristics Between PLC Patients With HBV Infection and Those Without HBV Infection

Compared to HCC patients without HBV infection, those with HBV infection were 8 years younger and showed higher proportions of AFP (≥20 ng/ml), the presence of liver cirrhosis, abnormal albumin (<40g/L), advanced BCLC stage, multiple tumor nodules, and vascular tumor thrombus as well as a higher male-to-female ratio. The seropositivity for anti-HCV antibody in HBV-free HCC patients was significantly higher than that in those with HBV infection (13.2% vs. 1.1%, $P<0.001$) (**Table 2**). ICC patients with HBV infection were 5 years younger and had a higher male-to-female ratio, higher proportions of AFP (≥20ng/ml), liver cirrhosis, incomplete tumor capsule, vascular tumor thrombus, and advanced BCLC stage compared to those without HBV infection (**Supplementary Table 2**). We then compared the clinical characteristics of HCC patients solely caused by HBV infection and those by HCV infection. Compared to HCV-HCC patients, HBV-HCC patients were 12 years younger and had a higher proportion in males, advanced BCLC stage, vascular tumor thrombus, multiple tumor nodules, larger tumor size (≥3 cm in diameter), and higher AFP (≥20ng/mL) (**Table 3**).

Risk Factors Related to the Overall Survival of Major Histotypes of PLC

In total, 7679 PLC patients who received curative surgery at the study medical centers were invited to participate in the follow-up study. Of those, 94 were excluded due to lack of survival data, and the remaining 7585 patients were included in the

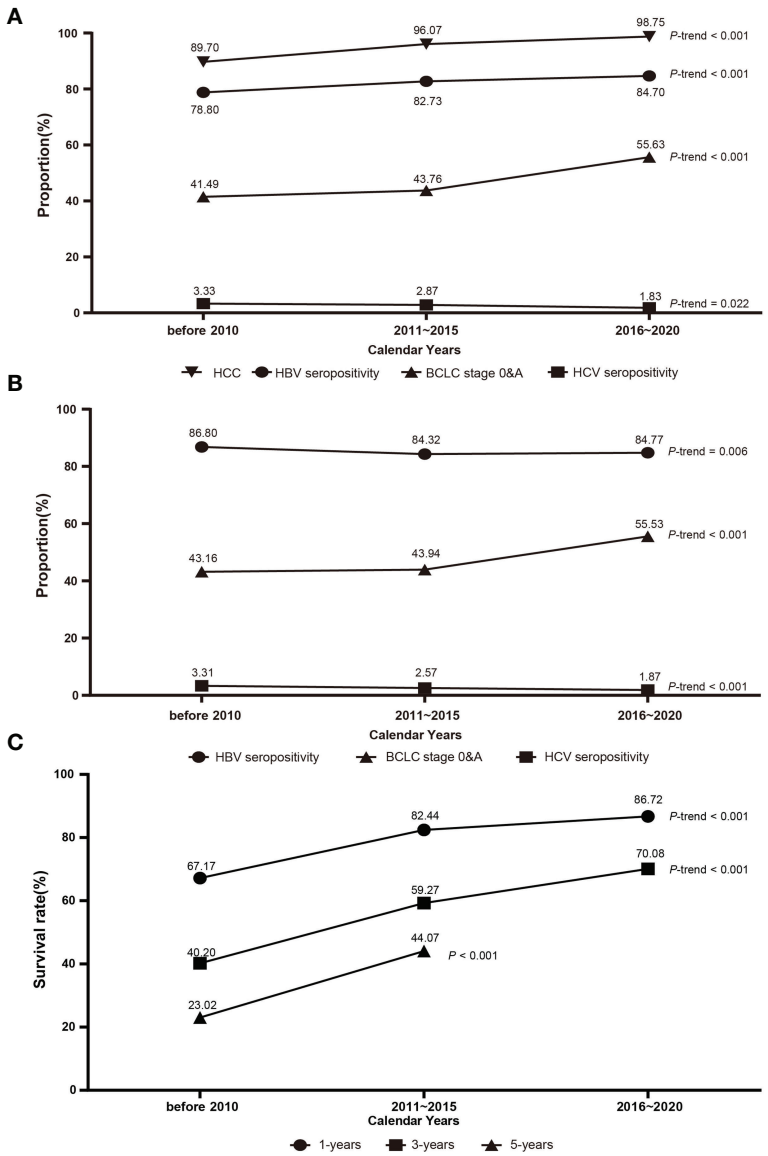


FIGURE 1 | The dynamic trend of clinical parameters of PLC patients from 2003 to 2020. **(A)** The proportions of HCC, HBV seropositivity, HCV seropositivity, and early BCLC stage in PLC. **(B)** The proportions of HBV seropositivity, HCV seropositivity, and early BCLC stage in HCC. **(C)** The dynamic trends of 1-year, 3-year, and 5-year survival rates. Linear trends were calculated by using Cochran-Armitage test. BCLC, Barcelona Clinic Liver Cancer; HBV, hepatitis B virus; HCV, hepatitis C virus; HCC, hepatocellular carcinoma.

survival analysis. The median follow-up time was 1.50 years, with an IQR of 0.75–3.17 years. Of the 7585 PLC patients, 2809 died of this malignancy during follow-up, with the 1-, 3-, and 5-year survival rates of 79.7%, 56.9%, and 41.3%, respectively. The 1-, 3-, and 5-year survival rates were 80.7%, 58.3%, and 42.1% in HCC; 48.7%, 24.2%, and 18.5% in ICC; and 68.6%, 30.1%, and 23.8% in CHC. Of patients in the four regions, patients in the west (Wuwei, Gansu) were excluded from the analysis because they did not have qualified histological and prognostic data. The 1-, 3-, and 5-year survival rates of HCC patients were 74.6%, 42.3%, and 25.9% in Shanghai; 85.5%, 67.3%, and 49.9% in Nanning; and 92.5%, 83.6%, and 69.4% in Beijing, respectively. In order to analyze the relationship between BCLC stages and postoperative survival of PLC patients, the survival rate was calculated according to the different BCLC stages. We confirmed that the survival rate decreased significantly with increasing BCLC stages (P<0.001). Compared to HCC patients at BCLC B&C stage, HCC patients at 0&A stage had a better postoperative prognosis; the same was true for ICC patients (**Table 4**).

The Cox proportional hazard model was applied to evaluate factors significantly associated with postoperative OS in HCC. The univariate Cox regression analysis identified 13 factors that were significantly associated with OS. The multivariate Cox

TABLE 2 | Comparison of demographic and clinical characteristics in HCC patients with information on HBV infection.

Variable	Total (n = 12,824)	Patients without HBV infection (n = 2,001)	Patients with HBV infection (n = 10,823)	*P*-value
Age (year)				
Medium (IQR)	53 (46-61)	60 (53-68)	52 (45-60)	
<40	1,356 (10.6)	103 (5.1)	1,253 (11.6)	<0.001
40–59	7,639 (59.6)	822 (41.1)	6,817 (63.0)	
≥60	3,829 (29.8)	1,076 (53.8)	2,753 (25.4)	
Gender				
Female	1,718 (13.4)	302 (15.1)	1,416 (13.1)	0.015
Male	11,106 (86.6)	1,699 (84.9)	9,407 (86.9)	
HCV				
Negative	9,710 (96.9)	1,480 (86.8)	8,230 (98.9)	<0.001
Positive	315 (3.1)	225 (13.2)	90 (1.1)	
Cirrhosis				
No	5,762 (50.7)	1,052 (67.9)	4,710 (48.0)	<0.001
Yes	5,610 (49.3)	498 (32.1)	5,112 (52.0)	
Ascites				
No	10,488 (96.6)	1,336 (95.5)	9,152 (96.8)	0.016
Yes	370 (3.4)	63 (4.5)	307 (3.2)	
BCLC stage				
0	451 (3.7)	47 (3.2)	404 (4.2)	<0.001
A	4,670 (39.6)	694 (47.7)	3,976 (41.4)	
B	4,018 (38.0)	518 (35.6)	3,500 (36.4)	
C	1,926 (18.7)	197 (13.5)	1,729 (18.0)	
Tumor thrombus				
No	6,812(65.0)	875 (68.1)	5,937 (64.5)	0.011
Yes	3,670(35.0)	409 (31.9)	3,261 (35.5)	
Tumor nodule				
Single	8,084 (80.1)	987 (84.2)	7,097 (79.6)	<0.001
Multiple	2,003 (19.9)	185 (15.8)	1,818 (20.4)	
Tumor diameter (cm)				
<3	1,644 (12.7)	175 (10.6)	1,298 (12.7)	0.007
≥3	9,950 (87.3)	1,469 (89.4)	8,652 (87.3)	
Tumor capsule				
Yes	7,129 (70.8)	830 (70.6)	6,299 (70.8)	0.867
No	2,942 (29.2)	346 (29.4)	2,596 (29.2)	
AFP (ng/ml)				
<20	4,417 (37.6)	800 (49.9)	3,617 (35.6)	<0.001
≥20	7,334 (62.4)	803 (50.1)	6,531 (64.4)	
Total bilirubin (μmol/L)				
≤23	9,483 (88.9)	1,152 (88.5)	8,331 (88.9)	0.635
>23	1,188 (11.1)	150 (11.5)	1,038 (11.1)	
Direct bilirubin (μmol/L)				
≤8	8,444 (81.5)	965 (81.4)	7,479 (81.5)	0.976
>8	1,921 (18.5)	220 (18.6)	1,701 (18.5)	
Albumin (g/L)				
≥40	6,555 (59.6)	912 (62.8)	5,643 (59.1)	0.007
<40	4,451 (40.4)	541 (37.2)	3,910 (40.9)	

Data are shown in n (%).

regression analysis indicated that HBV seropositivity (HR, 1.29; 95% CI, 1.10-1.21), incomplete tumor capsule (1.58; 1.41-1.77), vascular tumor thrombus (2.12; 1.90-2.36), tumor diameter (≥3 cm) (1.65; 1.29-2.12), more advanced BCLC stage (2.11; 1.85-2.41), AFP (≥20ng/ml) (1.69; 1.50-1.91), and direct bilirubin (>8μmol/L) (1.27; 1.11-1.45) independently contributed to shorter OS in HCC. Post-operative RFA (0.65; 0.53-0.79) and second resection (0.43; 0.34-0.55) significantly improved OS in HCC (**Table 5**).

In the survival analysis for ICC, the multivariate Cox regression analysis indicated that more advanced BCLC stage (1.72; 1.08-2.72) and AFP (≥20ng/ml) (1.58; 1.00-2.49) were independently associated with shorter OS, while reoperation (0.12; 0.02-0.89) was independently associated with longer OS in ICC (**Supplementary Table 3**).

Postoperative Survival of HCC Patients Solely Caused by HBV or HCV Infection

As HBV seropositivity independently increased the risk of OS in PLC in the multivariate Cox analysis, we evaluated the effect of HBV and HCV infection on the prognosis of HCC. Log-rank test was applied to compare the difference in OS between HBV-HCC and HCV-HCC. The result indicated that HBV-HCC patients

TABLE 3 | Comparison of demographic and clinical characteristics between HBV- and HCV-related HCC patients.

Variable	Total (n = 8,455)	Patients with HBV positive only (n = 8,230)	Patients with HCV positive only (n = 225)	*P*-value
Age (year)				
Medium (IQR)	52 (45-60)	52 (45-60)	64 (57-71)	
<40	864 (10.2)	861 (10.5)	3 (1.3)	<0.001
40–59	5,274 (62.4)	5,207 (63.3)	67 (29.8)	
≥60	2,317 (27.4)	2,162 (26.2)	155 (68.9)	
Gender				
Female	1,128 (13.3)	1,083 (13.2)	45 (20.0)	0.003
Male	7,327 (86.7)	7,147 (86.8)	180 (80.0)	
Cirrhosis				
No	3,906 (51.2)	3,838 (51.2)	68 (51.5)	0.945
Yes	3,720 (48.8)	3,656 (48.8)	64 (48.5)	
Ascites				
No	7,249 (97.9)	7,147 (98.0)	102 (94.4)	0.011
Yes	155 (2.1)	149 (2.0)	6 (5.6)	
BCLC stage				
0	364 (4.9)	357 (4.9)	7 (5.6)	<0.001
A	3,006 (40.5)	2,929 (40.1)	77 (61.1)	
B	2,926 (39.4)	2,892 (39.6)	34 (27.0)	
C	1,132 (15.2)	1,124 (15.4)	8 (6.3)	
Tumor thrombus				
No	4,659 (63.4)	4,583 (63.2)	76 (79.2)	0.001
Yes	2,689 (36.6)	2,669 (36.8)	20 (20.8)	
Tumor nodule				
Single	5,498 (81.1)	5,422 (81.0)	76 (85.4)	0.293
Multiple	1,285 (18.9)	1,272 (19.0)	13 (14.6)	
Tumor diameter (cm)				
<3	1,034 (13.9)	1,008 (13.8)	26 (20.6)	0.027
≥3	6,422 (86.1)	6,322 (86.2)	100 (79.4)	
AFP (ng/ml)				
<20	2,897 (36.6)	2,820 (36.3)	77 (49.0)	0.001
≥20	5,027 (63.4)	4,947 (63.7)	80(51.0)	
Tumor capsule				
Yes	5,183 (76.3)	5,112 (76.3)	71(79.8)	0.439
No	1,609 (23.7)	1,591 (23.7)	18 (20.2)	
Total bilirubin (μmol/L)				
≤23	6,475 (88.9)	6,394 (89.0)	81 (84.4)	0.152
>23	806 (11.1)	791 (11.0)	15 (15.6)	
Direct bilirubin (μmol/L)				
≤8	5,768 (81.7)	5,714 (81.8)	54 (73.0)	0.051
>8	1,293 (18.3)	1,273 (18.2)	20 (27.0)	
Albumin (g/L)				
≥40	4,338 (61.1)	4,339 (61.2)	49 (51.6)	0.056
<40	2,795 (38.9)	2,749 (38.8)	46 (48.4)	

Data are shown in n (%).

had an unfavorable prognosis compared to that of HCV-HCC (P<0.001) (**Figure 2A**). The PS matching with key baseline characteristics (age, gender, and BCLC stage) was applied to allow a common background for comparison between HBV-HCC and HCV-HCC, resulting in a matched sample size of 198 and 100, respectively. The result confirmed that HBV-HCC patients still had an unfavorable postoperative prognosis, compared to HCV-HCC patients (P<0.001) (**Figure 2B**).

TABLE 4 | Survival rate of HCC and ICC patients with different BCLC stage.

BCLC stage	Cases	1-year survival rate (%)	3-year survival rate (%)	5-year survival rate (%)	*P*-value
HCC					
0&A	3,343	92.4	77.2	64.8	<0.001[a]
B&C	3,697	69.8	40.5	29.2	
Subtotal	7,040	80.6	58.2	42.1	
ICC					
0&A	73	70.4	44.3	44.3	<0.001[a]
B&C	119	34.9	11.0	4.2	
Subtotal	192	48.7	24.2	18.5	

[a] *compared between 0&A stage and B&C stage.*

TABLE 5 | Univariate and multivariate Cox regression analysis for risk factors of overall survival in HCC.

Variable		No. (%) of participants (n =7257)	Univariate analysis		Multivariate analysis	
			HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Age	<40	859 (11.8)	1			
	40–59	4,324 (59.6)	1.19 (1.05-1.36)	0.009		
	≥60	2,074 (28.6)	1.10 (1.00-1.20)	0.044		
Gender	Female	936 (12.9)	1			
	Male	6,321 (87.1)	1.19 (1.06-1.34)	0.004		
HBV	Negative	1,104 (15.2)	1			
	Positive	6,136 (84.8)	1.50 (1.33-1.69)	<0.001	1.29 (1.10-1.21)	0.002
HCV	Negative	4,801 (97.2)	1			
	Positive	138 (2.8)	0.59 (0.43-0.80)	0.001		
Cirrhosis	No	3,296 (48.8)	1			
	Yes	3,459 (51.2)	0.96 (0.89-1.04)	0.325		
Ascites	No	6,402 (95.8)	1			
	Yes	279 (4.2)	0.93 (0.77-1.13)	0.481		
Tumor capsule	Yes	4,169 (68.6)	1			
	No	1,906 (31.4)	1.32 (1.21-1.43)	<0.001	1.58 (1.41-1.77)	<0.001
Tumor nodule	Single	4,821 (78.6)	1			
	Multiple	1,314 (21.4)	1.23 (1.12-1.35)	<0.001		
Tumor thrombus	No	4,215 (65.5)	1			
	Yes	2,222 (34.5)	3.20 (2.95-3.47)	<0.001	2.12 (1.90-2.36)	<0.001
BCLC stage	0&A	3,343 (47.5)	1			
	B&C	3,697 (52.5)	3.37 (3.10-3.68)	<0.001	2.11 (1.85-2.41)	<0.001
Tumor diameter (cm)	<3	919 (13.0)	1			
	≥3	6,150 (87.0)	2.85 (2.43-3.34)	<0.001	1.65 (1.29-2.12)	<0.001
AFP (ng/ml)	<20	2,548 (36.1)	1			
	≥20	4,504 (63.9)	2.17 (1.98-2.37)	<0.001	1.69(1.50-1.91)	<0.001
Total bilirubin (μmol/L)	≤23	5,756 (86.6)	1			
	>23	889 (13.4)	0.97 (0.86-1.10)	0.656		
Direct bilirubin (μmol/L)	≤8	5,048 (79.7)	1			
	>8	1,285 (20.3)	1.19 (1.07-1.31)	0.001	1.27(1.11-1.45)	<0.001
Albumin (g/L)	≥40	3,537 (53.4)	1			
	<40	3,082 (46.6)	0.96 (0.88-1.03)	0.258		
Post-operative TACE	No	1,744 (44.1)	1			
	Yes	2,210 (55.9)	1.02 (0.92-1.12)	0.717		
Post-operative RFA	No	3,596 (92.0)	1			
	Yes	312 (8.0)	0.57 (0.48-0.69)	<0.001	0.65(0.53-0.79)	<0.001
Reoperation	No	3,580 (92.0)	1			
	Yes	310 (8.0)	0.39 (0.32-0.49)	<0.001	0.43(0.34-0.55)	<0.001
Radiotherapy	No	3,751 (96.3)	1			
	Yes	145 (3.7)	1.05 (0.84-1.31)	0.668		
Chemotherapy	No	3,787 (94.6)	1			
	Yes	218 (5.4)	0.85 (0.66-1.09)	0.201		
Targeted therapy	No	3,712 (95.6)	1			
	Yes	171 (4.4)	1.06 (0.87-1.30)	0.555		

Prediction for Postoperative Prognosis in HCC

To evaluate if preoperative clinical parameters could predict postoperative prognosis of HCC, we developed a hazard risk prediction model consisting of independent preoperative prognostic factors. In the post-operative cohort of 7257 HCC patients, HCC patients were grouped randomly into a training cohort (n=3628) and a validation cohort (n=3629). All demographic and clinical characteristics were balanced between the two sub-cohorts (**Supplementary Table 4**). In the training

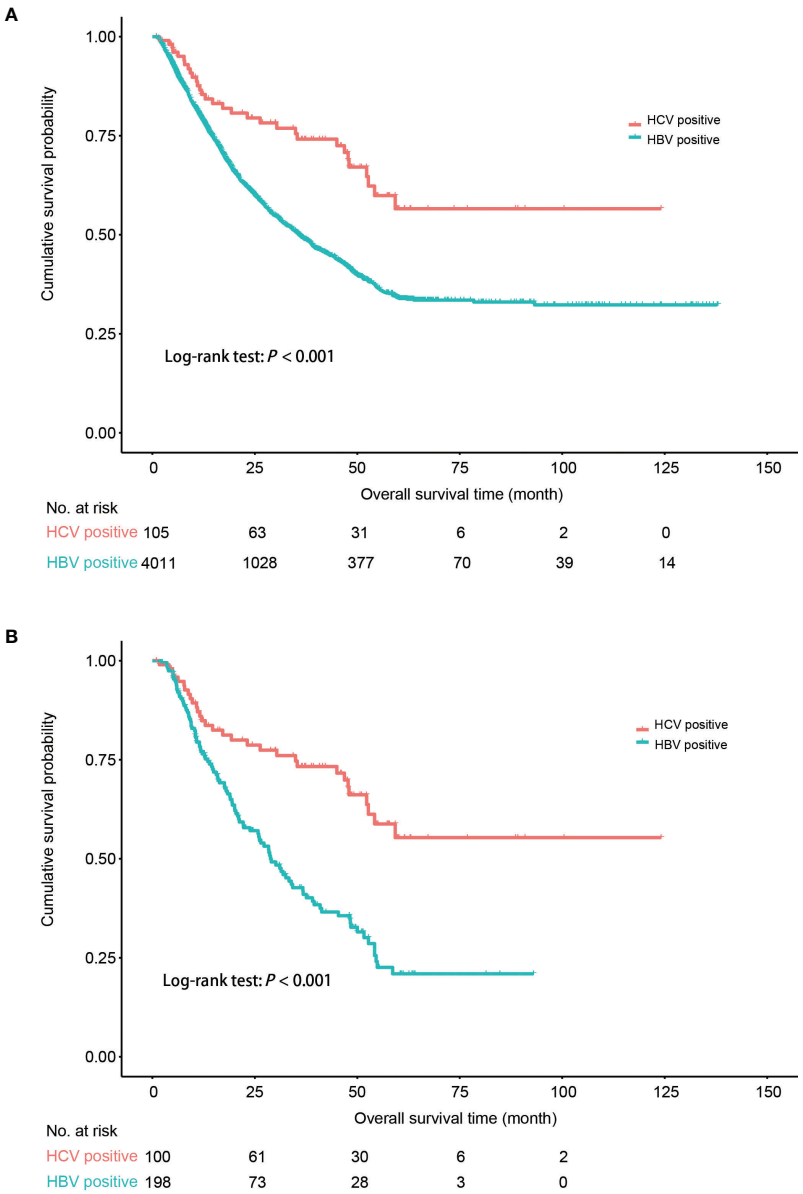


FIGURE 2 | Comparison of postoperative survival probability between HCC patients solely infected with HBV and HCC patients solely infected with HCV. **(A)** All the patients. **(B)** Patients following the 1:2 propensity score (PS) matching with age, gender, and BCLC stage. HBV, hepatitis B virus; HCV, hepatitis C virus; HCC, hepatocellular carcinoma. Kaplan–Meier curves were plotted to visualize the difference.

cohort, the multivariate Cox analysis indicated that age, gender, incomplete tumor capsule, vascular tumor thrombus, HBV positivity, tumor diameter, AFP, and advanced BCLC stage were independently related to OS in HCC. A nomogram composed of these factors is shown in **Figure 3A**. The C-index for the prediction of survival in the training cohort and the validation cohort was 0.735 (95% CI, 0.727–0.743) and 0.733 (95% CI, 0.725–0.741), respectively. Time-dependent ROC was applied to evaluate the power of the prognosis prediction model formulated in this study. The result indicated that the area under the curve (AUC) was 0.79 (95% CI, 0.77-0.81) for 1-year survival, 0.78 (0.76-0.80) for 3-year survival, and 0.75 (0.72-0.78) for 5-year survival in the training cohort (**Figure 3B**). In the validation cohort, the AUC for 1-, 3-, and 5-year survival was 0.79 (0.77-0.81), 0.77 (0.75-0.80), and 0.76 (0.73-0.78), respectively (**Figure 3C**). According to the medium of the risk score calculated from the established hazard risk prediction model, the training cohort was classified into a high-risk group and a low-risk group. The survival analysis showed that the low-risk group had better OS probability than the high-risk group (P<0.001). This grouping method was verified also in the validation cohort, and the result was similar to that of the training cohort (**Figures 3D, E**).

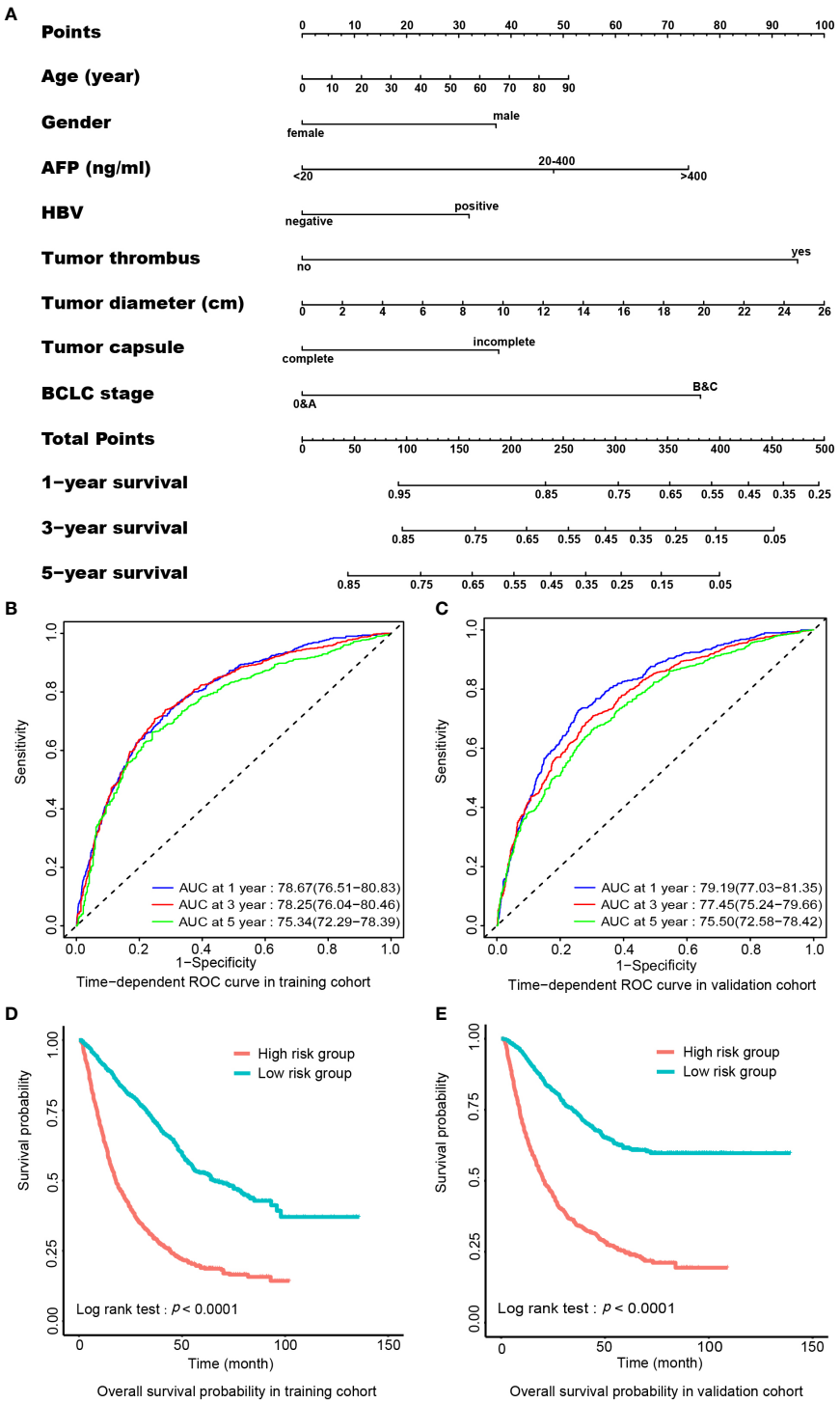


FIGURE 3 | Preoperative nomogram for predicting postoperative survival in HCC. **(A)** The nomogram. To use this nomogram, a patient's value is located on each variable axis, and a line represents the number of points received for each variable value. The sum of the score of each indicator is located on"Total Points" axis, and the total point represents the likelihood of postoperative survival of 1-, 3-, or 5-year shown on the survival axes. **(B)** AUC of time-dependent ROC curve for 1-, 3-, or 5-year survival in the training cohort. **(C)** AUC of time-dependent ROC curve for 1-, 3-, or 5-year survival in the validation cohort. **(D)** Comparison of OS probability between low- and high-risk groups according to total points from nomogram in the training cohort. **(E)** Comparison of overall survival probability between low- and high-risk group according to total points from nomogram in the validation cohort. AFP, α-fetoprotein; AUC, area under the curve; BCLC, Barcelona Clinic Liver Cancer; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; OS, overall survival; ROC, receiver operating characteristics.

The calibration plot was applied to evaluate the fitting degree of survival probability between the actual observation and prediction value calculated by the nomogram constructed in this study. The results displayed good fitness in the probability of 1-, 3-, and 5-year survivals both in the training cohort (**Supplementary Figures 3A, C, E**) and validation cohort (**Supplementary Figures 3B, D, F**).

DISCUSSION

In this study, we selected four medical centers distributed in the north, south, west, and east to represent the demographic, epidemiological, and clinical characteristics of PLC in China. The enrolled patients were mostly living in the provinces where study hospitals were located and near areas though they were from almost all provincial administrative regions of mainland China. The socioeconomic situations, living styles, and living environments of people from the four regions are mutually exclusive and different. The seropositivity of HBV in HCC patients was higher in the south and east than in the north and west. This is only partially coincident with the trend in their background HBV infection, as the prevalence of chronic HBV infection is lower in the north than in the south and east (21). The seropositivity of HBV in HCC in the west was lower than that in the east, although the prevalence of HBV was higher in the west than in the east (21). This is possibly because NAFLD, which is projected to become the leading cause of HCC in many countries (22), was more prevalent in the west (33.8%) than the other three regions (23). In addition, consumption of salted food containing N-nitroso compound and drinking water containing a high content of nitrate and nitrite were evident in Wuwei, Gansu, China (24). As the prevalence of PLC is closely related to geographic areas, socioeconomic state, and risk factor exposure, enrolled patients from the medical centers in the four regions should be highly representative among the current studies to characterize the overall risk factors, histotype composition, and prognostic factors in mainland China. We conclude that the seropositivity of HBV is 80.1% in PLC and 84.4% in HCC in mainland China; HCC, ICC, and CHC account for 93.0%, 4.3%, and 1.6% in PLC, respectively.

In this study, we found that the proportion of HCC in PLC increased consecutively from 89.70% before 2008 to 98.75% in 2016-2020, indicating the proportion of cholangiocarcinoma in PLC decreased correspondingly. Exposure to the risk factor of cholangiocarcinoma including liver fluke infection consecutively decreased. HBV seropositivity increased consecutively in PLC, but decreased in HCC, indicating that chronic HBV infection contributed increasingly to the occurrence of cholangiocarcinoma. It is possibly because HBV integration has been identified in 71.43% of ICCs (25). In this study, we found that the seropositivity of anti-HCV antibodies decreased consecutively, either in PLC or in HCC, although the incidence of HCV infection increased from 0.7 to 15.0 cases per 100,000 persons from 1997 to 2012 (26). This is possibly due to the fact that symptomatic hepatitis C has been treated in China over the past 20 years with medical insurance-covered interferon-α and ribavirin. In addition to interferon-α and ribavirin, direct-acting antivirals have fundamentally changed HCV-caused liver diseases, due to their high efficacy and tolerability (27, 28). HCC only derives from a diseased liver. HCV-induced carcinogenesis should be indirectly induced *via* multiple steps from chronic hepatitis to fibrosis, advanced fibrosis, and cirrhosis with somatic genetic/epigenetic alterations (27). In 1998, China enacted the blood donation law to strengthen the supervision of blood collection, organization, source management, and use of disposable syringes. Thereafter, the prevalence of hepatitis C in China decreased. These data support the observation of this study.

HBV infection led to an 8 year earlier onset in HCC, compared to HBV-free HCC. HBV-HCC patients had higher proportions of positive AFP, liver cirrhosis, advanced BCLC stage, multiple tumor nodules, and vascular tumor thrombus, indicating that HBV not only promotes the occurrence of HCC, but also promotes the recurrence of HCC. AFP can be upregulated by HBV X protein, which plays an important role in the aggressiveness of HCC by promoting HCC cells into stem cells and by activating the PI3K/mTOR signaling pathway (29). Liver cirrhosis is the result of an immune response to hepatic injury caused by chronic inflammation (30). HBV, especially its integrated forms in the human genome and its evolved forms generated in the long-term process of chronic infection, directly promotes the development of HCC (31–34). HBV replication, integration, and evolution also improve the recurrence and metastasis of HCC while long-term treatment of chronic HBV infection can reduce the development and postoperative recurrence of HBV-HCC (15, 35–39). As HCV-HCC is very rare in China, the large sample size in this study allows for identifying the difference in the clinical characteristics between HCV-HCC and HBV-HCC. Compared to HBV-HCC patients, HCV-HCC patients were 12 years older and had a lower proportion of some parameters indicating the aggressiveness and metastasis of HCC. As the BCLC stage is the major prognostic factor in HCC, the PS matching with age, gender, and BCLC stage was applied to allow a common background for comparison. The prognosis of HCV-HCC was also proven to be significantly better than that of HBV-HCC. The mechanism of HBV- and HCV-induced hepatocarcinogenesis should be different. HCV itself might not be directly oncogenic. Chronic HCV infection causes hepatic inflammation, necrosis, metabolic disorders, steatosis, regeneration, and cirrhosis, thus facilitating the development of HCC by creating an immunosuppressive tumorigenic environment and activating cancer stem-like cells by proinflammatory factors like plasminogen activator inhibitor-1 (40, 41). Other non-B and non-C risk factors including diabetes and NAFLD might also facilitate the development of HCC, mostly in elderly patients, and the overall survival rate is significantly better than that of HBV-HCC (42), possibly by inducing systemic and hepatic inflammation. Thus, HBV is directly carcinogenic. HBV replication, viral mutation, and integration into the host genome promote the development and progression of HCC. The non-HBV etiological factors including HCV infection promote the development of HCC

mostly by inducing non-resolving inflammation which leads to the development of tumors by promoting proliferative and survival signaling, inducing instability of genome, and subsequent angiogenesis. Of note, non-resolving inflammation is also an important factor for the development and progression of HBV-HCC. These data clearly indicate that HBV is more carcinogenic than HCV orany other cause. Antiviral treatment is effective in decreasing the occurrence and postoperative recurrence of HCC in HBV-infected patients (15, 43). Thus, the prophylactic and therapeutic effects of anti-HBV treatment on the development of HCC should be added into the current clinical guidelines.

In this study, we showed that the overall 1 -, 3 -, and 5-year OS rates of HCC patients were 80.7%, 58.3%, and 42.1%, respectively. Importantly, the survival rates decrease with increasing BCLC stages (**Table 4**). The 5-year OS rate of HCC patients in the early BCLC stage (stage 0 and A) was 64.8%. It has been shown that the 5-year survival rate of HCC patients at early BCLC (stage 0/A) is 60% to 90% (44–47). The 5-year OS rates of patients enrolled in Shanghai were lower than those in Nanning and Beijing, possibly because the BCLC stage of HCC patients enrolled in Shanghai was more advanced than those enrolled in the other two cities (**Table 1**). Shanghai is usually the last station of the medical tour for patients to seek the best treatments because Shanghai has the top medical facilities in HCC surgery. Surprisingly, we found that the overall 1 -, 3 -, and 5-year OS rates increased consecutively in HCC, which is consistent with the increasing proportions of early-stage HCC (**Figure 1**). The outcomes of this study strongly suggest the necessity of timely screening for HCC in the high-risk population, especially the high-risk HBV-infected subjects who carry a high viral load and HCC-risk HBV mutations (15, 34–36), to increase the detection of HCC at an early stage.

In this study, we demonstrated that HBV seropositivity, AFP, incomplete tumor capsule, tumor diameter, advanced BCLC stage, and vascular tumor thrombus were independently associated with an unfavorable prognosis in HCC by the multivariate regression analysis. These factors are measurable before surgery. We also developed a nomogram composed of these preoperative prognostic factors and confirmed that this nomogram was able to accurately predict an unfavorable postoperative prognosis in HCC, even for the 5-year OS rate (**Figure 3**). The high accuracy of the nomogram established by preoperative parameters of the large sample size for the prediction of postoperative prognosis suggests that this nomogram should be extremely important for clinical decision-making. Predictors in our prognosis-prediction nomogram differ a little from the current predictors in HCC. The predictors of reported nomograms for HCC prognosis were mostly extracted after surgical treatment, including MVI and resection margin (48, 49). Development of a nomogram to preoperatively estimate postoperative survival is also reported (50). However, the predictor is different from our study. We identified predictors from presurgical parameters including HBV seropositivity to determine if patients were suitable for surgical treatment. The Shanghai score, one of the clinical stages of liver cancer that used HBV information as predictors, has 14 predictors (51). Our nomogram has eight predictors that are easy to be used clinically. We also confirmed that post-operative RFA and reoperation after recurrence independently increased OS in HCC (**Table 5**). Thus, post-operative RFA and reoperation after recurrence should be options to improve the therapeutic regimen.

In this study, we also found that HBV infection was associated with 5-year earlier onset and higher AFP, liver cirrhosis, advanced BCLC stage, and vascular tumor thrombus in ICC. Published studies indicated that antiviral treatment decreased the risk and prolonged long-term survival in ICC (52, 53). This line of evidence reflects the role of HBV in generating inflammatory background from which ICC develops. In addition, HBV integration is frequently identified in ICC and other cancer types including non-Hodgkin lymphoma (24, 54). Significant associations of HBV seropositivity with leukemia, extrahepatic bile duct carcinoma, esophageal cancer, stomach cancer, and pancreatic cancer are also suggested (55). We hypothesize that weak antiviral and anti-cancer immunity predisposed by genetic and environmental exposure arouse cancer-promoting non-resolving inflammation, which facilitates the development of ICC and extrahepatic cancers.

There are several limitations in our study. First, the relatively weak risk factors of PLC, namely NAFLD, diabetes, nonalcoholic steatohepatitis, alcoholic liver disease, aflatoxin exposure, liver fluke exposure, family history, cigarette smoking, and alcohol consumption (5–7, 56) were not included because these data were incomplete in medical records. The contributions of these risk factors to HBV- or HCV-related HCC remain unknown. Second, clinical data related to curative surgery, pathological examination, and follow-up in Wuwei (the west) did not meet the criteria, and were therefore not included in the analysis, resulting in a loss of data. Third, compared to patients who did not join the follow-up study, patients who were followed up had a higher proportion of HBV positivity, high AFP (≥20ng/ml), poor liver function, multiple tumor nodules, incomplete tumor capsules, and late BCLC stage (**Supplementary Table 5**). These factors were mostly associated with an unfavorable prognosis in HCC. The postoperative survival might be underestimated. Fourth, ICC and CHC were not analyzed for prognosis prediction because of small sample sizes.

Conclusively, chronic HBV infection contributes to 84.4% of HCC in mainland China. HBV infection not only induces hepatocarcinogenesis but also promotes the aggressiveness of HCC. HCV-HCC onset is 12 years later than HBV-HCC and has a better prognosis than HBV-HCC. A significant postoperative survival benefit is obtained in patients at the early BCLC stage, highlighting the importance of screening for early-stage HCC in high-risk populations. Our prognosis prediction model constructed with preoperational parameters is important for clinical decision making.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Second Military Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

GC contributed to study design and supervision. HY, XB, WZ, WT, JC, and ZS contributed to patient enrolment, acquisition of data, and follow-up. HZ, JL, JY, MH, PZ, and JW conducted data organization, statistical analysis, and data interpretation. XL, CQ, MW, KL, YW, and ZZ contributed to data collection and follow-up. GC wrote the manuscript. All authors had access to the data and approved the final version of this manuscript.

FUNDING

This study was supported by the National Key Basic Research Program of China (973 Program) (grant numbers 2015CB554000 to GC), Key discipline from the "3-year public health promotion" program of Shanghai Municipal Health Commission to GC (GWV-10.1-XK17), the State Key Infection Disease Project of China (grant numbers 2017ZX10201201-006-001 to HZ), the National Natural Science Foundation of China (grant numbers 81520108021, 91529305,and 81673250 to GC), the Key Research and Development Project of Guangxi (grant number AB18050020 to HY and grant number AA18221001 to WT), the Natural Science Foundation of Guangxi (grant number 2018GXNSFDA050012 to HY), and Shanghai Wu Mengchao Medical Science Foundation (grant number JJHXM-2019042 to HY).

ACKNOWLEDGMENTS

We would like to thank the persons who joined the investigation and data collection. All the authors approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.906778/full#supplementary-material

Supplementary Figure 1 | Diagram of patients enrolled in this study.

Supplementary Figure 2 | The geographic distribution of the patients with primary liver cancer in the study.

Supplementary Figure 3 | The calibration curve for predicting postoperative survival with preoperative nomogram in HCC. **(A)** The calibration curve for predicting postoperative 1-year survival in the training cohort. **(B)** The calibration curve for predicting postoperative 1-year survival in the validation cohort. **(C)** The calibration curve for predicting postoperative 3-year survival in the training cohort. **(D)** The calibration curve for predicting postoperative 3-year survival in the validation cohort. **(E)** The calibration curve for predicting postoperative 5-year survival in the training cohort. **(F)** The calibration curve for predicting postoperative 5-year survival in the validation cohort.

REFERENCES

1. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The Ever-Increasing Importance of Cancer as a Leading Cause of Premature Death Worldwide. *Cancer* (2021) 127(16):3029–30. doi: 10.1002/cncr.33587
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: Cancer J Clin* (2021) 71 (3):209–49. doi: 10.3322/caac.21660
3. Wang S, Du X, Han X, Yang F, Zhao J, Li H, et al. Influence of Socioeconomic Events on Cause-Specific Mortality in Urban Shanghai, China, From 1974 to 2015: A Population-Based Longitudinal Study. *CMAJ Can Med Assoc J = J l'Association medicale Can* (2018) 190(39):E1153–e61. doi: 10.1503/cmaj.180272
4. Jiang D, Zhang L, Liu W, Ding Y, Yin J, Ren R, et al. Trends in Cancer Mortality in China From 2004 to 2018: A Nationwide Longitudinal Study. *Cancer Commun (London England)* (2021) 41(10):1024–36. doi: 10.1002/cac2.12195
5. Maucort-Boulch D, de Martel C, Franceschi S, Plummer M. Fraction and Incidence of Liver Cancer Attributable to Hepatitis B and C Viruses Worldwide. *Int J Cancer* (2018) 142(12):2471–7. doi: 10.1002/ijc.31280
6. Sagnelli E, Macera M, Russo A, Coppola N, Sagnelli C. Epidemiological and Etiological Variations in Hepatocellular Carcinoma. *Infection* (2020) 48(1):7–17. doi: 10.1007/s15010-019-01345-y
7. Rizvi S, Gores GJ. Pathogenesis, Diagnosis, and Management of Cholangiocarcinoma. *Gastroenterology* (2013) 145(6):1215–29. doi: 10.1053/j.gastro.2013.10.013
8. Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA Oncol* (2017) 3(12):1683–91. doi: 10.1001/jamaoncol.2017.3055
9. Yang F, Ma L, Yang Y, Liu W, Zhao J, Chen X, et al. Contribution of Hepatitis B Virus Infection to the Aggressiveness of Primary Liver Cancer: A Clinical Epidemiological Study in Eastern China. *Front Oncol* (2019) 9:370. doi: 10.3389/fonc.2019.00370
10. Wang M, Wang Y, Feng X, Wang R, Wang Y, Zeng H, et al. Contribution of Hepatitis B Virus and Hepatitis C Virus to Liver Cancer in China North Areas: Experience of the Chinese National Cancer Center. *Int J Infect Dis* (2017) 65:15–21. doi: 10.1016/j.ijid.2017.09.003
11. Mgaieth S, Kemp W, Gow P, Fink M, Lubel J, Nicoll A, et al. Impact of Viral Hepatitis Aetiology on Survival Outcomes in Hepatocellular Carcinoma: A Large Multicentre Cohort Study. *J Viral hepatit* (2017) 24(11):982–9. doi: 10.1111/jvh.12717
12. Younossi ZM, Stepanova M, Saab S, Ahmed A, Lam B, Srishord M, et al. The Impact of Viral Hepatitis-Related Hepatocellular Carcinoma to Post-Transplant Outcomes. *J Viral hepatit* (2016) 23(1):53–61. doi: 10.1111/jvh.12449
13. Li Q, Li H, Qin Y, Wang PP, Hao X. Comparison of Surgical Outcomes for Small Hepatocellular Carcinoma in Patients With Hepatitis B Versus Hepatitis C: A Chinese Experience. *J Gastroenterol Hepatol* (2007) 22 (11):1936–41. doi: 10.1111/j.1440-1746.2006.04619.x
14. Sinn DH, Gwak GY, Cho J, Paik SW, Yoo BC. Comparison of Clinical Manifestations and Outcomes Between Hepatitis B Virus- and Hepatitis C Virus-Related Hepatocellular Carcinoma: Analysis of a Nationwide Cohort. *PLoS One* (2014) 9(11):e112184. doi: 10.1371/journal.pone.0112184
15. Yin J, Li N, Han Y, Xue J, Deng Y, Shi J, et al. Effect of Antiviral Treatment With Nucleotide/Nucleoside Analogs on Postoperative Prognosis of Hepatitis

B Virus-Related Hepatocellular Carcinoma: A Two-Stage Longitudinal Clinical Study. *J Clin Oncol* (2013) 31(29):3647–55. doi: 10.1200/jco.2012.48.5896
16. Ueno M, Hayami S, Shigekawa Y, Kawai M, Hirono S, Okada K, et al. Prognostic Impact of Surgery and Radiofrequency Ablation on Single Nodular HCC ≤5 Cm: Cohort Study Based on Serum Hcc Markers. *J Hepatol* (2015) 63 (6):1352–9. doi: 10.1016/j.jhep.2015.07.013
17. Yin L, Li H, Li AJ, Lau WY, Pan ZY, Lai EC, et al. Partial Hepatectomy Vs. Transcatheter Arterial Chemoembolization for Resectable Multiple Hepatocellular Carcinoma Beyond Milan Criteria: A Rct. *J Hepatol* (2014) 61(1):82–8. doi: 10.1016/j.jhep.2014.03.012
18. Llovet JM, Brú C, Bruix J. Prognosis of Hepatocellular Carcinoma: The BCLC Staging Classification. *Semin liver Dis* (1999) 19(3):329–38. doi: 10.1055/s-2007-1007122
19. Harrell FE Jr, Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Stat Med* (1996) 15(4):361–87. doi: 10.1002/(sici)1097-0258(19960229)15:4<361::Aid-sim168>3.0.Co;2-4
20. Frank E, Harrell Jr. Rms: Regression Modeling Strategies, in: *R Package Version 3.4-0* (Accessed May 8, 2019).
21. Wang H, Men P, Xiao Y, Gao P, Lv M, Yuan Q, et al. Hepatitis B Infection in the General Population of China: A Systematic Review and Meta-Analysis. *BMC Infect Dis* (2019) 19(1):811. doi: 10.1186/s12879-019-4428-y
22. Ioannou GN. Epidemiology and Risk-Stratification of NAFLD-Associated HCC. *J Hepatol* (2021) 75(6):1476–84. doi: 10.1016/j.jhep.2021.08.012
23. Zhou F, Zhou J, Wang W, Zhang XJ, Ji YX, Zhang P, et al. Unexpected Rapid Increase in the Burden of NAFLD in China From 2008 to 2018: A Systematic Review and Meta-Analysis. *Hepatol (Baltimore Md)* (2019) 70(4):1119–33. doi: 10.1002/hep.30702
24. Luo HZ, Mi DH, Li YG, Su CY. Investigation of Risk Factors in the Liver Cancer of Wuwei City. *Cancer Res Clinic* (2010) 22(2):134–7. doi: 10.3760/cma.j.issn.1006-9801.2010.02.025
25. Li M, Du M, Cong H, Gu Y, Fang Y, Li J, et al. Characterization of Hepatitis B Virus DNA Integration Patterns in Intrahepatic Cholangiocarcinoma. *Hepatol Res* (2021) 51(1):102–15. doi: 10.1111/hepr.13580
26. Qin Q, Smith MK, Wang L, Su Y, Wang L, Guo W, et al. Hepatitis C Virus Infection in China: An Emerging Public Health Issue. *J Viral hepatit* (2015) 22 (3):238–44. doi: 10.1111/jvh.12295
27. Lee T-P, Bernstein D. The Impact of Direct-Acting Antivirals on Hepatitis C Associated Hepatocellular Carcinoma. *Hepatoma Res* (2020) 6:21. doi: 10.20517/2394-5079.2019.44
28. Yang M, Ma R, Huang Y, Wei L. Impact of Direct-Acting Antivirals on *de novo* Occurrence of Hepatocellular Carcinoma in Hepatitis C Virus Patients. *Hepatoma Res* (2020) 6:31. doi: 10.20517/2394-5079.2019.45
29. Zhu M, Li W, Lu Y, Dong X, Lin B, Chen Y, et al. HBx Drives Alpha Fetoprotein Expression to Promote Initiation of Liver Cancer Stem Cells Through Activating Pi3k/Akt Signal Pathway. *Int J Cancer* (2017) 140 (6):1346–55. doi: 10.1002/ijc.30553
30. Marcellin P, Asselah T. Long-Term Therapy for Chronic Hepatitis B: Hepatitis B Virus DNA Suppression Leading to Cirrhosis Reversal. *J Gastroenterol Hepatol* (2013) 28(6):912–23. doi: 10.1111/jgh.12213
31. Pu R, Liu W, Zhou X, Chen X, Hou X, Cai S, et al. The Effects and Underlying Mechanisms of Hepatitis B Virus X Gene Mutants on the Development of Hepatocellular Carcinoma. *Front Oncol* (2022) 12:836517. doi: 10.3389/fonc.2022.836517
32. Gao Q, Zhu H, Dong L, Shi W, Chen R, Song Z, et al. Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* (2019) 179(2):561–77.e22. doi: 10.1016/j.cell.2019.08.052
33. Chauhan R, Michalak TI. Earliest Hepatitis B Virus-Hepatocyte Genome Integration: Sites, Mechanism, and Significance in Carcinogenesis. *Hepatoma Res* (2021) 7:20. doi: 10.20517/2394-5079.2020.136
34. Yin J, Xie J, Liu S, Zhang H, Han L, Lu W, et al. Association Between the Various Mutations in Viral Core Promoter Region to Different Stages of Hepatitis B, Ranging of Asymptomatic Carrier State to Hepatocellular Carcinoma. *Am J Gastroenterol* (2011) 106(1):81–92. doi: 10.1038/ajg.2010.399
35. Yin J, Chen X, Li N, Han X, Liu W, Pu R, et al. Compartmentalized Evolution of Hepatitis B Virus Contributes Differently to the Prognosis of Hepatocellular Carcinoma. *Carcinogenesis* (2021) 42(3):461–70. doi: 10.1093/carcin/bgaa127
36. Chen X, Zhang M, Li N, Pu R, Wu T, Ding Y, et al. Nucleotide Variants in Hepatitis B Virus Pres Region Predict the Recurrence of Hepatocellular Carcinoma. *Aging* (2021) 13(18):22256–75. doi: 10.18632/aging.203531
37. Lin C-L, Kao J-H. Prevention of Hepatitis B Virus-Related Hepatocellular Carcinoma. *Hepatoma Res* (2021) 7:9. doi: 10.20517/2394-5079.2020.125
38. Kong J, Liang X, Zhang J, Zeng J, Liu J, Zeng J. Antiviral Therapy Improves Survival in Hepatocellular Carcinoma With Microvascular Invasion: A Propensity Score Analysis. *Digest Dis Sci* (2021). doi: 10.1007/s10620-021-07248-z
39. Hou JL, Zhao W, Lee C, Hann HW, Peng CY, Tanwandee T, et al. Outcomes of Long-Term Treatment of Chronic HBV Infection With Entecavir or Other Agents From a Randomized Trial in 24 Countries. *Clin Gastroenterol Hepatol* (2020) 18(2):457–67.e21. doi: 10.1016/j.cgh.2019.07.010
40. Khatun M, Ray R, Ray RB. Hepatitis C Virus Associated Hepatocellular Carcinoma. *Adv Cancer Res* (2021) 149:103–42. doi: 10.1016/bs.acr.2020.10.003
41. Nam DE, Angelucci A, Choi D, Leigh A, Seong HC, Hahn YS. Elevation of Plasminogen Activator Inhibitor-1 Promotes Differentiation of Cancer Stem-Like Cell State by Hepatitis C Virus Infection. *J Virol* (2021) 24;95(10): e02057–20. doi: 10.1128/jvi.02057-20
42. Xue X, Liao W, Xing Y. Comparison of Clinical Features and Outcomes Between HBV-Related and Non-B Non-C Hepatocellular Carcinoma. *Infect Agents Cancer* (2020) 15:11. doi: 10.1186/s13027-020-0273-2
43. Yin J, Wang J, Pu R, Xin H, Li Z, Han X, et al. Hepatitis B Virus Combo Mutations Improve the Prediction and Active Prophylaxis of Hepatocellular Carcinoma: A Clinic-Based Cohort Study. *Cancer Prev Res (Philadelphia Pa)* (2015) 8(10):978–88. doi: 10.1158/1940-6207.Capr-15-0160
44. Zhuang H, Zhou Z, Ma Z, Huang S, Gong Y, Zhang Z, et al. Prognostic Stratification Based on a Novel Nomogram for Solitary Large Hepatocellular Carcinoma After Curative Resection. *Front Oncol* (2020) 10:556489. doi: 10.3389/fonc.2020.556489
45. Hsiao CY, Hu RH, Ho CM, Wu YM, Lee PH, Ho MC. Surgical Resection Versus Radiofrequency Ablation for Barcelona Clinic Liver Cancer Very Early Stage Hepatocellular Carcinoma: Long-Term Results of a Single-Center Study. *Am J Surg* (2020) 220(4):958–64. doi: 10.1016/j.amjsurg.2020.03.017
46. Tsilimigras DI, Bagante F, Sahara K, Moris D, Hyer JM, Wu L, et al. Prognosis After Resection of Barcelona Clinic Liver Cancer (BCLC) Stage 0, A, and B Hepatocellular Carcinoma: A Comprehensive Assessment of the Current BCLC Classification. *Ann Surg Oncol* (2019) 26(11):3693–700. doi: 10.1245/s10434-019-07580-9
47. Liu H, Wang ZG, Fu SY, Li AJ, Pan ZY, Zhou WP, et al. Randomized Clinical Trial of Chemoembolization Plus Radiofrequency Ablation Versus Partial Hepatectomy for Hepatocellular Carcinoma Within the Milan Criteria. *BJS (Brit J Surg* (2016) 103(4):348–56. doi: 10.1002/bjs.10061
48. Liang L, Li C, Wang MD, Wang H, Zhou YH, Zeng YY, et al. Development and Validation of a Novel Online Calculator for Estimating Survival Benefit of Adjuvant Transcatheter Arterial Chemoembolization in Patients Undergoing Surgery for Hepatocellular Carcinoma. *J Hematol Oncol* (2021) 14(1):165. doi: 10.1186/s13045-021-01180-5
49. Ma L, Deng K, Zhang C, Li H, Luo Y, Yang Y, et al. Nomograms for Predicting Hepatocellular Carcinoma Recurrence and Overall Postoperative Patient Survival. *Front Oncol* (2022) 12:843589. doi: 10.3389/fonc.2022.843589
50. Wang YY, Xiang BD, Ma L, Zhong JH, Ye JZ, Wang K, et al. Development and Validation of a Nomogram to Preoperatively Estimate Post-Hepatectomy Liver Dysfunction Risk and Long-Term Survival in Patients With Hepatocellular Carcinoma. *Ann Surg* (2021) 274(6):e1209–e17. doi: 10.1097/sla.0000000000003803
51. Sun HC, Xie L, Yang XR, Li W, Yu J, Zhu XD, et al. Shanghai Score: A Prognostic and Adjuvant Treatment-Evaluating System Constructed for Chinese Patients With Hepatocellular Carcinoma After Curative Resection. *Chin Med J* (2017) 130(22):2650–60. doi: 10.4103/0366-6999.218019
52. Lee TY, Hsu YC, Yu SH, Lin JT, Wu MS, Wu CY. Effect of Nucleos(T)Ide Analogue Therapy on Risk of Intrahepatic Cholangiocarcinoma in Patients With Chronic Hepatitis B. *Clin Gastroenterol Hepatol* (2018) 16(6):947–54.e4. doi: 10.1016/j.cgh.2017.09.031
53. Lei Z, Xia Y, Si A, Wang K, Li J, Yan Z, et al. Antiviral Therapy Improves Survival in Patients With HBV Infection and Intrahepatic Cholangiocarcinoma Undergoing Liver Resection. *J Hepatol* (2018) 68 (4):655–62. doi: 10.1016/j.jhep.2017.11.015

54. Li M, Shen Y, Chen Y, Gao H, Zhou J, Wang Q, et al. Characterization of Hepatitis B Virus Infection and Viral DNA Integration in Non-Hodgkin Lymphoma. *Int J Cancer* (2020) 147(8):2199–209. doi: 10.1002/ijc.33027
55. Tian T, Song C, Jiang L, Dai J, Lin Y, Xu X, et al. Hepatitis B Virus Infection and the Risk of Cancer Among the Chinese Population. *Int J Cancer* (2020) 147(11):3075–84. doi: 10.1002/ijc.33130
56. Zampaglione L, Ferrari J, Pedica F, Goossens N. HCC in Metabolic Syndrome: Current Concepts and Future Directions. *Hepatoma Res* (2021) 7:55. doi: 10.20517/2394-5079.2021.22

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fonc.2022.895479

Associations of Pre-Diagnostic Serum Levels of Total Bilirubin and Albumin With Lung Cancer Risk: Results From the Southern Community Cohort Study

*Hyung-Suk Yoon, Xiao-Ou Shu, Chris Shidal, Jie Wu, William J. Blot, Wei Zheng and Qiuyin Cai**

Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, United States

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Aleš Dvořák, Charles University, Czechia
Jian Gu, University of Texas MD Anderson Cancer Center, United States

***Correspondence:**
Qiuyin Cai
qiuyin.cai@vanderbilt.edu

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *13 March 2022*
Accepted: *26 April 2022*
Published: *23 June 2022*

Citation:
Yoon HS, Shu XO, Shidal C, Wu J, Blot WJ, Zheng W and Cai Q (2022) Associations of Pre-Diagnostic Serum Levels of Total Bilirubin and Albumin With Lung Cancer Risk: Results From the Southern Community Cohort Study. Front. Oncol. 12:895479. doi: 10.3389/fonc.2022.895479

Background: Previous studies conducted among European and Asian decedents reported inverse associations of serum total bilirubin and albumin with lung cancer risk. Yet, no study has been conducted among African Americans or low-income European Americans.

Methods: This study included 522 incident lung cancer cases and 979 matched controls nested in the Southern Community Cohort Study, a cohort of predominantly low-income African and European Americans. Serum levels of total bilirubin and albumin, collected up to 11 years prior to case diagnoses, were measured by a clinical chemistry analyzer. Conditional logistic regression models were applied to evaluate the associations of total bilirubin and albumin with lung cancer risk.

Results: Overall, serum levels of total bilirubin ($OR_{T3\ vs.\ T1}$ = 0.96, 95% CI: 0.66-1.39) were not significantly associated with lung cancer risk. However, higher levels of serum total bilirubin were significantly associated with decreased risk of lung cancer among participants who were diagnosed within two years following sample collection ($OR_{T3\ vs.\ T1}$ = 0.36, 95% CI: 0.15-0.87) and among former/never smokers ($OR_{T3\ vs.\ T1}$ = 0.54, 95% CI: 0.32-0.93). Serum levels of albumin were significantly associated with decreased risk of lung cancer overall ($OR_{T3\ vs.\ T1}$ = 0.70, 95% CI: 0.50-0.98) and among African Americans ($OR_{T3\ vs.\ T1}$ = 0.62, 95% CI: 0.41-0.96), but not among European Americans.

Conclusion: Our results indicate that in a low-income African American and European American population, serum levels of total bilirubin may be related to lung cancer progression and differ by smoking status. Meanwhile, the association of serum albumin levels with lung cancer risk may differ by race. Further studies are warranted to confirm these results.

Keywords: lung cancer, total bilirubin (TB), albumin (ALB), smoking, Southern Community Cohort Study low-income populations, African Americans

INTRODUCTION

Lung cancer is the third most common type of cancer, and its incidence varies by race/ethnicity and geographical location (1, 2). Cigarette smoking is the primary risk factor for lung cancer, but the mechanisms of lung carcinogenesis are still not well understood (3).

Antioxidants are free radical scavengers that can prevent damage from excessive reactive oxygen species (4) and may act as protective factors in carcinogenic progress (5). Bilirubin and albumin have attracted increasing attention due to their potent antioxidant properties, which may help prevent cancer by reducing inflammation and tumor cell proliferation (6–10). Bilirubin is produced as a byproduct of heme degradation and mostly transported in blood tightly bound to albumin, but a very small fraction of bilirubin remains unbound, *i.e.*, free bilirubin, which has biological effects (11, 12). Several prospective studies investigating the associations of total bilirubin and albumin levels with lung cancer risk have been conducted (13–16), and a meta-analysis indicated an inverse association between total bilirubin levels and risk of lung cancer (17). However, this epidemiological evidence came mainly from studies conducted among European and Asian descendants. No study has been conducted among African Americans or socioeconomically disadvantaged populations who were experiencing elevated lung cancer rates (18).

In this study, we investigated the associations of pre-diagnostic serum levels of total bilirubin and albumin with lung cancer risk in the Southern Community Cohort Study (SCCS), a cohort study comprised of a large proportion of low-income African and European Americans. We further investigated whether the associations were modified by race, smoking status, and histological subtype.

METHODS

Study Population

The SCCS is a prospective cohort study designed to investigate the underlying causes of racial disparities in health outcomes. Detailed information on the SCCS is described in previous literature (19, 20). All study participants signed written informed consent before enrollment into the SCCS. Approximately 86% of the participants were recruited at community health centers (CHCs) that provide primary health care services for the primarily low-income and uninsured population. Trained interviewers conducted computer-assisted personal interviews to collect baseline data on demographic characteristics and potential health risk factors, including medical history, dietary habits, physical activity, smoking, alcohol consumption, and anthropometrics. Blood samples were collected for nearly half of CHC-enrolled participants and stored at our Vanderbilt freezer facilities.

Abbreviation: SCCS, Southern Community Cohort Study.

A total of 522 incident lung cancer cases (defined by the International Classification of Diseases, 10th Revision: ICD-10, C340–C349) with stored blood specimens were ascertained *via* linkage with state cancer registries operating in the 12-state study area and/or from the National Death Index mortality records, by November 2016, and included in the current study. Controls (N=979) with stored blood samples were randomly selected from cancer-free SCCS participants and individually matched to cases at a 2:1 or 1:1 ratio on age (± 2 years), race (African American or European American), and sex, as well as the date (± 6 months) and site (CHC) of study enrollment, and thus, the date of blood collection. Completely de-identified data were available for the current analysis.

Laboratory Assays

Blood samples were provided by SCCS participants during enrollment at CHCs. After a blood draw, samples were immediately refrigerated and shipped cold overnight to the Vanderbilt Epidemiology Center's Molecular Epidemiology Laboratory. Serum samples were isolated and stored at -80°C. An aliquot of serum samples was sent to the testing lab in dry ice for biomarker analyses. For samples included in the current study, comprehensive metabolic panels, including total bilirubin and albumin levels, were measured using the Beckman Coulter clinical chemistry analyzer (DXC 600) at the core research laboratory of the University of Washington's Kidney Research Institute. Quality control samples (3%) were included in the assays. The intra- and inter-assay coefficients of variation were 0.9% and 1.9%, respectively, for albumin; and 12.7% and 13.4%, respectively, for total bilirubin. Serum samples from each case-control set were analyzed in the same batch and adjacently to eliminate between-assay variability. The laboratory staff was blinded to the case-control status of serum samples and the identity of quality control samples included in the study.

Statistical Analysis

This study included 522 lung cancer incident cases and 979 matched controls, which consisted of African Americans (334 lung cancer cases and 629 matched controls) and European Americans (188 lung cancer cases and 350 matched controls). Baseline characteristics between lung cancer cases and matched controls were compared by the chi-squared test (for categorical variables) or the student *t*-test (for continuous variables). Serum levels of total bilirubin and albumin were categorized into race-sex specific tertiles among controls. We also categorized serum levels of total bilirubin and albumin into tertiles among all controls, and the results were similar to those using race-sex specific tertiles. Thus, we present results using race-sex specific tertiles among controls. In addition, we further calculated the total bilirubin/albumin ratio grouped into race-sex specific tertiles. For this procedure, total bilirubin (mg/dL) and albumin (g/dL) were converted to μmol/L. We conducted conditional logistic regression analyses to investigate the associations between tertiles of total bilirubin and albumin and lung cancer risk after adjusting for potential confounders, including baseline age, smoking status (current vs. former vs. never), pack-years (<30 vs. ≥30), alcohol consumption (heavy vs. moderate vs.

nondrinker), education (less than 11 years vs. completed high school vs. vocational/technical school vs. university degree or higher), household income (<$15,000 vs. $15,000-$24,999 vs. ≥$25,000), self-report of ever having been diagnosed with chronic obstructive pulmonary disease (COPD; no vs. yes), and body mass index (BMI; ≥ 30 kg/m^2 vs. 29.9-25.0 kg/m^2 vs. <25 kg/m^2). Stratified analyses were performed to investigate modifications by race (African American vs. European American) and the time between sample collection and lung cancer diagnosis (≤2 years vs. >2 years). We further conducted stratified analyses by smoking status and lung cancer histological subtype (adenocarcinoma, squamous cell, and small cell lung cancer). Interactions were evaluated by the likelihood ratio test. All statistical analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC, USA), with *p*-values of less than 0.05 being considered statistically significant.

RESULTS

Baseline characteristics of the study population are presented in **Table 1**. Lung cancer cases had a higher proportion of current smoking (74.3%) and higher pack-years (30.5 pack-years) than matched controls (46.5% and 19.2 pack-years, respectively). Compared with controls, lung cancer cases were more likely to be less educated, earn less income, be alcohol drinkers, have lower BMI, and have a history of COPD (**Table 1**).

Odds ratios (ORs) for lung cancer risk associated with serum levels of total bilirubin and albumin are presented in **Table 2** and **Supplementary Figure 1**. Overall, serum levels of total bilirubin ($OR_{T3\ vs.\ T1}$ = 0.96, 95% CI: 0.66-1.39) were not significantly associated with the risk of lung cancer. When the analyses were conducted separately for African Americans and European Americans, we observed a marginally significant interaction between race and serum total bilirubin levels (*p* interaction=0.05). Higher serum levels of albumin were significantly associated with decreased risk of lung cancer ($OR_{T3\ vs.\ T1}$ = 0.70, 95% CI: 0.50-0.98). The inverse association remained significant among African Americans ($OR_{T3\ vs.\ T1}$ = 0.62, 95% CI: 0.41-0.96) but not among European Americans ($OR_{T3\ vs.\ T1}$ = 0.73, 95% CI: 0.41-1.31) (**Table 2**). We also evaluated the associations of lung cancer risk with the total bilirubin/albumin ratio, a surrogate measurement of free bilirubin. As shown in **Supplementary Table 1**, no significant

TABLE 1 | Baseline characteristics of study population, SCCS.

	Cases (N = 522)	Controls (N = 979)	*p*-value[†]
Age (mean ± SD)	56.4 ± 9.0	56.3 ± 9.0	0.77
Race [N (%)]			
African Americans	334 (64.0)	629 (64.2)	0.92
European Americans	188 (36.0)	350 (35.8)	
Gender [N (%)]			
Men	289 (55.4)	540 (55.2)	0.94
Women	233 (44.6)	439 (44.8)	
Smoking Status [N (%)]			
Current	388 (74.3)	455 (46.5)	< 0.01
Former	104 (19.9)	240 (24.5)	
Never	30 (5.7)	284 (29.0)	
Pack-years[a] [median (IQR)]	30.5 (17.6-48.0)	19.2 (10.0-36.0)	< 0.01
Alcohol consumption [N (%)]			
Heavy	125 (24.0)	162 (16.5)	< 0.01
Moderate[b]	185 (35.4)	341 (34.8)	
Nondrinker	212 (40.6)	476 (48.6)	
Education [N (%)]			
Less than 11 years	247 (47.3)	388 (39.6)	< 0.01
Completed high school	173 (33.1)	345 (35.2)	
Vocational/technical school	88 (16.9)	171 (17.5)	
University degree or higher	14 (2.7)	75 (7.7)	
Household income [N (%)]			
< $15,000	377 (72.2)	624 (63.7)	< 0.01
$15,000 - $24,999	96 (18.4)	220 (22.5)	
≥ $25,000	49 (9.4)	135 (13.8)	
Self-report of COPD[c] [N (%)]			
No	432 (82.8)	885 (90.4)	< 0.01
Yes	90 (17.2)	94 (9.6)	
BMI [kg/m^2, median (IQR)]	25.8 (22.3-30.0)	28.2 (24.5-33.1)	< 0.01
Total Bilirubin [mg/dL, mean (SD)]	0.41 (0.22)	0.45 (0.32)	< 0.01
Albumin [g/dL, mean (SD)]	4.28 (0.37)	4.31 (0.36)	0.08

[†]*Estimated by the student t-test procedure for continuous variables and chi-square test for categorical variables.*
[a]*Included former/current smoker.*
[b]*Defined as > 0 but ≤ 2 drink/day for men or ≤1 drink/day for women.*
[c]*Ever diagnosed with emphysema or chronic bronchitis.*

TABLE 2 | Association of serum levels of total bilirubin and albumin with lung cancer risk.

	Median[a]	Cases	Controls	OR (95%CI)[†]	OR (95%CI)[‡]
Total Bilirubin (mg/dL)[b]					
Total population		**N=522**	**N=979**		
T1	0.20	162	287	1.00 (Ref.)	1.00 (Ref.)
T2	0.38	165	294	1.16 (0.82-1.63)	1.09 (0.76-1.56)
T3	0.60	195	398	1.07 (0.75-1.52)	0.96 (0.66-1.39)
p trend				0.85	0.71
African Americans		**N=334**	**N=629**		
T1	0.20	91	179	1.00 (Ref.)	1.00 (Ref.)
T2	0.37	128	214	1.48 (0.98-2.24)	1.49 (0.95-2.33)
T3	0.60	115	236	1.35 (0.86-2.12)	1.22 (0.75-1.98)
p trend				0.28	0.61
European Americans		**N=188**	**N=350**		
T1	0.20	71	108	1.00 (Ref.)	1.00 (Ref.)
T2	0.40	37	80	0.63 (0.33-1.21)	0.52 (0.26-1.04)
T3	0.60	80	162	0.66 (0.36-1.21)	0.57 (0.30-1.09)
p trend				0.25	0.15
***p* interaction[c]**				0.06	0.05
Albumin (g/dL)[b]					
Total population		**N=522**	**N=979**		
T1	3.90	173	283	1.00 (Ref.)	1.00 (Ref.)
T2	4.30	178	339	0.86 (0.62-1.18)	0.88 (0.63-1.23)
T3	4.60	171	357	0.78 (0.56-1.07)	0.70 (0.50-0.98)
p trend				0.12	0.04
African Americans		**N=334**	**N=629**		
T1	3.90	115	185	1.00 (Ref.)	1.00 (Ref.)
T2	4.30	114	221	0.83 (0.56-1.23)	0.90 (0.60-1.37)
T3	4.60	105	223	0.72 (0.49-1.08)	0.62 (0.41-0.96)
p trend				0.11	0.03
European Americans		**N=188**	**N=350**		
T1	4.00	58	98	1.00 (Ref.)	1.00 (Ref.)
T2	4.30	64	118	0.88 (0.50-1.55)	0.87 (0.48-1.56)
T3	4.70	66	134	0.85 (0.49-1.45)	0.73 (0.41-1.31)
p trend				0.55	0.28
***p* interaction[c]**				0.89	0.79

Analysis using conditional logistic regression models.

[a]*Median level of each tertile.*

[b]*Based on the race- and sex-specific tertiles among controls.*

[c]*p interaction between Total Bilirubin/Albumin and race with lung cancer risk.*

[†]*Adjustment for age, smoking status, and pack-years.*

[‡]*Adjustment for age, smoking status, pack-years, alcohol consumption, education, household income, history of COPD, and BMI.*

associations were found between the total bilirubin/albumin ratio and lung cancer risk, regardless of race.

We conducted analyses stratified by the time between sample collection and lung cancer diagnosis. Higher serum total bilirubin levels were significantly associated with decreased risk of lung cancer among participants diagnosed ≤2 years after blood collection ($OR_{T3\ vs.T1}$ = 0.36, 95% CI: 0.15-0.87; *p* trend=0.03) (**Table 3**). Higher albumin levels were also associated with decreased risk of lung cancer, although not significantly, among participants diagnosed ≤2 years after blood collection ($OR_{T3\ vs.T1}$ = 0.57, 95% CI: 0.27-1.18). No significant associations were observed in individuals who were diagnosed >2 years after blood collection for either total bilirubin or albumin (**Table 3**).

We further conducted analyses by smoking status. Serum total bilirubin levels were associated with decreased risk of lung cancer among former/never smokers ($OR_{T3\ vs.T1}$ = 0.54, 95% CI: 0.32-0.93) but not among current smokers (**Table 4**). The associations of serum albumin levels were similar between current smokers and former/never smokers. We found no significant differences across histological types (*i.e.*, adenocarcinoma, squamous cell lung cancer, and small cell lung cancer; data not shown).

DISCUSSION

This study investigated the associations of serum levels of total bilirubin and albumin with lung cancer risk among low-income Americans. Overall, total bilirubin was not associated with lung cancer risk, but higher serum levels of total bilirubin were associated with decreased risk of lung cancer within two years of blood collection and among former/never smokers. In addition, we found that higher levels of albumin were significantly inversely associated with lung cancer risk overall and among African Americans.

Previous studies among European and Asian decedents indicated an inverse association between total bilirubin and lung cancer risk. Horsfall and colleagues reported that a 0.1 mg/

TABLE 3 | Association of serum levels of total bilirubin and albumin with lung cancer risk by time between blood collection and lung cancer diagnosis.

	Median[a]	Cases	Controls	OR (95%CI)[†]	OR (95%CI)[‡]
≤ 2 years		**N=128**	**N=234**		
Total Bilirubin (mg/dL)[b]					
T1	0.20	70	106	1.00 (Ref.)	1.00 (Ref.)
T2	0.37	34	64	0.87 (0.44-1.72)	0.80 (0.38-1.67)
T3	0.60	24	64	0.50 (0.23-1.06)	0.36 (0.15-0.87)
p trend				0.07	0.03
Albumin (g/dL)[b]					
T1	4.00	30	52	1.00 (Ref.)	1.00 (Ref.)
T2	4.30	55	77	0.80 (0.38-1.67)	0.70 (0.32-1.52)
T3	4.60	43	105	0.60 (0.30-1.20)	0.57 (0.27-1.18)
p trend				0.14	0.14
> 2 years		**N=394**	**N=745**		
Total Bilirubin (mg/dL)[b]					
T1	0.20	92	181	1.00 (Ref.)	1.00 (Ref.)
T2	0.40	131	230	1.35 (0.90-2.04)	1.25 (0.81-1.94)
T3	0.60	171	334	1.36 (0.89-2.07)	1.17 (0.75-1.83)
p trend				0.22	0.64
Albumin (g/dL)[b]					
T1	3.90	143	231	1.00 (Ref.)	1.00 (Ref.)
T2	4.30	123	262	0.86 (0.60-1.22)	0.88 (0.60-1.28)
T3	4.60	128	252	0.84 (0.58-1.20)	0.74 (0.50-1.09)
p trend				0.35	0.12

Analysis using conditional logistic regression model.

[a]*Median level of each tertile.*

[b]*Based on the race- and sex-specific tertiles among controls.*

[†]*Adjustment for age, smoking status, and pack-years.*

[‡]*Adjustment for age, smoking status, pack-years, alcohol consumption, education, household income, history of COPD, and BMI.*

dL increase in serum bilirubin levels was associated with a lower risk of lung cancer in both men [incidence rate ratio (IRR)=0.92, 95% CI: 0.89-0.95] and women (IRR=0.89, 95% CI: 0.86-0.93) (9). In the EPIC-Heidelberg cohort study, higher levels of bilirubin showed a marginally significant association with a lower risk of lung cancer [hazard ratio (HR)=0.72, 95% CI: 0.51-1.00]. Wen and colleagues reported that bilirubin was inversely associated with lung cancer risk and mortality in Taiwanese male smokers (14). The Severance Cohort Study in Korea reported an increased risk of lung cancer with bilirubin levels from 0.2 to 0.7 mg/dL in men (HR=2.8, 95% CI: 1.8–4.2), compared with bilirubin levels ≥1.0 mg/dL (15). The authors also observed that one standard deviation (SD) increase in bilirubin was associated with a 30% decreased risk of lung cancer (HR=0.7, 95% CI: 0.5–0.9) among current smokers. A Japanese study using electronic medical records during a median 4.7-year follow-up demonstrated that serum bilirubin levels over 1.2 mg/dL were associated with a reduced risk of lung cancer among men (HR=0.47, 95% CI: 0.27-0.83) compared with ≤1.2 mg/dL (10). A large-scale study from Sweden also reported an inverse association between serum total

TABLE 4 | Association of serum levels of total bilirubin and albumin with lung cancer risk by smoking status.

	Current Smokers				Former and Never Smokers			
	Median[a]	Cases (N=388)	Controls (N=455)	OR (95%CI)[†]	Median[a]	Cases (N=134)	Controls (N=524)	OR (95%CI)[†]
Total Bilirubin (mg/dL)[b]								
T1	0.20	117	143	1.00 (Ref.)	0.20	45	144	1.00 (Ref.)
T2	0.39	118	137	1.06 (0.73-1.54)	0.37	47	157	0.90 (0.54-1.51)
T3	0.60	153	175	1.03 (0.73-1.45)	0.60	42	223	0.54 (0.32-0.93)
p trend				0.87				0.03
								p interaction: 0.19
Albumin (g/dL)[b]								
T1	3.90	128	137	1.00 (Ref.)	4.00	45	146	1.00 (Ref.)
T2	4.30	128	154	0.94 (0.66-1.34)	4.30	50	185	1.03 (0.61-1.72)
T3	4.60	132	164	0.83 (0.58-1.17)	4.60	39	193	0.70 (0.40-1.22)
p trend				0.28				0.22
								p interaction: 0.60

[a]*Median level of each tertile.*

[b]*Based on the race- and sex-specific tertiles among controls.*

[†]*Adjustment for age, sex, race, smoking status, pack-years, alcohol consumption, education, household income, history of COPD, and BMI.*

bilirubin and lung cancer risk ($HR_{Q4\ vs.Q1}$ = 0.50, 95% CI: 0.44-0.59) (17): this study showed that higher serum total bilirubin levels decreased the risk of lung cancer among non-smokers (HR=0.45, 95% CI: 0.24-0.86) but not among smokers (HR=0.84, 95% CI: 0.44-1.60). In addition, a meta-analysis of five cohort studies reported that high levels of bilirubin were associated with decreased risk of lung cancer (relative risk=0.69, 95% CI: 0.55-0.86) (17). A study from the UK Biobank reported a potential causal association between serum bilirubin and the incidence of lung cancer by Mendelian Randomization (21); this study suggested that genetically predicted serum bilirubin might protect individuals from exposure hazards to high levels of oxidants associated with cigarette smoking. Unlike previous studies, our study did not find a significant association between total bilirubin levels and lung cancer risk among total study participants. It is worth noting that our study was conducted among low-income African Americans and European Americans. Nevertheless, we found an inverse association of total bilirubin levels with lung cancer risk among cases diagnosed within two years after blood collection. This suggests that total bilirubin may be related to lung cancer progression and could potentially be a lung cancer biomarker. In addition, the inverse association of serum total bilirubin levels was only observed among former/never smokers in our study, indicating that total bilirubin may be a promising biomarker for non-current smokers.

Bilirubin exhibits potent antioxidant effects and has been shown to reduce age-related inflammation and metabolic deterioration in preclinical rodent models (22). The link between inflammation and cancer has long been recognized. Thus, endogenous compounds reducing inflammation have been generally hypothesized to have cancer-preventive properties. Additional mechanisms also indicate a protective role of bilirubin in cancer risk: bilirubin may act as an immunomodulatory agent and has been shown to suppress CD4 T cell responses (23). CD4-positive T cells, along with their related cytokines, are associated with lung cancer risk (24).

We found that higher albumin levels were associated with decreased risk of lung cancer overall and among African Americans. A previous study reported that higher levels of albumin had a marginally significant association with a lower risk of lung cancer (HR=0.72, 95% CI: 0.51-1.00) (13). Recently, the Kailuan cohort in China showed that pre-diagnostic albumin was inversely associated with lung cancer risk ($HR_{Q4\ vs.\ Q1}$ = 0.70, 95% CI: 0.52-0.95), but the inverse association became insignificant after excluding cases diagnosed within two years of enrollment (25). Another study also reported that one SD increment in albumin was inversely associated with lung cancer risk (HR=0.82, 95% CI: 0.72-0.94) among Koreans (26).

Albumin plays a role as a scavenger and antioxidant (8, 27–29), which may reflect the inflammatory and nutritional status (30). Albumin can be broken down in the cell, but it still provides amino acids for cell proliferation and matrix deposition (28). Low serum albumin levels are generally regarded as an indicator of severe inflammation or malnutrition (28, 31). The inverse correlations of serum albumin levels with C-reactive protein and tumor necrosis factor-α also support the association between serum albumin and inflammation (32–34). Given a strong link of chronic infection and inflammatory microenvironment to lung cancer (35), the inverse associations of serum albumin that we observed in the current study could be biologically plausible. However, future investigations are needed to further explore the biological role of albumin in the development of lung cancer and whether the molecular mechanisms underneath the albumin-inflammation association differ by race/ethnicity or socio-economic status.

The current study has several strengths. First, we conducted a population-based nested case-control study, including African Americans and low-income populations in the US. This could provide a unique insight into the associations of serum total bilirubin and albumin with lung cancer among underserved populations who are at a greater risk for lung cancer. Second, our study used blood samples collected before lung cancer diagnosis and treatment, thus minimizing potential reverse causality. Finally, comprehensive covariate information from the SCCS allowed us to adjust for major confounders. The limitations of our study should also be noted. Larger sample size was necessary for additional stratified analyses by smoking status and lung cancer histological subtypes. Despite the comprehensive adjustments for covariates, we could not completely rule out the influence of residual confounding variables, unmeasured potential confounders, or one-time measurements.

CONCLUSION

Our findings show that total bilirubin serum levels were inversely associated with short-term lung cancer incidence, particularly within two years of blood collection. These findings raise the possibility that lung cancer itself and/or its immediate clinical precursors may influence (lower) total bilirubin levels, leading to the appearance of a protective effect of the biomarkers upon lung cancer. However, the absence of associations beyond two years may imply that serum levels of total bilirubin are unrelated to lung cancer risk. In addition, total bilirubin was associated with decreased lung cancer risk among former/never smokers, indicating that total bilirubin may be a promising biomarker for non-current smokers. We also found that serum albumin was inversely associated with lung cancer risk overall and among African Americans. Further studies with a larger sample size are warranted to confirm our findings, especially on effect modification by race/ethnicity and smoking status.

DATA AVAILABILITY STATEMENT

Data used in the present study can be requested through the Southern Community Cohort Study online request system (https://ors.southerncommunitystudy.org/).

ETHICS STATEMENT

The SCCS protocol was reviewed and approved by the institutional review boards at Vanderbilt University Medical Center and Meharry Medical College. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

QC conceived of the study. WZ and WB provided study resources. QC and JW conducted laboratory analyses. HSY and CS performed statistical analyses. All authors contributed to interpretation of results. HSY, CS, XS, and QC drafted the manuscript. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the National Institutes of Health (T32CA160056, R01CA092447, and U01CA202979). Data collection and sample preparation were performed by the Survey and Biospecimen Shared Resource which is supported in part by the Vanderbilt-Ingram Cancer Center (P30CA68485).

ACKNOWLEDGMENTS

Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72,205. The Arkansas Central Cancer Registry is fully funded by a grant from the National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The authors thank the study participants and research staff of the Southern Community Cohort Study for their contribution to this study. We also thank Regina Courtney for laboratory management and Dr. Mary Shannon Byers and Ms. Kathleen Harmeyer for editing and manuscript preparation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.895479/full#supplementary-material

Supplementary Figure 1 | Association of serum levels of total bilirubin and albumin with lung cancer risk.

Supplementary Table 1 | Association of serum levels of total bilirubin/albumin ratio with lung cancer risk.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
2. American Lung Association. Too Many Cases, Too Many Deaths. In: *Lung Cancer in African Americans*. (Washington D.C: American Lung Association), (2010).
3. Ryan BM. Lung Cancer Health Disparities. *Carcinogenesis* (2018) 39:741–51. doi: 10.1093/carcin/bgy047
4. Pham-Huy LA, He H, Pham-Huy C. Free Radicals, Antioxidants in Disease and Health. *Int J BioMed Sci* (2008) 4:89–96.
5. Marnett LJ. Oxyradicals and DNA Damage. *Carcinogenesis* (2000) 21:361–70. doi: 10.1093/carcin/21.3.361
6. Peng YF, Goyal H, Xu GD. Serum Bilirubin has an Important Role in Multiple Clinical Applications. *J Lab Precis Med* (2017) 2:82–2. doi: 10.21037/jlpm.2017.09.08
7. Gupta D, Lis CG. Pretreatment Serum Albumin as a Predictor of Cancer Survival: A Systematic Review of the Epidemiological Literature. *Nutr J* (2010) 9:69. doi: 10.1186/1475-2891-9-69
8. Roche M, Rondeau P, Singh NR, Tarnus E, Bourdon E. The Antioxidant Properties of Serum Albumin. *FEBS Lett* (2008) 582:1783–7. doi: 10.1016/j.febslet.2008.04.057
9. Horsfall LJ, Rait G, Walters K, Swallow DM, Pereira SP, Nazareth I, et al. Serum Bilirubin and Risk of Respiratory Disease and Death. *JAMA* (2011) 305:691–7. doi: 10.1001/jama.2011.124
10. Inoguchi T, Nohara Y, Nojiri C, Nakashima N. Association of Serum Bilirubin Levels With Risk of Cancer Development and Total Death. *Sci Rep* (2021) 11:13224. doi: 10.1038/s41598-021-92442-2
11. Dudeja V, Ferrantella A, Fong Y. The Liver. In: *Sabiston Textbook of Surgery, 21st ed.* Philadelphia, PA: Elsevier (2021).
12. Adin CA. Bilirubin as a Therapeutic Molecule: Challenges and Opportunities. *Antioxidants* (2021) 10:1536. doi: 10.3390/antiox10101536
13. Kühn T, Sookthai D, Graf ME, Schübel R, Freisling H, Johnson T, et al. Albumin, Bilirubin, Uric Acid and Cancer Risk: Results From a Prospective Population-Based Study. *Br J Cancer* (2017) 117:1572–9. doi: 10.1038/bjc.2017.313
14. Wen CP, Zhang F, Liang D, Wen C, Gu J, Skinner H, et al. The Ability of Bilirubin in Identifying Smokers With Higher Risk of Lung Cancer: A Large Cohort Study in Conjunction With Global Metabolomic Profiling. *Clin Cancer Res* (2015) 21:193–200. doi: 10.1158/1078-0432.CCR-14-0748
15. Lim JE, Kimm H, Jee SH. Combined Effects of Smoking and Bilirubin Levels on the Risk of Lung Cancer in Korea: The Severance Cohort Study. *PloS One* (2014) 9:e103972. doi: 10.1371/journal.pone.0103972

16. Sprague BL, Trentham-Dietz A, Klein BEK, Klein R, Cruickshanks KJ, Lee KE, et al. Physical Activity, White Blood Cell Count, and Lung Cancer Risk in a Prospective Cohort Study. *Cancer Epidemiol Biomarkers Prev* (2008) 17:2714–22. doi: 10.1158/1055-9965.EPI-08-0042
17. Monroy-Iglesias MJ, Moss C, Beckmann K, Hammar N, Walldius G, Bosco C, et al. Serum Total Bilirubin and Risk of Cancer: A Swedish Cohort Study and Meta-Analysis. *Cancers (Basel)* (2021) 13:5540. doi: 10.3390/cancers13215540
18. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer Statistics for African American/Black People 2022. *CA Cancer J Clin* (2022) 72 (3):202–229. doi: 10.3322/caac.21718
19. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern Community Cohort Study: Establishing a Cohort to Investigate Health Disparities. *J Natl Med Assoc* (2005) 97:972–9.
20. Signorello LB, Hargreaves MK, Blot WJ. The Southern Community Cohort Study: Investigating Health Disparities. *J Health Care Poor Underserved* (2010) 21:26–37. doi: 10.1353/hpu.0.0245
21. Horsfall LJ, Burgess S, Hall I, Nazareth I. Genetically Raised Serum BilirubinLevels and Lung Cancer: A Cohort Study and Mendelian RandomisationUsing UK Biobank.*Thorax* (2020) 75(11):955–64. doi: 10.1136/thoraxjnl-2020-214756.
22. Zelenka J, Dvořák A, Alán L, Zadinová M, Haluzík M, Vítek L. Hyperbilirubinemia Protects Against Aging-Associated Inflammation and Metabolic Deterioration. *Oxid Med Cell Longev* (2016) 2016:6190609. doi: 10.1155/2016/6190609
23. Liu Y, Stewart KN, Bishop E, Marek CJ, Kluth DC, Rees AJ, et al. Unique Expression of Suppressor of Cytokine Signaling 3 Is Essential for Classical Macrophage Activation in Rodents *In Vitro* and *In Vivo*. *J Immunol* (2008) 180:6270–8. doi: 10.4049/jimmunol.180.9.6270
24. Liao C, Yu ZB, Meng G, Wang L, Liu QY, Chen LT, et al. Association Between Th17-Related Cytokines and Risk of Non-Small Cell Lung Cancer Among Patients With or Without Chronic Obstructive Pulmonary Disease. *Cancer* (2015) 121 Suppl 17:3122–9. doi: 10.1002/cncr.29369
25. Yang Z, Zheng Y, Wu Z, Wen Y, Wang G, Chen S, et al. Association Between Pre-Diagnostic Serum Albumin and Cancer Risk: Results From a Prospective Population-Based Study. *Cancer Med* (2021) 10:4054–65. doi: 10.1002/cam4.3937
26. Kim YR, Choi CK, Lee YH, Choi SW, Kim HY, Shin MH, et al. Association Between Albumin, Total Bilirubin, and Uric Acid Serum Levels and the Risk of Cancer: A Prospective Study in a Korean Population. *Yonsei Med J* (2021) 62:792–8. doi: 10.3349/ymj.2021.62.9.792
27. Taverna M, Marie AL, Mira JP, Guidet B. Specific Antioxidant Properties of Human Serum Albumin. *Ann Intensive Care* (2013) 3:4. doi: 10.1186/2110-5820-3-4
28. Soeters PB, Wolfe RR, Shenkin A. Hypoalbuminemia: Pathogenesis and Clinical Significance. *JPEN J Parenter Enteral Nutr* (2019) 43:181–93. doi: 10.1002/jpen.1451
29. Merlot AM, Kalinowski DS, Richardson DR. Unraveling the Mysteries of Serum Albumin-More Than Just a Serum Protein. *Front Physiol* (2014) 5:299/abstract. doi: 10.3389/fphys.2014.00299
30. Eckart A, Struja T, Kutz A, Baumgartner A, Baumgartner T, Zurfluh S, et al. Relationship of Nutritional Status, Inflammation, and Serum Albumin Levels During Acute Illness: A Prospective Study. *Am J Med* (2020) 133:713–722.e7. doi: 10.1016/j.amjmed.2019.10.031
31. Kondrup J, Allison SP, Elia M, Vellas B, Plauth M. Educational and Clinical Practice Committee, European Society of Parenteral and Enteral Nutrition (ESPEN). ESPEN Guidelines for Nutrition Screening 2002. *Clin Nutr* (2003) 22:415–21. doi: 10.1016/S0261-5614(03)00098-0
32. Ishida S, Hashimoto I, Seike T, Abe Y, Nakaya Y, Nakanishi H. Serum Albumin Levels Correlate With Inflammation Rather Than Nutrition Supply in Burns Patients: A Retrospective Study. *J Med Invest* (2014) 61:361–8. doi: 10.2152/jmi.61.361
33. Ohwada H, Nakayama T, Kanaya Y, Tanaka Y. Serum Albumin Levels and Their Correlates Among Individuals With Motor Disorders at Five Institutions in Japan. *Nutr Res Pract* (2017) 11:57–63. doi: 10.4162/nrp.2017.11.1.57
34. Oe Y, Mochizuki K, Miyauchi R, Misaki Y, Kasezawa N, Tohyama K, et al. Plasma TNF-α Is Associated With Inflammation and Nutrition Status in Community-Dwelling Japanese Elderly. *J Nutr Sci Vitaminol (Tokyo)* (2015) 61:263–9. doi: 10.3177/jnsv.61.263
35. Tan Z, Xue H, Sun Y, Zhang C, Song Y, Qi Y. The Role of Tumor Inflammatory Microenvironment in Lung Cancer. *Front Pharmacol* (2021) 12:688625. doi: 10.3389/fphar.2021.688625

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fonc.2022.913929

Impact of Marital Status on Prognosis of Patients With Invasive Breast Cancer: A Population-Based Study Using SEER Database

Dechuang Jiao[†], *Youzhao Ma*[†], *Jiujun Zhu*, *Hao Dai*, *Yue Yang*, *Yajie Zhao*, *Xuhui Guo and Zhenzhen Liu*[*]

Department of Breast Disease, Henan Breast Cancer Center, Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital, Zhengzhou, China

OPEN ACCESS

Edited by:
Paolo Boffetta,
Icahn School of Medicine at Mount Sinai, United States

Reviewed by:
Qin Wenxing,
Shanghai Changzheng Hospital, China
Changjun Wang,
Peking Union Medical College Hospital (CAMS), China

***Correspondence:**
Zhenzhen Liu
zlyyliuzhenzhen0800@zzu.edu.cn

[†]*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *07 April 2022*
Accepted: *20 June 2022*
Published: *22 July 2022*

Citation:
Jiao D, Ma Y, Zhu J, Dai H, Yang Y, Zhao Y, Guo X and Liu Z (2022) Impact of Marital Status on Prognosis of Patients With Invasive Breast Cancer: A Population-Based Study Using SEER Database. Front. Oncol. 12:913929. doi: 10.3389/fonc.2022.913929

Objective: This study aimed to investigate the prognostic roles of marital status in patients with invasive breast cancer. Method: We extracted the data of patients with invasive breast cancer who were diagnosed during 2010–2015 and had complete staging and molecular typing from the Surveillance, Epidemiology, and End Results (SEER)-18 database. Kaplan–Meier curve method and Cox regression analysis were performed to investigate the differences in breast cancer–specific survival (BCSS) and overall survival (OS) in the total population and various subgroups with different marital statuses.

Results: Among the 324,062 patients with breast cancer in this study, 55.0%, 40.0%, and 5.0% were married, unmarried, and unknown, respectively; 51.8%, 32.2%, 10.5%, and 5.5% were patients with Stages I, II, III, and IV breast cancer, respectively. The 5-year BCSS and OS of married patients were 92.6% and 88.1%, respectively, higher than those of unmarried patients (88.3% and 78.1%, $P < 0.001$). After adjustment for sex, age, T and N stages, histological grade, insurance status, race, year of diagnosis, and molecular subtypes, married status was an independent predictor of better BCSS [hazard ratio (HR) = 0.775, 95% confidence interval (CI) = 0.753–0.797, $P < 0.001$) and OS (HR = 0.667, 95% CI = 0.653–0.681, $P < 0.001$). After multivariate analysis of various subgroups of sex, age, stage, histological grade, insurance status, race, and molecular subtype, married status was an independent predictor of better BCSS in all subgroups except for Grade IV, age < 35 years, and uninsured subgroups. Marital status was an independent predictor of better OS in all subgroups except the subgroup with age <35 years.

Conclusions: In conclusion, marital status was an independent prognostic factor for breast cancer. The unmarried patients with breast cancer had a worse prognosis, except for the subgroup with age <35 years. Hence, unmarried patients with breast cancer and age ≥35 years may need additional psychosocial and emotional support to achieve more prolonged survival, besides active treatment of primary disease.

Keywords: breast cancer, marital status, prognosis, surveillance, epidemiology and end results (SEER)

INTRODUCTION

Psychosocial factors are closely related to the occurrence and prognosis of malignant tumors, while marital status is one of the most critical psychosocial factors affecting the occurrence and development of malignant tumors (1, 2). Previous studies revealed that the married population had a healthier lifestyle, including a healthy diet, physical exercise, and regular physical examination, which might be intermediate factors in cancer prevention (3). Marital status is closely related to the prognosis of multiple malignant tumors (4–8). Married patients may receive more emotional and financial support, get more standardized and complete medical treatment, and obtain a better prognosis (9–11). In 2020, only 50% of American residents were married, which was a decrease of 9% in the last 25 years. Moreover, this downward trend has always existed. Hence, the relationship between marital status and cancer prevention and treatment is worthy of further research.

As a systematic disease, breast cancer has been considered one of the most affected cancers by marital status. The pain caused by widowhood or divorce and a series of following unhealthy lifestyles are associated with the onset of breast cancer (3, 12). Meanwhile, the lack of experience of pregnancy and lactation in unmarried women may also be related to breast cancer (13–16). Previous studies indicated that unmarried patients with breast cancer were usually diagnosed at an advanced stage and had more depressive symptoms. Additionally, compared with unmarried patients, married patients received more reasonable and standardized treatment (4).

Chen et al. (8) reported that marital status affected the prognosis of patients with breast cancer by affecting the stage at the time of diagnosis. However, the prognosis of breast cancer is affected by various factors. Indeed, other clinical and social indicators, including age, race, insurance status, and sex, are closely related to the marital status and the prognosis of breast cancer. Whether and how these factors affect the relationship between marital status and prognosis remain elusive. Additionally, the molecular typing and patients with the first diagnosis of Stage IV breast cancer were not included in previous studies on the relationship between marital status and prognosis. This study investigated the relationship between marital status and prognosis in patients with different molecular typing and stages.

MATERIALS AND METHODS

We extracted the data of marital status and other clinicopathological factors of patients aged 18 years or older and with breast cancer diagnosed from January 1, 2010, to December 31, 2015, from the Surveillance, Epidemiology, and End Results (SEER)-18 database released in April 2021. The inclusion criteria were as follows: (a) the known tumor stage (AJCC 7th edition), (b) the known molecular subtype, (c) invasive cancers, and (d) Stage I–III breast cancer subjected to surgical treatment or Stage IV breast cancer. Patients who did not meet the aforementioned inclusion criteria were excluded from our study. We excluded 15,913 cases with the unknown tumor stage, 25,887 cases with the unknown molecular typing, 95 cases of carcinoma *in situ* and 14,172 cases with Stages I–III without surgical treatment (**Figure 1**). Finally, a total of 324,062 cases were included in this study.

The marital status was divided into married and unmarried (e.g., single, divorced, separated, widowed, and domestic partner). The tumor stages were divided into four groups: Stage I, Stage II, Stage III, and Stage IV. The molecular typing was divided into four groups: HR+/HER2-, HR+/HER2+, HR-/HER2-, and HR-/HER2-. The race was divided into three groups: the white, the black, and other races (including American Indian or Alaska Native and Asian or Pacific Islander). The age was divided into three groups: <35 years old group, ≥35 years and <65 years old group, and ≧65 years old group. The histological grading was divided into four groups: well differentiated (Grade I), moderately differentiated (Grade II), poorly differentiated (Grade III), and undifferentiated and anaplastic (Grade IV). The insurance status was divided into three groups: uninsured, insured, and Medicaid.

The differences in clinicopathological indexes among groups were analyzed by the Pearson χ^2 test.BCSS time was defined as the time from the diagnosis to the death due to breast cancer, and OS time was defined as the time from the diagnosis to the death due to any cause. The Kaplan–Meier survival curve was used to estimate the survival rate, while the log-rank test was used to compare the survival differences among groups. A Cox proportional hazards model was constructed for univariate and multivariate analyses and the generation of HR value and 95% CI. All statistical analyses were performed using IBM SPSS Statistics, version 22.0 (IBM Corp.). All tests were two sided, with P <0.05 indicating a statistically significant difference.

RESULTS

Clinicopathologic Characteristics of Study Participants

A total of 324,062 cases were involved in this study, with an average age of 61.4 ± 13.3 years (median = 62.0;range:18-104 years) and a mean follow-up of 41.4 ± 21.6 months (median= 40.0 months; range: 0–83 months). The number of death events related to breast cancer was 22,274 (6.9%), and the total number of death events was 39,337 (12.1%). Further, 178,153 (55.0%), 129,549 (40.0%), and 16,360 (5.0%) of them were married, unmarried, and unknown, respectively. Patients with Stages I, II, III, and IV breast cancer accounted for 51.8%, 32.2%, 10.5%, and 5.5%, respectively. For other clinicopathological indexes, the number and proportion of patients in each subgroup are shown in **Table 1**.

Survival Analysis of All Participants According to Marital Status

Compared with unmarried patients, married patients exhibited significantly higher 5-year BCSS (92.6% vs 88.3%, P < 0.001) and 5-year OS (88.1% vs 78.1%, P < 0.001) (**Figure 2**).

Abbreviations: SEER, Surveillance, epidemiology, and end results; BCSS, Breast cancer–specific survival; OS, Overall survival; HR, Hazard ratio; HR+, Hormone receptor positive; HR-, Hormone receptor negative; HER2, Human epidermal growth factor receptor-2.

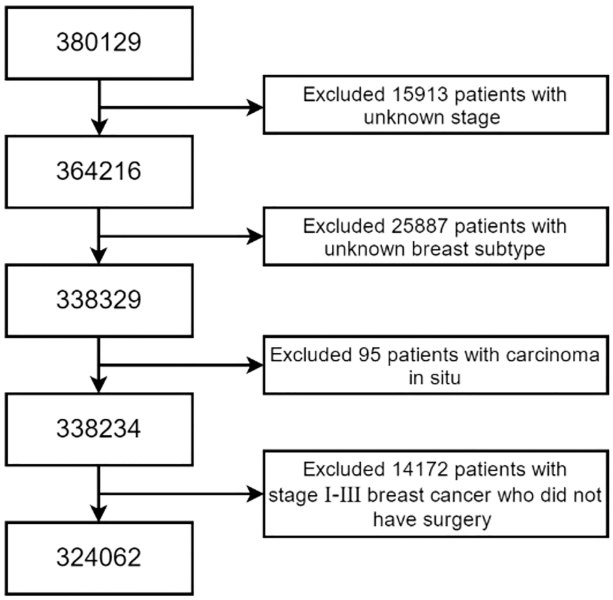


FIGURE 1 | Subject selection.

Based on the univariate analysis, marital status, sex, T stage, N stage, ER status, PR status, HER2 status, molecular typing, historical grading, age, insurance status, and year of diagnosis were associated with BCSS and OS. Compared with the unmarried subgroup, the HR of BCSS of the married subgroup was 0.583 (95% CI = 0.67–0.599, $P < 0.001$), and the HR of OS was 0.498 (95% CI = 0.488–0.509, $P < 0.001$). According to the multivariate analysis, after adjusted for sex, T stage, N stage, ER status, PR status, HER2 status, histological grading, age, insurance status, and year of diagnosis, the marital status was still an independent predictor of BCSS and OS. Compared with the unmarried subgroup, the HR of BCSS of the married subgroup was 0.775 (95% CI = 0.753–0.797, $P < 0.001$), and the HR of OS was 0.667 (95% CI = 0.653–0.681, $P < 0.001$) (**Table 2**).

Survival Analysis for Subgroups of Sex, Stage, Subtype, and Age

The study included 2317 male patients, accounting for only 0.8%. The 5-year BCSS (86.7%) of male participants was lower than that of female participants (90.8%). The five-year BCSS of married and unmarried male and female patients was (89.9% vs 79.1%, $P < 0.001$) and (92.6% vs 88.4%, $P < 0.001$), respectively (**Figure 3**).

The 5-year BCSS in subgroups with ages between 35 and 65 years (91.4%) was higher than that in the subgroups with age ≥65 years (90.2%) and <35 years (85.1%). In each age group, the 5-year BCSS of married patients was better than that of unmarried patients (**Figure 4**). Patients with low histological grading had a higher 5-year BCSS. For different histological grading, married patients had a higher 5-year BCSS (**Figure 4**).

In this study, breast cancer was divided into Stages I, II, III, and IV, and the 5-year BCSS was 98.0%, 92.5%, 77.7%, and 33.0%, respectively. Significant differences were observed among groups ($P < 0.001$). For different subgroups, the 5-year BCSS of married and unmarried patients was (98.3% vs 97.6%, $P < 0.001$), (93.8% vs 91.0%, $P < 0.001$), (80.9% vs 73.7%, $P < 0.001$), and (37.7% vs 28.8%, $P < 0.001$), respectively, as shown in **Figure 5**.

Among the four breast cancer subtypes, HR+/HER2− subtype showed the highest 5-year BCSS rate (93.0%).HR+/HER2+ subtype showed a higher 5-year BCSS rate (90.0%) than HR−/HER2+ (84.9%) and HR−/HER2− subtypes (79.3%). HR−/HER2− subtype showed the lowest 5-year BCSS rate. Among all four types of breast cancer, the 5-year BCSS of married patients and unmarried patients was (94.5% vs 90.9%, $P < 0.001$), (92.4% vs 86.8%, $P < 0.001$), (87.9% vs 80.5%, $P < 0.001$), and (81.8% vs 76.1%, $P < 0.001$), as shown in **Figure 6**.

Multivariable and Interaction Analyses of Subgroups Corresponding to Different Clinical-Pathological Factors

In each subgroup of sex, stage, molecular typing, histological grading, age, and insurance status, Cox regression analysis was conducted with BCSS and OS as the observation endpoints. The other five variables and marital status were involved in the Cox model. Being married was an independent predictor of better BCSS, except for Grade IV, age <35 years, and uninsured subgroups. Except for histological grading, all the other variables had interaction effects with marital status (**Figure 7**).

Moreover, being married was an independent predictor of better OS, except for age <35 years subgroup. All the variables had interaction effects with marital status (**Figure 8**).

TABLE 1 | Clinicalpathologic characteristics of subjects according to marital status.

Characteristics	Marital status			P	Total
	Unmarried No.(%)	Married No.(%)	Unknown No.(%)		No. (%)
All	129549 (40.0)%	178153 (55.0%)	16360 (5.0%)		324062 (100%)
Gender				<0.001	
Female	128860 (99.5%)	176525 (99.1%)	16224 (99.2%)		321609 (99.2%)
Male	689 (0.5%)	1628 (0.9%)	136 (0.8%)		2453 (0.8%)
T				<0.001	
T0/1	73540 (57.3%)	111427 (62.9%)	9955 (61.7%)		194922 (60.6%)
T2	39834 (31.0%)	50804 (28.7%)	4724 (29.3%)		95362 (29.7%)
T3	8388 (6.5%)	9729 (5.5%)	834 (5.2%)		18951 (5.9%)
T4	6536 (5.1%)	5066 (2.9%)	631 (3.9%)		12233 (3.8%)
N				<0.001	
N0	87260 (67.8%)	122699 (69.1%)	11327 (69.9%)		221286 (68.6%)
N1	29187 (22.7%)	40425 (22.8%)	3584 (22.1%)		73196 (22.7%)
N2	7412 (5.8%)	8875 (5.0%)	798 (4.9%)		17085 (5.3%)
N3	4810 (3.7%)	5516 (3.1%)	503 (3.1)		10829 (3.4%)
Stage				<0.001	
I	63358 (48.9%)	95886 (53.8%)	8608 (52.6%)		167852 (51.8%)
II	42439 (32.8%)	56761 (31.9%)	5144 (31.4%)		104344 (32.2%)
III	14691 (11.3%)	17713 (9.9%)	1593 (9.7%)		33997 (10.5%)
IV	9061 (7.0%)	7793 (4.4%)	1015 (6.2%)		17869 (5.5%)
ER				<0.001	
Negative	22100 (17.1%)	28900 (16.2%)	2714 (16.6%)		53714 (16.6%)
Positve	107437 (82.9%)	149242 (83.8%)	13643 (83.4%)		270322 (83.4%)
PR				<0.001	
Negative	36363 (28.1%)	46936 (26.4%)	4520 (27.7%)		87819 (27.2%)
Positve	92947 (71.9%)	130803 (73.6%)	11796 (72.3%)		235546 (72.8%)
HER2				<0.001	
Negative	110843 (85.6%)	151282 (84.9%)	14052 (85.9%)		276177 (85.2%)
Positve	18706 (14.4%)	26871 (15.1%)	2308 (14.1%)		47885 (14.8%)
Subtypes				<0.001	
HR+/HER2-	95701 (73.9%)	132093 (74.1%)	12192 (74.5%)		239986 (74.1%)
HR+/HER2+	13080 (10.1%)	18930 (10.6%)	1615 (9.9%)		33625 (10.4%)
HR-/HER2+	5626 (4.3%)	7941 (4.5%)	693 (4.2%)		14260 (4.4%)
HR-/HER2-	15142 (11.7%)	19189 (10.8%)	1860 (11.4%)		36191 (11.2%)
Grade				<0.001	
I	28176 (22.7%)	40987 (23.9%)	3709 (24.0%)		72872 (23.4%)
II	55191 (44.4%)	76459 (44.5%)	6995 (45.3%)		138645 (44.5%)
III	40462 (32.6%)	53773 (31.3%)	4692 (30.4%)		98927 (31.8%)
IV	393 (0.3%)	452 (0.3%)	39 (0.3%)		884 (0.3%)
Race				<0.001	
White	98503 (76.4%)	146752 (82.8%)	12621 (78.7%)		257876 (80.0%)
Black	21123 (16.4%)	12077 (6.8%)	2058 (12.8%)		35258 (10.9%)
Other races	9367 (7.3%)	18493 (10.4%)	1367 (80.5%)		29227 (9.1%)
Age group (years)				<0.001	
<35	2662 (2.1%)	3030 (1.7%)	271 (1.7%)		5963 (1.8%)
≥35,<65	61765 (47.7%)	112146 (62.9%)	8812 (53.9%)		182723 (56.4%)
≥65	65122 (50.3%)	62977 (35.3%)	7277 (44.5%)		135376 (41.8%)
Insurance				<0.001	
Uninsured	2583 (2.0%)	2091 (1.2%)	255 (1.8%)		4929 (1.5%)
Insured	103186 (80.4%)	162623 (92.1)	12443 (88.0%)		278252 (87.2%)
Medicaid	22576 (17.6%)	11931 (6.8%)	1442 (10.2%)		35949 (11.3%)
Year of diagnosis				<0.001	
2010	20115 (15.5%)	27354 (15.4%)	2231 (13.6%)		49700 (15.3%)
2011	20869 (16.1%)	28543 (16.0%)	3010 (18.4%)		52422 (16.2%)
2012	21442 (16.6%)	29412 (16.5%)	2955 (18.1%)		53809 (16.6%)
2013	22001 (17.0%)	30331 (17.0%)	2681 (16.4%)		55013 (17.0%)
2014	22222 (17.2%)	30728 (17.2%)	2847 (17.4%)		55797 (17.2%)
2015	22900 (17.7%)	31785 (17.8%)	2636 (16.1%)		57321 (17.7%)

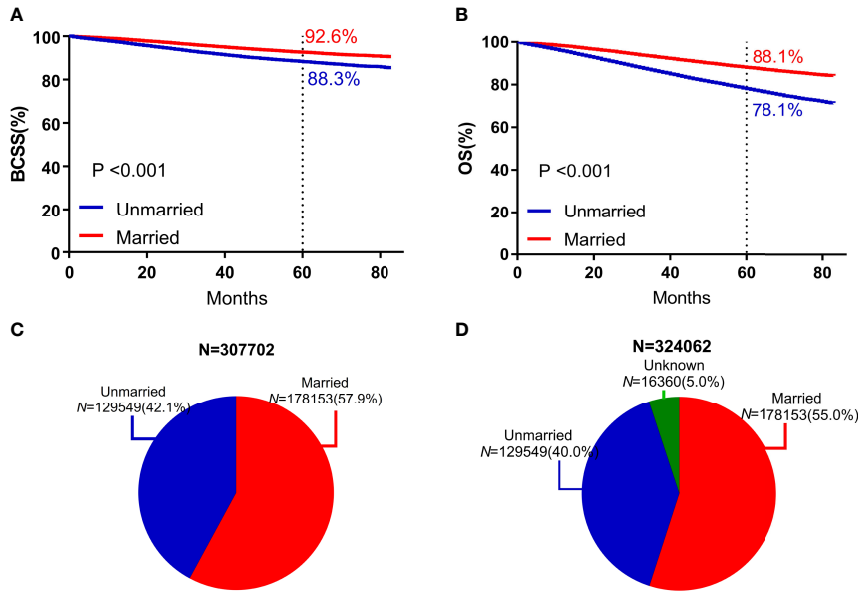


FIGURE 2 | Survival curves and subject proportions according to marital statuses. BCSS **(A)** and OS **(B)** according to marital statuses were depicted. Subject proportions according to marital statuses without the unknown marital status **(C)** and with the unknown marital status **(D)** were also depicted.

DISCUSSION

This study investigated the relationship between marital status and BCSS and OS in 324,062 patients with invasive breast cancer in the SEER database. First, by analyzing the relationship between marital status and survival in each subpopulation, we found that marital status was an independent prognostic factor among different stages of breast cancer. Therefore, the effect of marital status on prognosis could not be explained simply by the stage at the time of diagnosis. Then, our results indicated that the effect of marital status on survival might be different in different subgroups with breast cancer. Especially, the marital status of patients with breast cancer aged <35 had no significant effect on survival. Additionally, we also found that for different sex, molecular typing, and insurance status, marital status was an independent prognostic factor for breast cancer.

Aizer et al. (4) analyzed the survival rates of patients diagnosed as the top 10 tumors with the highest tumor-related mortality in the SEER database from 2004 to 2008. They found that marital status could affect the tumor stage, treatment, and tumor-related death at the time of diagnosis. Compared with unmarried patients with breast cancer, married patients had fewer advanced lesions at the time of diagnosis [odds ratio (OR) = 0.60, 95% CI = 0.58–0.63], higher proportion of receiving treatment (OR = 1.54, 95% CI = 1.42–1.66), and lower breast cancer–related mortality (HR = 0.78, 95% CI = 0.74–0.81).

Chen et al. (8) performed mediation analyses to investigate the intermediate factors of marital status affecting the survival of patients with cancer, demonstrating that marital status could affect the survival of patients by affecting the stage of breast cancer at the time of diagnosis. This study found that the proportion of married patients with the advanced stage at the time of diagnosis was lower than that of unmarried patients (4.4% vs 7.0%).

For patients with different stages of breast cancer, the relationship between marital status and survival was analyzed; the survival of married patients was better than that of unmarried patients in each stage. Consequently, we speculated that the marital status affected the survival of patients with breast cancer not just by influencing the stage at the time of diagnosis. Patients with different marital statuses might have different mental states and living habits. These factors might also affect the ability of patients with breast cancer to reintegrate into society, besides impacting the emotional recovery and even postoperative recovery (17–19).

Previous studies demonstrated that marital status was closely related to survival for aged patients with breast cancer (20). This study was the first to report that the relationship between marital status and survival was not consistent in each age subgroup. For patients with breast cancer aged <35 years, the prognosis of married and unmarried patients showed no differences. This was not consistent with previous studies (21). The analysis showed that the different correction factors added in the two studies contributed to the difference in the final results. Stage and molecular typing were involved in our multivariate analysis, both of which were closely related to the prognosis of breast cancer. Previous studies showed that compared with patients with breast cancer aged <35 years had an advanced stage, and the proportion of HER2-positive and triple-negative breast cancer was higher (22). Patients with early breast cancer aged <35 years had worse 5- and 10-year survival (23, 24), and marital status was irrelevant to this conclusion. Adekolujo at al. (25) analyzed the marital status

TABLE 2 | Univariable and multivariable analyses regarding BCSS and OS.

Characteristics	BCSS						OS					
	Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
	HR	95%CI	P	HR	95%CI	P	HR	95%CI	P	HR	95%CI	P
Marital status			<0.001			<0.001			<0.001			<0.001
Unmarried	ref			ref			ref			ref		
Married	0.583	0.567-0.599	<0.001	0.775	0.753-0.797	<0.001	0.498	0.488-0.509	<0.001	0.667	0.653-0.681	<0.001
Unknown	0.813	0.766-863	<0.001	0.919	0.865-0.977	0.007	0.78	0.746-0.815	<0.001	0.875	0.837-0.916	<0.001
Gender			<0.001			<0.001			<0.001			<0.001
Female	ref			ref			ref			ref		
Male	1.479	1.301-1.681	<0.001	1.247	1.096-1.419	0.001	1.896	1.740-2.065	<0.001	1.54	1.413-1.679	<0.001
T			<0.001			<0.001			<0.001			<0.001
T0/1	ref			ref			ref			ref		
T2	3.976	3.833-4.123	<0.001	2.392	2.301-2.485	<0.001	2.265	2.212-2.319	<0.001	1.775	1.731-1.821	<0.001
T3	9.073	8.679-9.485	<0.001	4.347	4.143-4.561	<0.001	4.034	3.900-4.172	<0.001	2.822	2.721-2.927	<0.001
T4	24.082	23.115-25.089	<0.001	8.389	8.003-8.793	<0.001	10.287	9.976-10.609	<0.001	5.305	5.119-5.498	<0.001
N			<0.001			<0.001			<0.001			<0.001
N0	ref			ref			ref			ref		
N1	3.582	3.469-3.699	<0.001	2.094	2.024-2.168	<0.001	1.942	1.897-1.987	<0.001	1.443	1.408-1.480	<0.001
N2	6.068	5.817-6.330	<0.001	2.558	2.444-2.677	<0.001	3.01	2.909-3.114	<0.001	1.718	1.657-1.782	<0.001
N3	11.048	10.598-11.517	<0.001	3.67	3.505-3.842	<0.001	5.141	4.968-5.320	<0.001	2.419	2.329-2.512	<0.001
Stage			<0.001						<0.001			
I	ref						ref					
II	3.899	3.711-4.097	<0.001				1.866	1.816-1.918	<0.001			
III	13.158	12.528-13.820	<0.001				4.125	4.004-4.250	<0.001			
IV	67.848	64.776-71.066	<0.001				18.371	17.873-18.884	<0.001			
ER			<0.001			<0.001			<0.001			<0.001
Negative	ref			ref			ref			ref		
Positve	0.328	0.319-0.337	<0.001	0.726	0.699-0.755	<0.001	0.474	0.463-0.484	<0.001	0.761	0.738-0.786	<0.001
PR			<0.001			<0.001			<0.001			<0.001
Negative	ref			ref			ref			ref		
Positve	0.34	0.331-0.349	<0.001	0.609	0.587-0.632	<0.001	0.491	0.481-0.500	<0.001	0.712	0.692-0.732	<0.001
HER2			<0.001			<0.001			<0.001			<0.001
Negative	ref			ref			ref			ref		
Positve	1.342	1.298-1.389	<0.001	0.696	0.672-0.721	<0.001	1.109	1.079-1.140	<0.001	0.758	0.737-0.779	<0.001
Subtype			<0.001						<0.001			
HR+/HER2-	ref						ref					
HR+/HER2+	1.468	1.405-1.533	<0.001				1.15	1.111-1.189	<0.001			
HR-/HER2+	2.415	2.294-2.542	<0.001				1.629	1.561-1.700	<0.001			
HR-/HER2-	3.432	3.327-3.540	<0.001				2.326	2.269-2.385	<0.001			
Grade			<0.001			<0.001			<0.001			<0.001
I	ref			ref			ref			ref		
II	2.852	2.684-3.031	<0.001	1.763	1.658-1.875	<0.001	1.517	1.469-1.567	<0.001	1.157	1.119-1.196	<0.001
III	7.578	7.147-8.035	<0.001	2.747	2.581-2.923	<0.001	2.748	2.662-2.837	<0.001	1.531	1.477-1.586	<0.001
IV	10.812	9.202-12.702	<0.001	3.052	2.593-3.592	<0.001	3.791	3.326-4.319	<0.001	1.764	1.546-2.012	<0.001
Race			<0.001			<0.001			<0.001			<0.001
White	ref			ref			ref			ref		
Black	1.897	1.833-1.963	<0.001	1.197	1.156-1.240	<0.001	1.548	1.506-1.591	<0.001	1.138	1.106-1.170	<0.001
Other races	0.804	0.763-0.848	<0.001	0.786	0.745-0.829	<0.001	0.697	0.668-0.726	<0.001	0.742	0.711-0.773	<0.001
Age group (years)			<0.001			<0.001			<0.001			<0.001
<35	ref			ref			ref			ref		
≥35,<65	0.604	0.557-0.655	<0.001	1.038	0.957-1.126	0.372	0.691	0.641-0.745	<0.001	1.056	0.979-1.139	0.161
≥65	0.727	0.670-0.789	<0.001	1.683	1.550-1.828	<0.001	1.489	1.382-1.605	<0.001	2.715	2.517-2.929	<0.001
Insurance			<0.001			<0.001			<0.001			<0.001
Uninsured	ref			ref			ref			ref		
Insured	0.391	0.362-0.421	<0.001	0.645	0.597-0.697	<0.001	0.56	0.524-0.599	<0.001	0.631	0.589-0.675	<0.001
Medicaid	0.842	0.778-0.913	<0.001	0.87	0.803-0.943	0.001	1.032	0.962-1.107	0.374	0.916	0.854-0.983	0.015
Year of diagnosis			0.005			0.625			<0.001			0.313
2010	ref			ref			ref			ref		
2011	1.004	0.965-1.044	0.857	1.022	0.983-1.063	0.272	0.994	0.966-1.024	0.706	1.002	0.973-1.032	0.885
2012	0.965	0.926-1.006	0.094	0.994	0.953-1.036	0.77	0.985	0.955-1.017	0.355	0.996	0.965-1.027	0.791
2013	0.971	0.929-1.015	0.191	1.025	0.980-1.071	0.286	0.965	0.933-0.998	0.04	0.988	0.954-1.022	0.469
2014	0.943	0.897-0.991	0.021	1.008	0.959-1.060	0.757	0.955	0.919-0.992	0.019	0.988	0.950-1.027	0.529
2015	0.903	0.850-0.960	0.001	0.991	0.932-1.054	0.778	0.899	0.857-0.943	<0.001	0.947	0.902-0.993	0.026

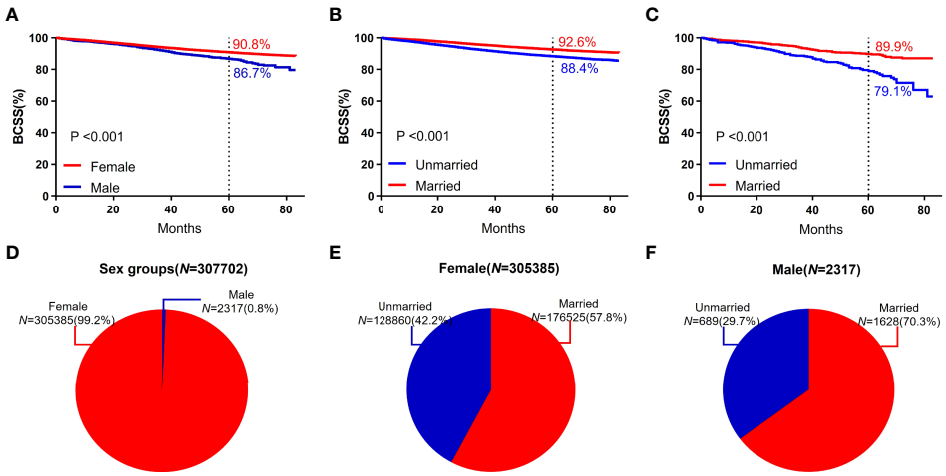


FIGURE 3 | BCSS curves according to the sex groups. **(A)** and subject proportions of the sex groups **(D)**. Survival curves of each marital status for female **(B)** and male **(C)** and subject proportions of each marital status for female **(E)** and male **(F)** were depicted.

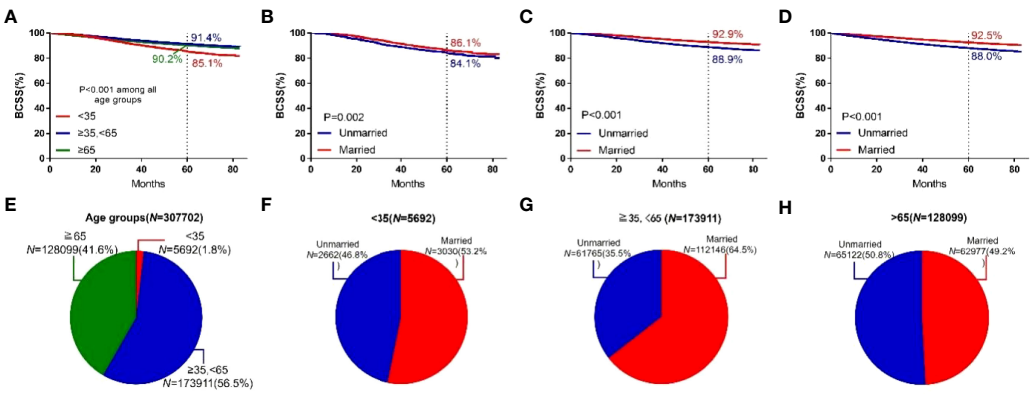


FIGURE 4 | BCSS curves according to the age groups. **(A)** and subject proportions of the age groups **(E)**. Survival curves of each marital status for <35 years old **(B)**, ≥35, <65years old **(C)** and ≧65years old **(D)** and subject proportions of each marital status for <35 years old **(F)**, ≥35, <65years old **(G)** and ≧65years old **(H)** were depicted.

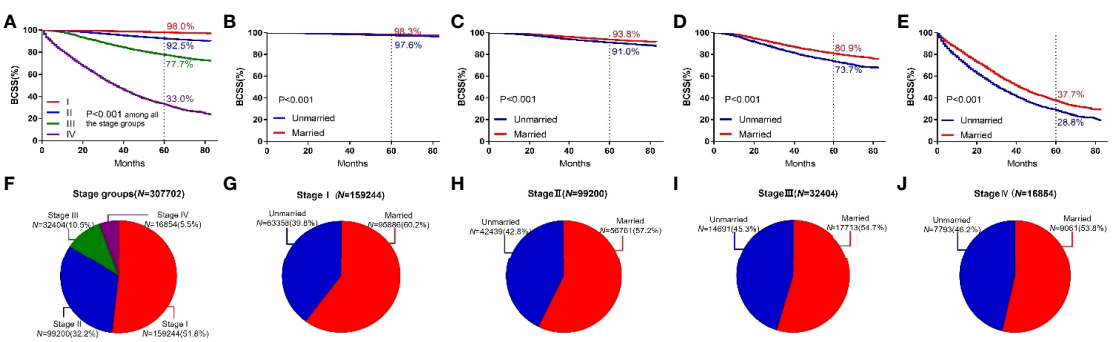


FIGURE 5 | BCSS curves according to the stage groups. **(A)** and subject proportions of the stage groups **(F)**. Survival curves of each marital status for stage I **(B)**, stage II **(C)**, stage III **(D)** and stage IV **(E)** and subject proportions of each marital status for stage I **(G)**, stage II **(H)**, stage III **(I)** and stage IV **(J)** were depicted.

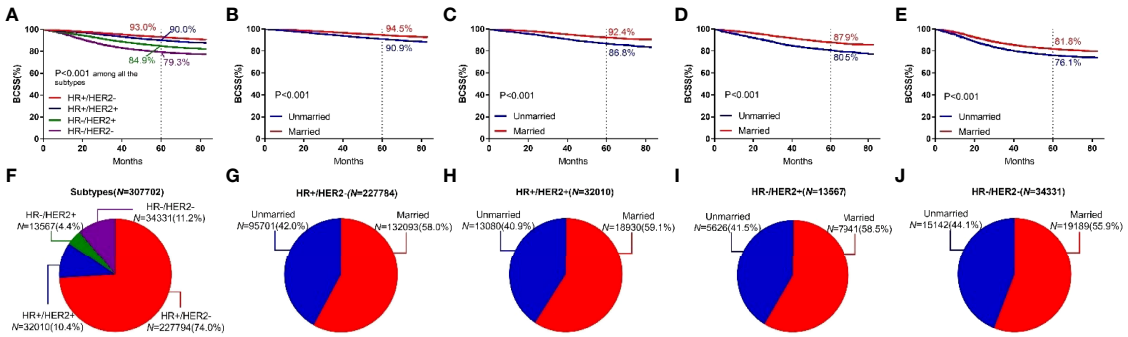


FIGURE 6 | BCSS curves according to the subtypes. **(A)** and subject proportions of the subtypes **(F)**. Survival curves of each marital status for HR+/HER2- **(B)**, HR+/HER2+ **(C)**, HR-/HER2+ **(D)** and HR-/HER2- **(E)** and subject proportions of each marital status for HR+/HER2- **(G)**, HR+/HER2+ **(H)**, HR-/HER2+ **(I)** and HR-/HER2- **(J)** were depicted.

and survival of 3761 patients diagnosed with breast cancer in the SEER database from 1990 to 2011, and found that unmarried patients had an advanced stage and worse diagnosis. In this study, the marital status might have had a more significant impact on the survival of male patients with breast cancer than on female patients (BCSS, HR: 0.533 vs. 0.776). Compared with uninsured and Medicaid patients with breast cancer, insured patients with breast cancer had an earlier stage and better prognosis (26, 27). Molecular typing (28), race (29, 30), and histological grading (31) were also independent factors for the prognosis of breast cancer. In this study, we found that the influence of marital status on prognosis was independent of these factors.

This study also had limitations. First, this was a retrospective study, which had inherent defects, including data bias. Then, little information was available about breast cancer treatment in the SEER database, and hence it was impossible to know the impact of marital status on treatment selection and compliance. Breast cancer does not contribute to the marital breakdown (32). However, a disharmonious partnership in marriage could delay the postoperative recovery of breast cancer and cause a worse prognosis (33). Hence, the prognosis of even married patients

Subgroup	No.of Patient	No.of Event (%)	HR(95%CI)	P value	P value for interaction
Overall	307702	21075(6.8%)	0.774(0.752-0.796)	<0.001	
Sex					<0.001
Female	305385	20851(6.8%)	0.776(0.754-0.799)	<0.001	
Male	2317	224(9.7%)	0.533(0.401-0.707)	<0.001	
Stage					<0.001
I	159244	2116(1.3%)	0.777(0.711-0.849)	<0.001	
II	99200	4994(5.0%)	0.785(0.740-0.832)	<0.001	
III	32404	5230(16.1%)	0.814(0.769-0.862)	<0.001	
IV	16854	8735(51.8%)	0.788(0.754-0.824)	<0.001	
Subtypes					<0.001
HR+/HER2-	227794	11620(5.1%)	0.741(0.713-0.770)	<0.001	
HR+/HER2+	32010	2300(7.2%)	0.746(0.684-0.814)	<0.001	
HR-/HER2+	13567	1556(11.5%)	0.749(0.674-0.832)	<0.001	
HR-/HER2-	34331	5599(16.3%)	0.872(0.825-0.922)	<0.001	
Grade					0.95
I	69163	1168(1.7%)	0.747(0.683-0.816)	<0.001	
II	131650	6202(4.7%)	0.716(0.680-0.755)	<0.001	
III	94235	11380(12.1%)	0.82(0.789-0.853)	<0.001	
IV	845	156(18.5%)	0.841(0.591-1.198)	0.338	
Race					<0.001
White	245255	15796(6.4%)	0.759(0.734-0.784)	<0.001	
Black	33200	3863(11.6%)	0.866(0.807-0.929)	<0.001	
Other races	27860	1399(5.0%)	0.79(0.708-0.882)	<0.001	
Age group(years)					<0.001
<35	5692	587(10.3%)	0.906(0.761-1.079)	0.267	
≥35,<65	173911	11105(6.4%)	0.855(0.822-0.890)	<0.001	
≥65	128099	9383(7.3%)	0.695(0.666-0.725)	<0.001	
Insurance					<0.001
Uninsured	4674	663(14.2%)	0.859(0.730-1.011)	0.067	
Insured	265809	16112(6.1%)	0.763(0.739-0.788)	<0.001	
Medicaid	34507	4073(11.8%)	0.828(0.772-0.888)	<0.001	

FIGURE 7 | Analyses of BCSS (breast cancer-specific survival) for married vs unmarried patients in each subgroup after adjusted by other clinicalpathological factors. Hazard Ratio (HR) estimates for BCSS are indicated by rectangles, and 95% confidence intervals (95% CI) are indicated by the crossing horizontal lines.

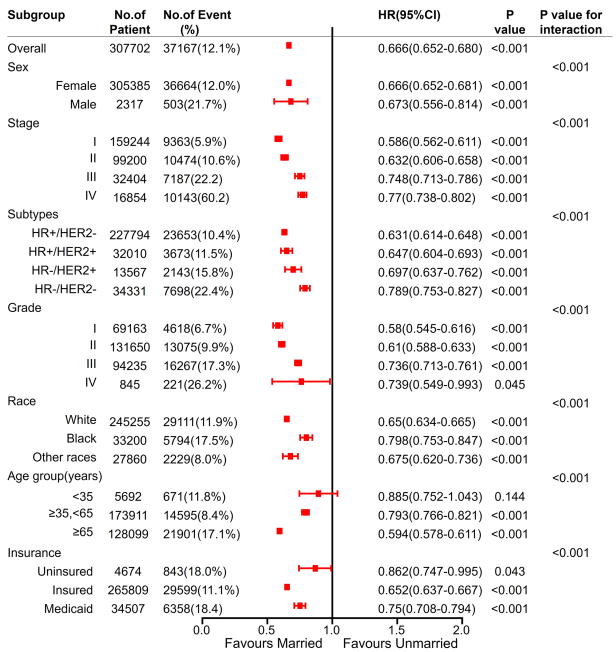


FIGURE 8 | Analyses of OS (Overall survival) for married vs unmarried patients in each subgroup after adjusted by other clinicalpathological factors. Hazard Ratio (HR) estimates for BCSS are indicated by rectangles, and 95% confidence intervals (95% CI) are indicated by the crossing horizontal lines.

varied due to different partnerships. Emotional support was beneficial to the survival of unmarried patients (34). However, this study could not answer how additional psychosocial and emotional supports might improve the survival of unmarried patients and the degree of improvement. Nevertheless, this study might be valuable for understanding the relationship between marital status and prognosis. Due to the large sample size, we could analyze the subgroup of each clinicopathological index by the univariate and multivariate analysis. The final conclusion might help understand how different clinicopathological indexes affect the relationship between marital status and prognosis, and assist the unmarried population who really need emotional intervention.

CONCLUSIONS

This study found for the first time that among patients with different stages of breast cancer, married ones had a better prognosis. Therefore, the effect of marital status on prognosis could not be explained simply by the stage of breast cancer at the time of diagnosis. We also found for the first time that in the subgroup of patients aged <35 years, marital status was not associated with the prognosis of breast cancer. Further in-depth research is needed to clarify this phenomenon. Additionally, it was reported for the first time that marital status was an independent prognosis factor for breast cancer regardless of the molecular typing. These conclusions provide a basis for us to further understand the relationship between marital status and prognosis of breast cancer, and also highlight the necessity of providing emotional support to unmarried patients with breast cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because the data accrued was anonymized at source by the SEER-18 database, therefore no consent was deemed necessary. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

DJ participated in the study design and writing the draft manuscript. YM participated in data collection, collation, and statistical analysis. JZ and HD contributed to study design, results interpretation, and manuscript editing. YY, YZ, and XG participated in the analysis of the results and the revision of the manuscript. ZL conceived the study, interpreted data, and edited the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Baum A, Posluszny D. Health Psychology: Mapping Biobehavioral Contributions to Health and Illness. *Annu Rev Psychol* (1999) 50:137–63. doi: 10.1146/annurev.psych.50.1.137
2. Soler-Vila H, Kasl S, Jones B. Prognostic Significance of Psychosocial Factors in African-American and White Breast Cancer Patients: A Population-Based Study. *Cancer* (2003) 98(6):1299–308. doi: 10.1002/cncr.11670
3. Williams K, Umberson D. Marital Status, Marital Transitions, and Health: A Gendered Life Course Perspective. *J Health Soc Behav* (2004) 45(1):81–98. doi: 10.1177/002214650404500106
4. Aizer A, Chen M-H, McCarthy EP, Mendu ML, Koo S, Wilhite TJ, et al. Marital Status and Survival in Patients With Cancer. *J Clin Oncol Off J Am Soc Clin Oncol* (2013) 31(31):3869–76. doi: 10.1200/JCO.2013.49.6489
5. Fosså S, Cvancarova M, Chen L, Allan AL, Oldenburg J, Peterson DR, et al. Adverse Prognostic Factors for Testicular Cancer-Specific Survival: A Population-Based Study of 27,948 Patients. *J Clin Oncol Off J Am Soc Clin Oncol* (2011) 29(8):963–70. doi: 10.1200/JCO.2010.32.3204
6. Wang L, Wilson SE, Stewart DB, Hollenbeak CS. Marital Status and Colon Cancer Outcomes in US Surveillance, Epidemiology and End Results Registries: Does Marriage Affect Cancer Survival by Gender and Stage? *Cancer Epidemiol* (2011) 35(5):417–22. doi: 10.1016/j.canep.2011.02.004
7. Al-Badawi I, Munkarah AR, Tulbah A, Babic II, Husaini HA, Ahmad S, et al. A Detailed Study of Patients and Tumor Characteristics of Epithelial Ovarian Cancer in Saudi Women. *Int J Gynecol Cancer Off J Int Gynecol Cancer Soc* (2013) 23(3):456–60. doi: 10.1097/IGC.0b013e318284aafe
8. Chen Z, Yang KB, Zhang YZ, Wu CF, Wen DW, Lv JW, et al. Assessment of Modifiable Factors for the Association of Marital Status With Cancer-Specific Survival. *JAMA Netw Open* (2021) 4(5):e2111813. doi: 10.1001/jamanetworkopen.2021.11813
9. Gao Z, Ren F, Song H, Wang Y, Wang Y, Gao Z, et al. Marital Status and Survival of Patients With Chondrosarcoma: A Population-Based Analysis. *Med Sci Monit Int Med J Exp Clin Res* (2018) 24:6638–48. doi: 10.12659/MSM.911673
10. Reyes Ortiz C, Freeman JL, Kuo YF, Goodwin JS. The Influence of Marital Status on Stage at Diagnosis and Survival of Older Persons With Melanoma. *J Gerontol Ser A Biol Sci Med Sci* (2007) 62(8):892–8. doi: 10.1093/gerona/62.8.892
11. Goodwin J, Hunt WC, Key CR, Samet JM.. The Effect of Marital Status on Stage, Treatment, and Survival of Cancer Patients. *JAMA* (1987) 258 (21):3125–30. doi: 10.1001/jama.1987.03400210067027
12. Lee S, Cho E, Grodstein F, Kawachi I, Hu FB, Colditz GA, et al. Effects of Marital Transitions on Changes in Dietary and Other Health Behaviours in US Women. *Int J Epidemiol* (2005) 34(1):69–78.
13. Kobayashi S, Sugiura H, Ando Y, Shiraki N, Yanagi T, Yamashita H, et al. Reproductive History and Breast Cancer Risk. *Breast Cancer (Tokyo Japan)* (2012) 19(4):302–8. doi: 10.1007/s12282-012-0384-8
14. Vishwakarma G, Ndetan H, Nandini Das D, Gupta G, Suryavanshi, Mehta A, et al. Reproductive Factors and Breast Cancer Risk: A Meta-Analysis of Case-Control Studies in Indian Women. *South Asian J Cancer* (2019) 8(2):80–4. doi: 10.4103/sajc.sajc_317_18
15. Wu H, Yang HI, Lin PH, Chen CJ, Santella RM, Terry MB, et al. Reproductive and Environmental Exposures and the Breast Cancer Risk in Taiwanese Women. *Sci Rep* (2021) 11(1):15656. doi: 10.1038/s41598-021-95290-2
16. Brouckaert O, Rudolph A, Laenen A, Keeman R, Bolla MK, Wang Q, et al. Reproductive Profiles and Risk of Breast Cancer Subtypes: A Multi-Center Case-Only Study. *Breast Cancer Res BCR* (2017) 19(1):119. doi: 10.1186/s13058-017-0909-3
17. Katz J, Kertai MD, Cooter M, Greenup RA, Hwang S. Risk Factors for Acute Pain and Its Persistence Following Breast Cancer Surgery. *Pain* (2005) 119:16–25. doi: 10.1016/j.pain.2005.09.008
18. McDonough M, Sabiston C, Wrosch C. Predicting Changes in Posttraumatic Growth and Subjective Well-Being Among Breast Cancer Survivors: The Role of Social Support and Stress. *Psycho-Oncology* (2014) 23(1):114–20. doi: 10.1002/pon.3380
19. Mystakidou K, Tsilika E, Parpa E, Kyriakopoulos D, Malamos N, Damigos D, et al. Personal Growth and Psychological Distress in Advanced Breast Cancer. *Breast (Edinburgh Scotland)* (2008) 17(4):382–6. doi: 10.1016/j.breast.2008.01.006
20. Osborne C, Ostir GV, Du X, Peek MKGoodwin JS. The Influence of Marital Status on the Stage at Diagnosis, Treatment, and Survival of Older Women With Breast Cancer. *Breast Cancer Res Treat* (2005) 93(1):41–7. doi: 10.1007/s10549-005-3702-4
21. Zhai Z, Zhang F, Zheng Y, Zhou L, Tian T, Lin S. Effects of Marital Status on Breast Cancer Survival by Age, Race, and Hormone Receptor Status: A Population-Based Study. *Cancer Med* (2019) 8(10):4906–17. doi: 10.1002/cam4.2352
22. Colleoni M, Rotmensz N, Robertson C, Orlando L, Viale G, Renne G, et al. Very Young Women (<35 Years) With Operable Breast Cancer: Features of Disease at Presentation. *Ann Oncol Off J Eur Soc Med Oncol* (2002) 13(2):273–9. doi: 10.1093/annonc/mdf039
23. Liu Y, Jiang YZ, Yu KD, Shao ZM. Different Patterns in the Prognostic Value of Age for Breast Cancer-Specific Mortality Depending on Hormone Receptor Status: A SEER Population-Based Analysis. *Ann Surg Oncol* (2015) 22 (4):1102–10. doi: 10.1245/s10434-014-4108-5
24. Ahn S, Son BH, Kim SW, Kim SI, Jeong J, Ko S-S, et al. Poor Outcome of Hormone Receptor-Positive Breast Cancer at Very Young Age Is Due to Tamoxifen Resistance: Nationwide Survival Data in Korea–a Report From the Korean Breast Cancer Society. *J Clin Oncol Off J Am Soc Clin Oncol* (2007) 25 (17):2360–8. doi: 10.1200/JCO.2006.10.3754
25. Adekolujo O, Tadisina S, Koduru U, Gernand J, Smith SJ, Kakarala RR, et al. Impact of Marital Status on Tumor Stage at Diagnosis and on Survival in Male Breast Cancer. *Am J Men's Health* (2017) 11(4):1190–9. doi: 10.1177/1557988316669044
26. Franzoi M, Schwartsmann G, de Azevedo SJ, Geib G, Zaffaroni F, Liedke PER, et al. Differences in Breast Cancer Stage at Diagnosis by Ethnicity, Insurance Status, and Family Income in Young Women in the USA. *J Racial Ethnic Health Disparities* (2019) 6(5):909–16. doi: 10.1007/s40615-019-00591-y
27. Hsu C, Wang X, Habif DVJr, Ma CX, Johnson KJ. Breast Cancer Stage Variation and Survival in Association With Insurance Status and Sociodemographic Factors in US Women 18 to 64 Years Old. *Cancer* (2017) 123(16):3125–31. doi: 10.1002/cncr.30722
28. Hwang K, Kim J, Jung J, Chang JH, Chai YJ, Oh SW, et al. Impact of Breast Cancer Subtypes on Prognosis of Women With Operable Invasive Breast Cancer: A Population-Based Study Using SEER Database. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2019) 25(6):1970–9. doi: 10.1158/1078-0432.CCR-18-2782
29. Ko N, Hong S, Winn RA, Calip GS. Association of Insurance Status and Racial Disparities With the Detection of Early-Stage Breast Cancer. *JAMA Oncol* (2020) 6(3):385–92. doi: 10.1001/jamaoncol.2019.5672
30. Cho B, Han Y, Lian M, Colditz GA, Weber JD, Ma C, et al. Evaluation of Racial/Ethnic Differences in Treatment and Mortality Among Women With Triple-Negative Breast Cancer. *JAMA Oncol* (2021) 7(7):1016–23. doi: 10.1001/jamaoncol.2021.1254
31. Rakha E, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic Significance of Nottingham Histologic Grade in Invasive Breast Carcinoma. *J Clin Oncol Off J Am Soc Clin Oncol* (2008) 26(19):3153–8. doi: 10.1200/JCO.2007.15.5986
32. Laitala V, Saarto T, Einiö EK, Martikainen P, Silventoinen K. Early-Stage Breast Cancer Is Not Associated With the Risk of Marital Dissolution in a Large Prospective Study of Women. *Br J Cancer* (2015) 113(3):543–7. doi: 10.1038/bjc.2015.216
33. Yang H, Schuler T. Marital Quality and Survivorship: Slowed Recovery for Breast Cancer Patients in Distressed Relationships. *Cancer* (2009) 115(1):217–28. doi: 10.1002/cncr.23964
34. Wittenberg L, Yutsis M, Taylor S, Giese-Davis J, Bliss-Isberg C, Star P. Marital Status Predicts Change in Distress and Well-Being in Women Newly Diagnosed With Breast Cancer and Their Peer Counselors. *Breast J* (2010) 16(5):481–9. doi: 10.1111/j.1524-4741.2010.00964.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fonc.2022.904292

Disease Burden, Risk Factors, and Trends of Leukaemia: A Global Analysis

Junjie Huang[1], Sze Chai Chan[1], Chun Ho Ngai[1], Veeleah Lok[2], Lin Zhang[3,4], Don Eliseo Lucero-Prisno III[5], Wanghong Xu[6], Zhi-Jie Zheng[7], Edmar Elcarte[8], Mellissa Withers[9] and Martin C. S. Wong[1,4,7*]on behalf of NCD Global Health Research Group Association of Pacific Rim Universities (APRU)*

[1] The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2] Department of Global Public Health, Karolinska Institute, Karolinska University Hospital, Stockholm, Sweden, [3] School of Population and Global Health, The University of Melbourne, Carlton, VIC, Australia, [4] School of Public Health, The Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [5] Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, United Kingdom, [6] School of Public Health, Fudan University, Shanghai, China, [7] Department of Global Health, School of Public Health, Peking University, Beijing, China, [8] College of Nursing, University of the Philippines, Manila, Philippines, [9] Department of Preventive Medicine, Institute for Global Health, University of Southern California, Los Angeles, CA, United States

OPEN ACCESS

Edited by:
Dana Kristjansson, Norwegian Institute of Public Health (NIPH), Norway

Reviewed by:
Wafaa M. Rashed, Children's Cancer Hospital, Egypt
Pierluigi Cocco, Università di Cagliari, Italy

***Correspondence:**
Mellissa Withers
mwithers@usc.edu
Martin C. S. Wong
wong_martin@cuhk.edu.hk

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *25 March 2022*
Accepted: *07 June 2022*
Published: *22 July 2022*

Citation:
Huang J, Chan SC, Ngai CH, Lok V, Zhang L, Lucero-Prisno DE III, Xu W, Zheng Z-J, Elcarte E, Withers M and Wong MCS (2022) Disease Burden, Risk Factors, and Trends of Leukaemia: A Global Analysis. Front. Oncol. 12:904292. doi: 10.3389/fonc.2022.904292

Leukaemia accounted for approximately 2.5% of all new cancer incidence and 3.1% of cancer-related mortality. The investigation of its risk factors and epidemiologic trends could help describe the geographical distribution and identify high-risk population groups. This study aimed to evaluate the global incidence, mortality, associated risk factors, and temporal trends of leukaemia by sex, age, and country. We extracted incidence and mortality of leukaemia from *GLOBOCAN*, *CI5*, *WHO mortality database*, *NORDCAN*, and *SEER*. We searched the *WHO Global Health Observatory* data repository for the age-standardised prevalence of lifestyle and metabolic risk factors. We tested the trends by calculating Average Annual Percentage Change (AAPC) from Joinpoint regression. The age-standardized rate of incidence and mortality were 5.4 and 3.3 per 100,000 globally. The incidence and mortality of leukaemia were associated with Human Development Index, Gross Domestics Products per capita, prevalence of smoking, physical activity, overweight, obesity, and hypercholesterolaemia at the country level. Overall, more countries were showing decreasing trends than increasing trends in incidence and mortality. However, an increasing trend of leukaemia incidence was found in Germany, Korea, Japan, Canada and the United Kingdom (AAPC, 2.32-0.98) while its mortality increased in the Philippines, Ecuador, Belarus, and Thailand (AAPC, 2.49-1.23). There was a decreasing trend of leukaemia for the past decade while an increase in incidence and mortality was observed in some populations. More intensive lifestyle modifications should be implemented to control the increasing trends of leukaemia in regions with these trends. Future studies may explore the reasons behind these epidemiological transitions.

Keywords: Leukaemia, incidence, mortality, risk factors, temporal trend

INTRODUCTION

Leukaemia is a blood-related malignancy characterized by transformed hematopoietic progenitors and diffuse infiltration of bone marrow. The main types of leukaemia include acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL), and chronic myeloid leukaemia (CML). Globally in 2020, leukaemia accounted for approximately 2.5% and 3.1% of all new cancer incidence and mortality, respectively (1). The risk of leukaemia varies among populations of different ages, sexes, and geographical locations (2). Such disparities could be attributable to the difference in the prevalence of different environmental and genetic risk factors for leukaemia.

Risk factors for leukaemia include smoking, exposure to certain chemicals, chemotherapy in the past, radiation exposure, rare congenital diseases, certain blood disorders, family history, age, and gender (3). Due to the recent development of novel therapeutic strategies and targeted drugs, the overall survival of leukaemia patients has shown remarkable improvements (4). The epidemiology of leukaemia may have changed over time and may vary by different population groups. Therefore, it is imperative to examine the global disease distribution, risk factors, and trends of leukaemia to inform the development of its preventive strategies tailored for different countries.

Prior studies are limited to certain countries or captured temporal trends using relatively old data (5–7). Furthermore, none comprehensively determined the lifestyle and metabolic risk factors for leukaemia at a country level. This study aims to: 1) investigate the most updated global incidence and mortality of leukaemia by region, sex, and income level 2) explore the global dietary and socioeconomic factors in differentiating trends in leukaemia incidence and mortality worldwide; and 3) examine the recent incidence and mortality trends of leukaemia for the recent past decade among groups of different ages, sexes, and countries.

METHODS

Data Sources

To retrieve updated statistics on cancer on a global scale, various databases were accessed and explored. The *GLOBOCAN* database was accessed for comprehensive records of the most updated incidence and mortality of leukaemia for 185 countries (8). The Human Development Index (HDI) for each country was extracted from the United Nations (9). Data on gross domestic products (GDP) per capita were retrieved from the World Bank. The prevalence of risk factors for each country was collected from the *Global Health Observatory* (*GHO*) (10), including the prevalence of current smoking, physical inactivity, overweight, obesity, diabetes, and hypercholesterolaemia. For trend analysis, data from the yearly incidence of leukaemia was extracted from the *Cancer Incidence in Five Continents I-X plus (CI5Plus)* for 48 countries. *CI5* is a global cancer database developed by the International Association of Cancer Registries, where the age and sex-associated cancer incidence from different countries can be found to facilitate direct comparison of cancer incidence based on demographic characteristics (11). Data on leukaemia mortality were retrieved from the WHO Mortality Database, where the number of cancer-related deaths is collected (12). In addition, the *Nordic Cancer Registries (NORDCAN) (*13 14) and the *Surveillance, Epidemiology, and End Results (SEER) (*13) were retrieved to obtain the latest leukaemia incidence and mortality data of Northern European countries and the United States, respectively. In our analysis, leukaemia was defined using the International Classification of Diseases 10th revision (ICD-10) C91-95. The countries were divided into nine regions in the trend analysis for presentation, including Asia, Oceania, Northern America, Southern America, Northern Europe, Western Europe, Southern Europe, Eastern Europe, and Africa. For easier reference and comparison, incidence and mortality across countries and age groups were presented in the form of age-standardized rates (ASRs) per 100,000 after adjustment according to the Segi-Doll standardized population.

Statistical Analysis

To present the global incidence and mortality of leukaemia, choropleth maps were constructed. Associations between HDI, GDP per capita, and potential risk factors and incidence and mortality of leukaemia for each country were examined by univariable linear regression analysis for men and women separately. Beta coefficients (β) and the corresponding 95% confidence intervals (CI) were calculated from the regression. The β estimates measure the degree of change in ASR of incidence or mortality of leukaemia per unit increase in the prevalence of risk factors. The corresponding Average Annual Percentage Change (AAPC) for different regions and countries were then calculated for the temporal trend of cancer incidence and mortality of leukaemia on a global scale (14). In trend analysis with transitions, AAPC is preferred over annual percentage change (APC) because it considers the length of the time segment and it does not assume linearity (15). The AAPCs were estimated using Joinpoint regression analysis software, which is developed by the Surveillance, Epidemiology, and End Results Program (SEER) under the United States National Cancer Institute. As a normal practice in epidemiology research for cancer, data of a period of 10 years were used. The ASRs had undergone a logarithmic transformation and related standard errors had been calculated. They were then used to calculate the AAPC and the 95% Confidence Interval (CI) for all countries and both sexes. The epidemiological trends of incidence and mortality are indicated by the AAPC, with a positive AAPC indicating an increasing trend and vice versa. The 95% CI can be used as an indicator to assess the stability of the trend: an interval overlapping with 0 signifies a stable trend without significant temporal change. In this study, the incidence and mortality of the entire population were examined. The incidence rates of different age groups (below 15, between 15-49, 50 or above, and 0-85+) were compared to evaluate the role of ages; results from

both sexes in each group were separately assessed to investigate the role of sex in leukaemia.

RESULTS

Global Incidence of Leukaemia in 2020

A total of 474,519 new cases of leukaemia were reported in 2020 (**Figure 1**). The global age-standardized rate of incidence was 5.4 per 100,000 and there was an almost five-fold variation worldwide. North America (ASR = 10.9), Australia and New Zealand (ASR = 10.4), Western Europe (ASR = 8.5), and Northern Europe (ASR = 8.5) had the highest incidence, whereas the lowest incidence was found in Middle Africa (ASR = 2.2), Western Africa (ASR = 2.3), and Eastern Africa (ASR = 3.3). As far as the sex-specific age-standardized rate is concerned, the ASR of men (6.3) was 40% higher than that of women (4.5) worldwide, and larger differences could be found in regions with higher ASRs. Moreover, it was found that countries with higher income levels had a higher incidence; high-income countries (ASR=8.4) had incidence 1.5 times higher than low-income countries (ASR=3.4).

Global Mortality of Leukaemia in 2020

In terms of mortality, 311,594 related deaths were reported in 2020. There was a smaller regional difference worldwide in mortality from leukaemia, as the majority of regions in Asia, Europe, America, and Australia and New Zealand reported mortality of around 2.5-4.0 per 100,000. It is worth mentioning that Western Asia (ASR = 4.6) had the highest mortality, 40% higher than the world average (ASR = 3.3). Sex was also a pivotal factor in mortality as men (ASR = 4.0) had a mortality of almost 50% higher than women (ASR = 2.7). Regarding the discrepancy in mortality among countries with different levels of income, countries with higher income [upper-middle income (ASR = 3.6), high income (ASR = 3.2)] had around 30% higher risk than lower income (low income (ASR = 2.8) and low-middle income (ASR = 2.7) countries).

Associations Between Risk Factors and Burden of Leukaemia

Among men, higher ASR of incidence of leukaemia was associated with a higher HDI (β=1.27, CI 1.05 to 1.49), GDP per capita (β=0.75, CI 0.56 to 0.95), and higher prevalence of inactivity (β=0.11, CI 0.06 to 0.16), overweight (β=0.10, CI 0.08 to 0.12), obesity (β=0.23, CI 0.18 to 0.27), and hypercholesterolaemia (β=0.31, CI 0.26 to 0.37; **Figure 2**). For women, higher incidence was associated with a higher HDI (β=0.82, CI 0.66 to 0.98), GDP per capita (β=0.45, CI 0.31 to 0.58), and higher prevalence of smoking (β=0.08, CI 0.05 to 0.11), inactivity (β=0.05, CI 0.03 to 0.08), overweight (β=0.06, CI 0.04 to 0.08), obesity (β=0.08, CI 0.04 to 0.11), and hypercholesterolaemia (β=0.22, CI 0.18 to 0.27). Among men, higher ASR of mortality of leukaemia was associated with a higher HDI (β=0.25, CI 0.15 to 0.36), and higher prevalence of smoking (β=0.02, CI 0.001 to 0.03), inactivity (β=0.03, CI 0.01 to 0.05), overweight (β=0.02, CI 0.02 to 0.03), obesity (β=0.05, CI 0.03 to 0.07), and hypercholesterolaemia (β=0.05, CI 0.02 to 0.08; **Figure 3**). For women, higher mortality was associated with a higher HDI (β=0.37, CI 0.24 to 0.50) and

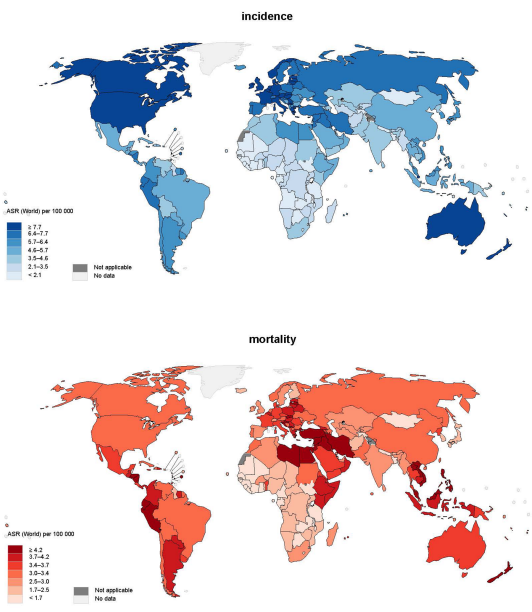


FIGURE 1 | Global incidence and mortality of leukaemia, both sexes, all ages, in 2020.

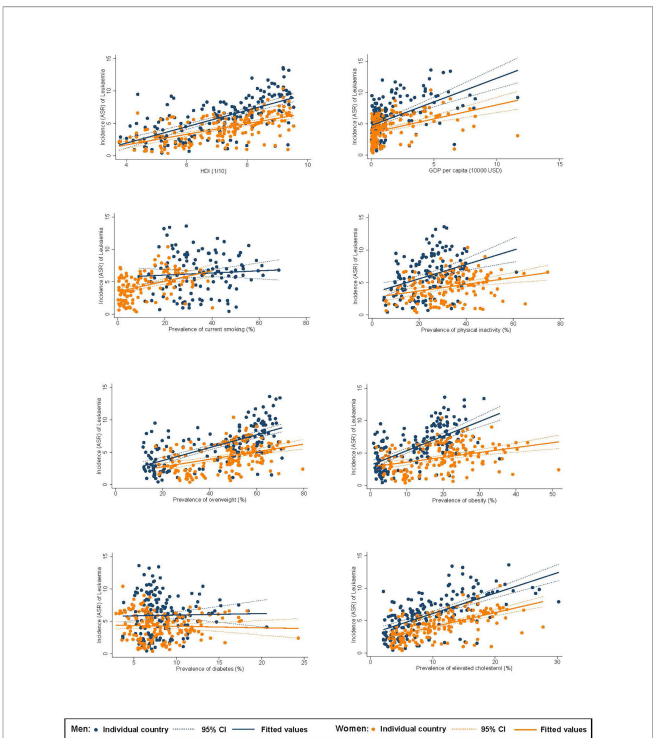


FIGURE 2 | Associations between risk factors and incidence of leukaemia.

higher prevalence of smoking (β=0.03, CI 0.01 to 0.06), inactivity (β=0.03, CI 0.01 to 0.05), overweight (β=0.04, CI 0.02 to 0.05), obesity (β=0.04, CI 0.02 to 0.07), and hypercholesterolaemia (β=0.09, CI 0.06 to 0.13).

Temporal Trends of Leukaemia

The incidence and mortality trends of leukaemia for each country between 1980 and 2017 are shown in **Supplementary Figure 1**, and the trend regression is presented in **Supplementary Figure 2**. Overall, more countries were showing decreasing trends than increasing trends in incidence in all age groups and both sexes, and such discrepancy was particularly significant for women. As for mortality, more countries were showing decreasing trends than increasing trends for both sexes, and this discrepancy was particularly significant for European countries.

Incidence Trends of Individuals Aged 0-85+

Among men, five countries showed increasing trends in incidence (**Figure 4**), including, in descending order, Germany (AAPC = 2.32, 95% CI [0.50, 4.17], p value = 0.019), Korea (AAPC = 1.78, 95% CI [0.66, 2.91], p value = 0.006), Japan (AAPC = 1.19, 95% CI [0.37, 2.02], p value = 0.010), Canada (AAPC = 1.10, 95% CI [0.21, 2.00], p value = 0.010) and the United Kingdom (AAPC = 0.98, 95% CI [0.26 to 1.70], p value = 0.014). By contrast, seven countries had decreasing trends, with Brazil (AAPC = -14.10, 95% CI [-18.72, -9.21], p value < 0.001), Costa Rica (AAPC = -7.18, 95% CI [-11.13, -3.06], p value = 0.004), and the Philippines (AAPC = -3.76, 95% CI [-5.34, -2.15], p value = 0.001) reported the most drastic decreases. Among women, only the United Kingdom (AAPC = 1.23, 95% CI [0.26 to 2.20], p value = 0.019) reported an increasing trend in incidence. Conversely, nine countries reported decreasing trend, as Brazil (AAPC = -16.04, 95% CI [-25.94, -4.82],

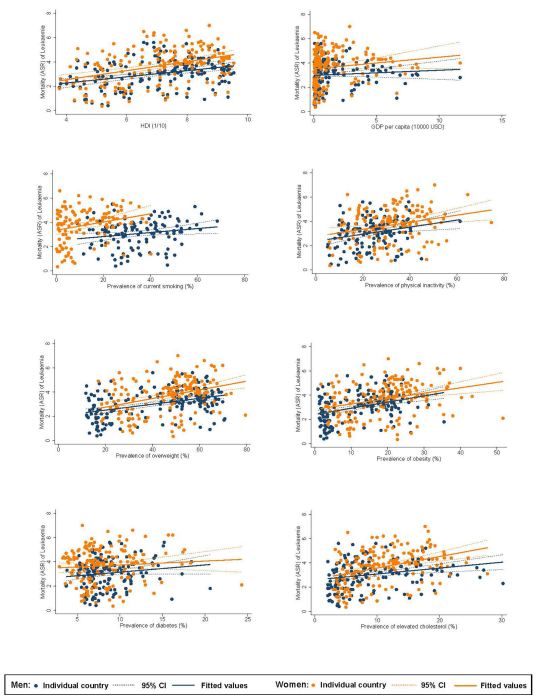


FIGURE 3 | Associations between risk factors and mortality of leukaemia.

p value = 0.012), Costa Rica (AAPC = -7.45, 95% CI [-11.67, -3.03], p value = 0.005), and Cyprus (AAPC = -6.36, 95% CI [-9.97, -2.61], p = 0.005) showed the most significant decrease.

Mortality Trends of Individuals Aged 0-85+

Considering male patients, 22 countries, including 15 European countries, showed significant decreasing trends in mortality

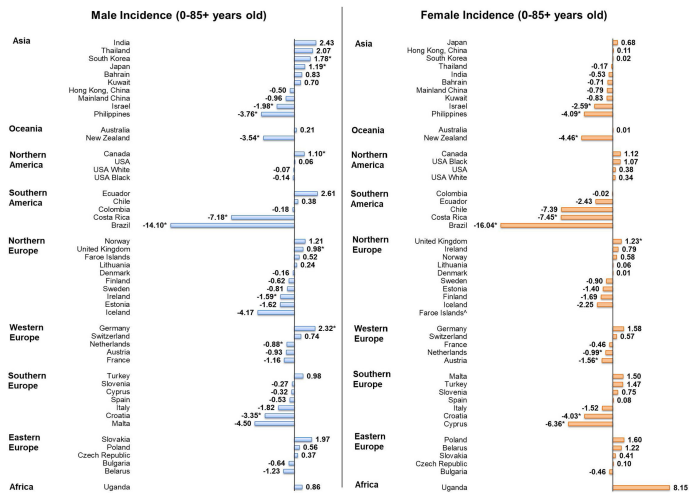


FIGURE 4 | AAPC of incidence of leukaemia in individuals aged 0 – 85+ years old. * p values < 0.05.

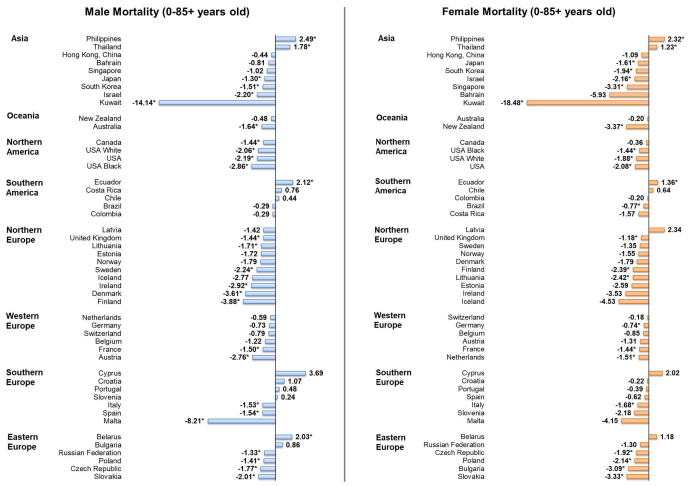


FIGURE 5 | AAPC of mortality of leukaemia in individuals aged 0 – 85+ years old. * p values < 0.05.

(**Figure 5**). Countries with the most significant decrease were Kuwait (AAPC = -14.14, 95% CI [-24.51, -2.36], p = 0.026), Malta (AAPC = -8.21, 95% CI [-14.30 to -1.68], p value = 0.021) and Finland (AAPC = -3.88, 95% CI [-5.80, -1.93], p value = 0.002). In contrast, four countries reported significant increasing trends, which included the Philippines (AAPC = 2.49, 95% CI [1.03, 3.97], p value = 0.004), Ecuador (AAPC = 2.12, 95% CI [0.55, 3.72], p value = 0.014), Belarus (AAPC = 2.03, 95% CI [0.58, 3.50], p value = 0.012), and Thailand (AAPC = 1.78, 95% CI [1.07, 2.49], p value < 0.001). Considering female patients, 19 countries showed significant decreasing trends in mortality and only three countries showed significant increasing trends. Kuwait (AAPC = -18.48, 95% CI [-26.76, -9.26], p = 0.002), New Zealand (AAPC = -3.37, 95% CI [-6.31, -0.35], p = 0.033) and Slovakia (AAPC = -3.33, 95% CI [-5.20, -1.43], p value = 0.004) were the countries showing the most drastic decrease, while the increasing trends were reported in the Philippines (AAPC = 2.32, 95% CI [0.94, 3.72], p value = 0.005), Ecuador (AAPC = 1.36, 95% CI [0.04, 2.69], p value = 0.045), and Thailand (AAPC = 1.23, 95% CI [0.45, 2.02], p value = 0.007).

Incidence Trends of Individuals in Specific Age Groups

For men aged 50 or above, five countries showed significant increases in incidence and eight countries showed significant decreases, with India (AAPC = 6.59, 95% CI [2.84, 10.49], p value = 0.003) showing the largest increase and Brazil (AAPC = -16.04, 95% CI [-23.35, -8.04], p value < 0.001) showing the largest decrease (**Supplementary Figure 3**). For the younger men aged between 15-49, four countries reported significant decreases and no countries reported significant increases, Brazil (AAPC = -17.70; 95% CI [-26.62, -7.71]; p value = 0.004) showed the largest decrease (**Supplementary Figure 4**). For the youngest age group of boys aged 14 or below, New Zealand (AAPC = 7.38; 95% CI [4.54, 10.29]; p value < 0.001) and Korea (AAPC = 3.11; 95% CI [0.89, 5.39]; p value = 0.012) reported significant increasing trends while decreasing trends were found in Brazil (AAPC = -8.96; 95% CI [-15.7, -1.68]; p value = 0.023) and Philippines (AAPC = -4.52; 95% CI [-7.78, -1.15]; p value = 0.015) (**Supplementary Figure 5**). For women aged 50 or above, only the United Kingdom (AAPC = 1.97, 95% CI [1.09, 2.86], p value = 0.001) reported a significant increase in incidence and 11 countries reported significant decreases, with Costa Rica showing the most significant decrease (AAPC = -15.50, 95% CI [-22.03, -8.42], p value = 0.001); for the younger women aged between 15-49, four countries and two countries reported significant increases and decreases respectively, in which Uganda (AAPC = 9.12; 95% CI [1.12, 17.75]; p value = 0.03) had the largest increase and Brazil (AAPC = -18.76; 95% CI [-33.92, -0.12]; p value = 0.049) had the largest decrease. For girls aged 14 or below, only Belarus (AAPC = 5.71; 95% CI [2.02, 9.54]; p value = 0.007) showed a significant increasing trend whereas decreasing trends were observed in Bahrain (AAPC = -11.87; 95% CI [-22.16, -0.23]; p value = 0.047) and three other countries.

DISCUSSION

Summary of Major Findings

This analysis provides the most updated evaluation of the global burden, risk factors, and epidemiologic trends of leukaemia by age, sex, and country using data from cancer registries. We have several major findings: 1) the highest incidence and mortality rates of leukaemia were observed in countries with higher income and among males; 2) higher incidence and mortality of leukaemia were associated with a higher HDI, GDP per capita, prevalence of smoking, inactivity, overweight, obesity, and hypercholesterolaemia; 3) there was an overall decreasing trend of leukaemia for the recent past decade, while an increasing incidence and mortality were observed in some populations, including men and younger individuals.

Explanation of Findings and Relationship With Literature

There was a disparity in the distribution of leukaemia across different regions in 2020. The study found the highest incidence of leukaemia in North America, Australia and New Zealand, Western Europe, and Northern Europe. The burden of leukaemia was associated with HDI and GDP per capita at the country level. These findings are generally consistent with a previous study which concluded regions with high Socio-demographic Index (SDI) usually had higher ASRs of leukaemia (7). The reasons behind this phenomenon remain unexplored but may be related to genetics (16). Other possible factors may include a higher prevalence of environmental and lifestyle risk factors, metabolic diseases, and level of technology and capacity of detection for leukaemia in more developed regions (19 20). Also, low- and middle- income countries (LMICs) often have a lower life expectancy, the relatively lower mortality may be attributable to the occurrence of competing causes of deaths. The incidence and mortality of leukaemia were higher in men than in women, which was likely due to a higher level of exposure to leukaemia-related risk factors among men, including lifestyle risk factors (e.g., smoking) and occupational factors (e.g., ionizing and non-ionizing radiation, hydrocarbons and pesticides) (17–21).

This study found some preventable and common lifestyle and metabolic risk factors associated with the incidence and mortality of leukaemia at a country level, including smoking, physical inactivity, overweight, obesity, and hypercholesterolaemia. The results are generally supported by the findings of previous individual-level observational research on the association between these risk factors and risk of leukaemia. For instance, a meta-analysis of 23 studies showed that current and ever smokers have 40% (RR 1.40, 95% CI 1.22–1.60) and 25% (RR 1.25, 95% CI 1.15–1.36) increased risk of developing AML when compared with non-smokers (22). A large study of 1.44 million participants concluded high vs low levels of leisure-time physical activity were associated with lower risks of myeloid leukaemia (HR, 0.80; 95% CI, 0.70-0.92) (23). Another meta-analysis found the RRs of leukaemia were 1.14 [95% confidence interval (CI), 1.03–1.25] for overweight participants and 1.39 (95% CI, 1.25–1.54) for obese participants (24). Evidence also suggested there is a high incidence of hypercholesterolemia in CLL patients (25).

For the recent past decade, there was an overall decreasing trend of leukaemia incidence and mortality. The attributable factors to this favourable trend may include: 1) progress in therapies for leukaemia and their associated treatment-related prognosis; 2) reduction in exposure to environmental risk factors and smoking; 3) decrease in childhood leukaemia; 4) increase in intake of folate and vitamin supplementation during pregnancy; and 5) expanded genetic screening for high-risk germline mutations (5, 26–31). Nevertheless, we also observed a significant increase in leukaemia incidence and mortality in some populations. The incidence increases in more developed countries may be likely due to continuous improvement in the technology and capacity of detection for leukaemia so that more leukaemia cases were diagnosed and recorded. On the contrary, the mortality increases in less developed countries are more concerning since this unfavourable trend may be driven by the increasing prevalence of risk factors for leukaemia in these regions. More intensive risk modifications are therefore required for these countries.

Limitations

There are some limitations to the current study. Firstly, there could be under-reporting of the incidence and mortality of leukaemia in the developing countries due to the underdevelopment of infrastructure and mechanism of cancer reporting in these regions. Secondly, numbers might have been overestimated for some countries since their figures were represented by cancer registries of the major cities. Thirdly, a direct comparison between different countries could be difficult as the cancer registration might have changed over time. However, this limitation is of less concern when we compared the incidence and mortality of leukaemia according to age and sex groups within the same region. Furthermore, there was a lack of analysis on the trend of the different subtypes of leukaemia. As the geographical distribution, risk factors, and epidemic trends could vary by different subtypes of leukaemia, this information bare important implications for diseases prevention. Lastly, linking exposure at a country-level to individuals and controlling for confounders may be difficult based on the ecological epidemiological design of the study. Possible confounders may include the prolonged life expectancy in countries with higher HDI, while the association between BMI and hypercholesterolaemia and the burden of leukaemia might be confounded by the more accurate diagnostic procedures in countries with a higher HDI. Therefore, the findings between the exposure and the trends should be interpreted with caution.

Conclusions

The incidence and mortality of leukaemia has been decreasing for the past decade likely due to the recent development of novel therapeutic strategies and targeted drugs for leukaemia. However, an increasing trend of leukaemia incidence was found in Germany, Korea, Japan, Canada and the United Kingdom while mortality increased in Ecuador, Belarus, Thailand, and the Philippines. Intensive lifestyle modifications including further smoking reduction, physical activity, weight control, and optimal management of hypercholesterolaemia might be beneficial to reduce the risk of leukaemia, especially among men and younger individuals. It is also important to improve early detection, treatment, surveillance, and quality of life for patients with leukaemia. Lastly, further longitudinal research is required to explore the reasons behind these epidemiologic trends observed and give more insights into the specific aetiology and prognosis of leukaemia by different subtypes. Study should be done to confirm the association between the lifestyle factors and the risk of leukaemia at an individual level.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

This study was approved by the Survey and Behavioural Research Ethics Committee, the Chinese University of Hong Kong (No. SBRE-20-332).

AUTHOR CONTRIBUTIONS

JH and MCSW participated in the conception of the research ideas, study design, interpretation of the findings, writing of the first draft of the manuscript, and provided intellectual input to the translational aspects of the study. SC, CN, and VL retrieved information from the relevant databases, performed the statistical analysis, and presented the methodology and results. MW, LZ, DL-P, WX, Z-JZ, and EE made critical revisions on the manuscripts and provided expert opinions on implications of the study findings. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank the staff at the World Health Organization and its collaborators who prepared these publicly available data. This study is based on publicly available data and solely reflects the opinion of its authors and not that of the World Health Organization.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.904292/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Lim JY, Bhatia S, Robison LL, Yang JJ. Genomics of Racial and Ethnic Disparities in Childhood Acute Lymphoblastic Leukemia. *Cancer* (2014) 120 (7):955–62. doi: 10.1002/cncr.28531
3. Bispo JAB, Pinheiro PS, Kobetz EK. Epidemiology and Etiology of Leukemia and Lymphoma. *Cold Spring Harb Perspect Med* (2020) 10(6). doi: 10.1101/cshperspect.a034819
4. Dombret H, Gardin C. An Update of Current Treatments for Adult Acute Myeloid Leukemia. *Blood* (2016) 127(1):53–61. doi: 10.1182/blood-2015-08-604520
5. Dong Y, Shi O, Zeng Q, Lu X, Wang WJ, Li Y, et al. Leukemia Incidence Trends at the Global, Regional, and National Level Between 1990 and 2017. *Exp Hematol Oncol* (2020) 9(1):14. doi: 10.1186/s40164-020-00170-6
6. Ou Z, Yu D, Liang Y, He W, Li Y, Zhang M, et al. Analysis of the Global Burden of Disease Study Highlights the Trends in Death and Disability-Adjusted Life Years of Leukemia From 1990 to 2017. *Cancer Commun* (2020) 40(11):598–610. doi: 10.1002/cac2.12094
7. Lin X, Wang J, Huang X, Wang H, Li F, Ye W, et al. Global, Regional, and National Burdens of Leukemia From 1990 to 2017: A Systematic Analysis of the Global Burden of Disease 2017 Study. *Aging (Albany NY)* (2021) 13 (7):10468–89. doi: 10.18632/aging.202809
8. International Agency for Research on Cancer WHO. *Data Visualization Tools for Exploring the Global Cancer Burden in 2020* (2020). Available at: https://gco.iarc.fr/today/home (Accessed 10 May 2021).
9. United Nations. *Human Development Classification, 2018 Statistical Update* (2018). Available at: http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf (Accessed 30 July 2021).
10. World Health Organization. *Global Health Observatory Data Repository* (2016). Available at: https://apps.who.int/gho/data/node.main (Accessed 10 May 2021).
11. Bray F CM, Mery L, Piñeros M, Znaor A, Zanetti R and Ferlay J. *Cancer Incidence in Five Continents, Vol. XI (Electronic Version) Lyon: International Agency for Research on Cancer*. Available at: https://ci5.iarc.fr (Accessed 10 May 2021).
12. World Health Organization. *WHO Mortality Database* (2020). Available at: https://www.who.int/healthinfo/mortality_data/en/ (Accessed 10 May 2021).
13. U.S. Department of Health and Human Services NIoH. *Surveillance, Epidemiology, and End Results (SEER) Program* (2020). Available at: https://seer.cancer.gov/about/ (Accessed 10 May 2021).
14. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation Tests for Joinpoint Regression With Applications to Cancer Rates. *Stat Med* (2000) 19(3):335–51. doi: 10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.0.co;2-z
15. Clegg LX, Hankey BF, Tiwari R, Feuer EJ, Edwards BK. Estimating Average Annual Percent Change in Trend Analysis. *Stat Med* (2009) 28(3):3670–82. doi: 10.1002/sim.3733
16. Stieglitz E, Loh ML. Genetic Predispositions to Childhood Leukemia. *Ther Adv Hematol* (2013) 4(4):270–90. doi: 10.1177/2040620713498161
17. Colamesta V, D'Aguanno S, Breccia M, Bruffa S, Cartoni C, La Torre G, et al. Do the Smoking Intensity and Duration, the Years Since Quitting, the Methodological Quality and the Year of Publication of the Studies Affect the Results of the Meta-Analysis on Cigarette Smoking and Acute Myeloid Leukemia (AML) in Adults? *Crit Rev Oncol Hematol* (2016) 99:376–88. doi: 10.1016/j.critrevonc.2016.01.003
18. Belson M, Kingsley B, Holmes A. Risk Factors for Acute Leukemia in Children: A Review. *Environ Health Perspect* (2007) 115(1):138–45. doi: 10.1289/ehp.9023
19. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* (2017) 3(4):524–48. doi: 10.1001/jamaoncol.2016.5688
20. Fitzmaurice C, Akinyemiju TF, Al Lami FH, Alam T, Alizadeh-Navaei R, Allen C, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* (2018) 4(11):1553–68. doi: 10.1001/jamaoncol.2018.2706
21. Bonner MR, Williams BA, Rusiecki JA, Blair A, Freeman LEB, Hoppin JA, et al. Occupational Exposure to Terbufos and the Incidence of Cancer in the Agricultural Health Study. *Cancer Causes Control* (2010) 21(6):871–7. doi: 10.1007/s10552-010-9514-9
22. Fircanis S, Merriam P, Khan N, Castillo JJ. The Relation Between Cigarette Smoking and Risk of Acute Myeloid Leukemia: An Updated Meta-Analysis of Epidemiological Studies. *Am J Hematol* (2014) 89(8):E125–E32. doi: 10.1002/ajh.23744
23. Moore SC, Lee I-M, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of Leisure-Time Physical Activity With Risk of 26 Types of

Cancer in 1.44 Million Adults. *JAMA Internal Med* (2016) 176(6):816–25. doi: 10.1001/jamainternmed.2016.1548
24. Larsson SC, Wolk A. Overweight and Obesity and Incidence of Leukemia: A Meta-Analysis of Cohort Studies. *Int J Cancer* (2008) 122(6):1418–21. doi: 10.1002/ijc.23176
25. Chow S, Buckstein R, Spaner DE. A Link Between Hypercholesterolemia and Chronic Lymphocytic Leukemia. *Leuk Lymphoma* (2016) 57(4):797–802. doi: 10.3109/10428194.2015.1088651
26. Danckert B FJ, HL H, TB J, Khan S, JE Køtlum, Ólafsdóttir E, et al. *NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.2* (Accessed 10 May 2021).
27. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint Å, et al. NORDCAN–A Nordic Tool for Cancer Information, Planning, Quality Control and Research. *Acta Oncol* (2010) 49(5):725–36. doi: 10.3109/02841861003782017
28. Mahoney MC, Moysich KB, McCarthy PL Jr., McDonald RC, Stepanenko VF, Day RW, et al. The Chernobyl Childhood Leukemia Study: Background & Lessons Learned. *Environ Health* (2004) 3(1):12. doi: 10.1186/1476-069x-3-12
29. Metayer C, Petridou E, Aranguré JM, Roman E, Schuz J, Magnani C, et al. Parental Tobacco Smoking and Acute Myeloid Leukemia: The Childhood Leukemia International Consortium. *Am J Epidemiol* (2016) 184(4):261–73. doi: 10.1093/aje/kww018
30. Metayer C, Milne E, Dockerty JD, Clavel J, Pombo-de-Oliveira MS, Wesseling C, Spector LG, et al. Maternal Supplementation With Folic Acid and Other Vitamins and Risk of Leukemia in Offspring: A Childhood Leukemia International Consortium Study. *Epidemiology* (2014) 25(6):811–22. doi: 10.1097/ede.0000000000000141
31. Ogasawara A, Matsushita H, Tanaka Y, Shirasugi Y, Ando K, Asai S, et al. A Simple Screening Method for the Diagnosis of Chronic Myeloid Leukemia Using the Parameters of a Complete Blood Count and Differentials. *Clin Chim Acta* (2019) 489:249–53. doi: 10.1016/j.cca.2018.08.038

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 August 2022
DOI 10.3389/fonc.2022.901098

OPEN ACCESS

EDITED BY
Dana Kristjansson,
Norwegian Institute of Public Health (NIPH), Norway

REVIEWED BY
Yan-Shen Shan,
National Cheng Kung University Hospital, Taiwan
João Botelho,
Egas Moniz Interdisciplinary Research Center, Portugal
Hee Sam Na,
Pusan National University, South Korea
Marwa Abbass,
Cairo University, Egypt
Kristina Bertl,
Malmö University, Sweden
Nina Shenoy,
Nitte University, India

*CORRESPONDENCE
Han Sang Kim
MODEERFHS@yuhs.ac
Inkyung Jung
IJUNG@yuhs.ac

†These authors share first authorship

‡These authors share senior authorship

SPECIALTY SECTION
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

RECEIVED 21 March 2022
ACCEPTED 01 August 2022
PUBLISHED 23 August 2022

CITATION
Kim EH, Nam S, Park CH, Kim Y, Lee M, Ahn JB, Shin SJ, Park YR, Jung HI, Kim B-I, Jung I and Kim HS (2022) Periodontal disease and cancer risk: A nationwide population-based cohort study.
Front. Oncol. 12:901098.
doi: 10.3389/fonc.2022.901098

Periodontal disease and cancer risk: A nationwide population-based cohort study

Eun Hwa Kim[1†], Sunghyun Nam[2], Chung Hyun Park[3], Yitak Kim[2], Myeongjee Lee[1], Joong Bae Ahn[4], Sang Joon Shin[4], Yu Rang Park[2], Hoi In Jung[5], Baek-Il Kim[5], Inkyung Jung[6*‡] and Han Sang Kim[4,7,8*†‡]

[1]Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea, [2]College of Medicine, Yonsei University, Seoul, South Korea, [3]Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea, [4]Yonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea, [5]Department of Preventive Dentistry and Public Oral Health, BK21 FOUR Project, Yonsei University College of Dentistry, Seoul, South Korea, [6]Division of Biostatistics, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea, [7]Graduate School of Medical Science, Brain Korea 21 Project, Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul, South Korea, [8]Institute for Innovation in Digital Healthcare (IIDH), Severance Hospital, Seoul, South Korea

Background: Although emerging evidence suggests that periodontitis might increase the risk of cancer, comorbidity and lifestyle behaviors, such as smoking and body mass index (BMI), may have confounded this reported association. This study aimed to investigate whether chronic periodontitis is associated with cancer risk using a large, nationwide database.

Methods: We conducted a population-based, retrospective cohort study using data from the Korean National Health Insurance Cohort Database obtained between January 2003 and December 2015. We included 713,201 individuals without a history of cancer who were followed up to 10 years. Confounding factors included demographic factors (age, sex, income, and residential area), lifestyle behaviors (smoking history and BMI), and comorbidities, such as hypertension, diabetes, heart failure, and pulmonary disease, using the Charlson Comorbidity Index. Multivariable Cox regression analysis was applied to estimate the adjusted hazard ratio (aHR) for cancer risk.

Results: Of the 713,201 participants, 53,075 had periodontitis and were placed in the periodontitis group; the remaining 660,126 individuals were included as the control group. Overall, the cumulative incidence of cancer in the periodontitis group was 2.2 times higher than that in the control group. The periodontitis group had an increased risk of total cancer compared to the control group after adjusting for age, sex, comorbidities, BMI, and smoking history (aHR, 1.129; 95% confidence interval [CI], 1.089-1.171; $P<0.0001$). When examining specific cancer types, significant associations were also observed between periodontitis and stomach cancer (aHR, 1.136; 95% CI, 1.042-1.239; $P=0.0037$), colon cancer (aHR, 1.129; 95% CI, 1.029-1.239; $P=0.0105$), lung

cancer (aHR, 1.127; 95% CI, 1.008-1.260; *P*=0.0353), bladder cancer (aHR, 1.307; 95% CI, 1.071-1.595; *P*=0.0085), thyroid cancer (aHR, 1.191; 95% CI, 1.085-1.308; *P*=0.0002), and leukemia (aHR, 1.394; 95% CI, 1.039-1.872; *P*=0.0270). There was no significant association between the development of secondary malignancy and periodontitis in cancer survivors who were alive 5 years after they were diagnosed with the primary malignancy.

Conclusions: Periodontal disease, including periodontitis, was associated with increased risk of cancer, which persisted after controlling for confounding factors. Further prospective research is warranted to establish a causal relationship.

KEYWORDS

periodontal disease, cancer risk, cohort study, periodontitis, oral inflammation

Introduction

Periodontal disease is an inflammatory disorder of the periodontal tissue induced by dysbiotic plaque. It can range from a mild form, such as gingivitis, to a more severe, destructive form, such as periodontitis, which occurs as a result of the destruction of the attachment apparatus, including the alveolar bone, the periodontal ligament that subsequently to tooth loss (1, 2). Periodontitis is an evolving disease and a recently updated classification framework based on a staging and grading system incorporating severity, tooth loss, and management complexity (3). The global prevalence of periodontal disease is 20-50%, and approximately 10% of the global population is affected by severe periodontitis (4–6). Recently, intensive efforts have been made to elucidate the effects of the dysbiotic oral microbiome on various systemic diseases, including cardiovascular disease and cancer (7).

Previous observational reports and meta-analyses reported that the presence of periodontal disease positively correlates with an increased risk of total cancer and site-specific cancers (8–12). Recent prospective studies have reported increases in the overall cancer risk associated with periodontal disease of 14% to 24%, and the association was not attenuated even after adjustment for known risk factors, such as smoking (10, 12, 13). Although the methodology to define periodontal disease is not consistent across studies, multiple population-based studies have shown a consistent relationship between periodontitis and cancer risk, and the risk seems to increase significantly in proportion to disease severity (13). However, our understanding of the relationship between periodontal disease and site-specific cancer risk is limited, which makes it difficult to reach a consensus. A meta-analysis that reviewed 14 cohort and 20 case-control studies reported positive associations between periodontitis and oral, lung, and pancreatic cancers (12). Other recent cohort studies have shown positive associations between periodontitis and esophageal, breast, lung, gallbladder, and colorectal cancers and melanoma (10, 13). However, these conflicting relationships in specific tumor types may also be explained by the differences in study populations, cohort sizes, study designs, particularly the use of various clinical measures to classify periodontal disease, and the statistical effects of confounding variables.

Few prospective studies have investigated the relationship between periodontitis and overall and site-specific cancer incidences. In the present study, we aimed to examine the association between periodontal disease, including periodontitis, and the risks of total and site-specific cancers using the National Health Insurance Service-Health Examine Cohort data. Furthermore, we evaluated whether the risk of developing a secondary cancer would be different in patients with periodontal disease, including periodontitis.

Materials and methods

Study participants and design

We conducted a population-based, retrospective cohort study using data from the Korean National Health Insurance Cohort Database obtained between January 2003 and December 2015. All patients in the database older than 1 year of age were included in the cohort. Patients diagnosed with any form of cancer during the washout period (2003-2005) were excluded. Those without a cancer history who visited a dental clinic two or more than two times within one year and were diagnosed with periodontitis under those ICD-10 codes (K05.2, K05.3, K05.4, K05.5, and K05.6) between January 2003 and December 2005,

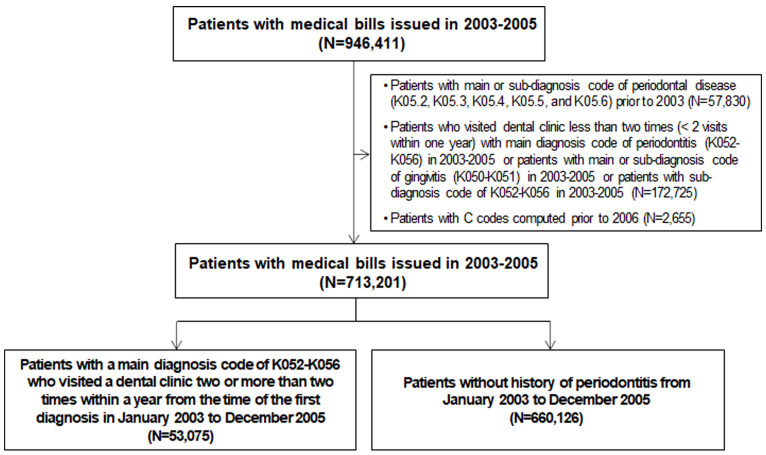


FIGURE 1
Flowchart of selection of the study population from the National Health Insurance Service-Health Screening Cohort database (n=713,201).

were included in the periodontitis group. As a control group, subjects have no history of periodontitis between 2003 and 2015. We excluded patients receiving a periodontal diagnosis prior to 2003 in the cohorts (Figure 1). A dentist performed an oral examination, and periodontitis was assessed using the Community Periodontal Index (CPI). Periodontal disease was defined as a CPI score≥3. The study population was followed up from the index date (January 2006) to the date of cancer, death, or the end of the study (December 2015). The study was approved by the Institutional Review Board (4-2019-0616).

Study outcomes and definitions

The main outcome of this study was the association between periodontitis and overall and site-specific cancer risks. As certain types of cancers, such as breast cancer and cancers of reproductive organs, were not differentiated in the Korean National Health Insurance Database, these were placed under the category of "others." These included breast, cervical, vulvar, vaginal, ovarian, and endometrial cancers in female patients and prostate, testicular, and penile cancers in male patients. The International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes were used to designate the patients diagnosed with periodontal disease (K05.2 through K05.6) and those diagnosed with cancer, defined as a new claim for primary diagnosis of cancer (C code and V193). Patients with newly diagnosed cancer were registered with a special certification, code V193, from September 2005 for expanded benefit coverage in Korea.

The cancer occurrence date was defined as the date of a newly developed primary cancer from the National Health Insurance System (NHIS) database. The presence of periodontal disease was identified when ICD-10 codes for acute periodontitis (K05.2), chronic periodontitis (K05.3), periodontitis (K05.4), other periodontal diseases (K05.5), or unspecified periodontal disease (K05.6), as previous studies (14, 15). Other periodontal diseases (K05.5) or unspecified periodontal diseases (K05.6) can include acute or chronic periodontitis (K05.2, K05.3, and K05.4).

To evaluate the development of secondary cancer, we evaluated whether the cancer survivors were diagnosed with another primary cancer different from the primary cancer type based on ICD-10 codes using C codes representing cancer diagnosis. Individuals who had not survived five years from their first primary cancer were excluded.

Confounding variables

Confounding factors included demographic factors (age, sex, income, and residential area), lifestyle behaviors (smoking history and body mass index [BMI]), and comorbidities, as defined by the Charlson Comorbidity Index (CCI) (16). A healthy examination questionnaire obtained smoking status in the patient cohort between 2003 and 2005. Smoking status was categorized into none, former smoker, and current smoker regardless of the amount of smoking. CCI was calculated based on the ICD-10 codes according to previous studies (17, 18). Briefly, CCI corresponds to the sum of the weights of the current comorbidities for each patient. Comorbidities with corresponding weights include myocardial infarction within the six months prior to surgery (1), congestive heart failure (1), peripheral vascular disease or rest pain (1), any history of

cerebrovascular accident (1), dementia (1), chronic obstructive pulmonary disease (1), connective tissue disease (1), peptic ulcer disease (1), diabetes mellitus (1), moderate to severe chronic kidney disease (2), hemiplegia (2), leukemia (2), malignant lymphoma (2), ascites or esophageal varices (3), disseminated cancer (6), and acquired immune deficiency syndrome (6). The point values were summed for a total number.

Statistical analysis

The Kolmogorov-Smirnov normality test was used to evaluate if variables are normally distributed. Continuous variables were expressed as median with interquartile range (IQR) and compared using the Mann–Whitney U test when the data did not follow the normal distribution. Categorical variables, such as sex, level of income, or residential area, were presented as numbers (%) and were compared using the chi-squared test or Fisher's exact test. Regarding categorical variables with missing values (such as BMI and smoking status), we treated missing values as a valid missing category. The Kaplan-Meier method was used to estimate the cumulative risk of cancer. Multivariable Cox proportional hazards regression analysis was used to estimate the adjusted hazard ratio (aHR) and 95% confidence intervals (CIs). The model was adjusted for potential confounders such as age, sex, comorbidities, BMI, and smoking history. Since the 77% of patients had missing value for BMI and smoking history, missing value were grouped into a "missing" category. Analyses were performed using the SAS Enterprise Guide version 7.1 (SAS Institute, Inc., Cary, NC, USA) and two-sided P value < 0.05 was considered statistically significant.

Results

Baseline characteristics of the study population

Among the 713,201 participants, 53,075 (7.4%) had periodontitis, and 660,126 (92.6%) were included as healthy controls between January 2003 to December 2005 (Figure 1). In the subjects with periodontitis (n=53,075), the median age was 49 years; 49.6% were males; 6.4% had BMI more than 25; 10.8% were current smokers (Table 1). Current smokers were nearly twice in the periodontitis group compared to the control (10.8% vs. 5.5%, respectively). Regarding socioeconomic status, the proportions of people living below or equal to 50% of the median income and people living at 51-80% of the median income were 3.7% and 2.2% higher in the control group than in the periodontitis group, respectively. In contrast, the proportion of people living above or equal to 80% of the median income was 6.0% higher in the periodontitis group (25.7%) than in the control group (31.7%). Both groups had nearly half of the participants living in the capital or metropolitan cities (control: 46.1% and periodontitis: 48.9%), respectively. Multivariate Cox regression analyses of potential confounding factors for cancer development in this study cohort (n=713,201) was shown in Supplementary Table 1. Female sex (aHR, 0.761; 95% CI, 0.740-0.783; P < 0.0001), current smoker (aHR, 1.187; 95% CI, 1.127-1.250; P = 0.0013) or former smoker (aHR, 1.127; 95% CI, 1.048-1.211; P < 0.0001), and subjects living in capital city (aHR, 1.037; 95% CI, 1.002-1.073; P = 0.0378) were potential confounders for cancer development (Supplementary Table 1).

Table 1 Baseline characteristics of the study population (n = 713,201).

Characteristics	Control (n = 660,126)	Periodontitis (n = 53,075)	*P* value
Age, years	31 (15-46)	49 (39-60)	<0.0001
CCI score	0 (0-1)	1 (0-2)	<0.0001
Sex (%)			<0.0001
Male	325,796 (49.6)	27,310 (51.5)	
Female	334,330 (50.4)	25,765 (48.5)	
BMI, kg/m^2			<0.0001
< 20	89,951 (13.6)	2,172 (4.1)	
20 ~ 25	80,501 (12.2)	13,088 (24.7)	
≥ 25	41,969 (6.4)	7,534 (14.2)	
Missing	519,496 (78.7)	30,281 (57.1)	
Smoking			<0.0001
None	89,951 (13.6)	14,300 (26.9)	
Former	12,049 (1.8)	2,444 (4.6)	
Current	36,498 (5.5)	5,732 (10.8)	
Missing	521,628 (79.0)	30,599 (57.7)	
Level of income			<0.0001
≤ 50% (lower income)	256,969 (39.5)	18,910 (35.8)	
51~80%	226,055 (34.8)	17,244 (32.6)	
≥ 81% (higher income)	166,857 (25.7)	16,748 (31.7)	
Residential area			<0.0001
Capital city	133,337 (20.2)	11,610 (21.9)	
Urban area	171,117 (25.9)	14,350 (27.0)	
Rural area	255,672 (53.9)	27,115 (51.1)	

Incidences of overall cancer incidence in patients with periodontitis

Next, we evaluated whether periodontitis is associated with increased overall cancer incidence over time. First, we calculated cumulative cancer incidence over time in the subjects with periodontitis, compared to the control group for ten years. One minus the Kaplan-Meier estimate provided an estimate of the cumulative cancer incidence over time in Figure 2.

10.3389/fonc.2022.901098

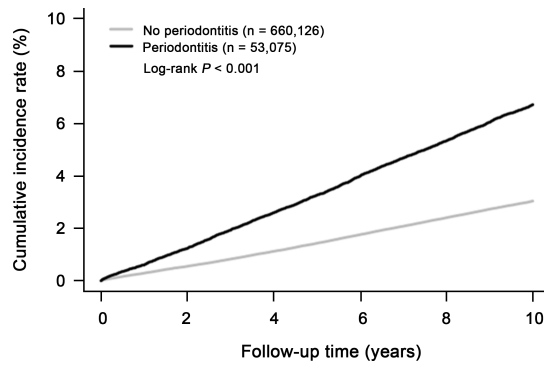


FIGURE 2
Cumulative incidence of overall cancer in subjects with periodontitis (n=53,075) and without periodontitis (n=660,126).

Interestingly, the cumulative incidence of cancer in the periodontitis group was 2.2 times higher than that in the control group over ten years (Figure 2; log-rank test, $P < 0.001$). The increased incidence rate showed a linear trend over time in periodontitis group (Figure 2).

Risks of organ-specific cancers in patients with periodontitis

To compare the risk of cancer development in the periodontitis and control groups, multivariable Cox proportional hazards regression analysis was performed (Table 2). The model was adjusted for potential confounders, such as age, sex, comorbidities defined by the Charlson Comorbidity Index, BMI, smoking history, the level of income, and residential areas (Table 2). The overall cancer risk was significantly higher in the periodontitis group than in the control group (aHR 1.129; 95% CI, 1.089-1.171; $P < 0.0001$). Further, diagnosis with periodontitis was associated with increased risks of stomach cancer (aHR, 1.136; 95% CI, 1.042-1.239; $P = 0.0037$), colon cancer (aHR, 1.129; 95% CI, 1.029-1.239; $P = 0.0105$), lung cancer (aHR, 1.127; 95% CI, 1.008-1.260; $P = 0.0353$), bladder cancer (aHR, 1.307; 95% CI, 1.071-1.595; $P = 0.0085$), thyroid cancer (aHR, 1.191; 95% CI, 1.085-1.308; $P = 0.0002$), and leukemia (aHR, 1.394; 95% CI, 1.039-1.872; $P = 0.0270$). Cumulative cancer incidence over time was shown in primary cancer of the stomach, colon, lung, bladder, thyroid, and leukemia, respectively (Supplementary Figure 1), suggesting that patients with periodontitis increased overall cancer incidence compared to the control group.

Table 2 Adjusted risks of total and organ-specific cancers in patients with periodontitis after correcting confounding factors.

Type of cancer (ICD code)	Number of events			Adjusted HR(95% CI)	*P* value
	Total	Control	Periodontitis		
Lip, oral cavity, pharynx (C00-C14)	372	324	48	0.914 (0.670-1.249)	0.5730
Esophagus (C15)	314	262	52	1.098 (0.812-1.486)	0.5441
Stomach (C16)	3,920	3,278	642	1.136 (1.042-1.239)	0.0037**
Colon (C18-C20)	3,416	2,868	548	1.129 (1.029-1.239)	0.0105*
Liver (C22)	2,002	1,673	329	1.108 (0.982-1.250)	0.0962
Gallbladder, biliary tract (C23-C24)	565	469	96	1.171 (0.937-1.464)	0.1655
Pancreas (C25)	674	572	102	1.012 (0.815-1.256)	0.9152
Larynx (C32)	157	130	27	1.024 (0.673-1.559)	0.9102
Lung (C33-C34)	2,259	1,874	385	1.127 (1.008-1.260)	0.0353*
Kidney (C64)	490	404	86	1.249 (0.984-1.584)	0.0676
Bladder (C67)	602	475	127	1.307 (1.071-1.595)	0.0085**
Brain, CNS (C70-C72)	402	353	49	1.152 (0.848-1.566)	0.3662
Thyroid (C73)	4,399	3,871	528	1.191 (1.085-1.308)	0.0002**
Hodgkin lymphoma (C81)	37	34	3	0.956 (0.283-3.229)	0.9425
Non-Hodgkin lymphoma (C82-C86, C96)	497	430	67	1.045 (0.803-1.359)	0.7439
Multiple myeloma (C90)	137	114	23	1.189 (0.754-1.874)	0.4567
Leukemia (C91-C95)	399	344	55	1.394 (1.039-1.872)	0.0270*
Other malignant neoplasms (remainder of C00-C96)	5,322	4,565	757	1.102 (1.019-1.192)	0.0148*
Total	**23,281**	**19,778**	**3,503**	**1.129 (1.089-1.171)**	**<0.0001****

*P < 0.05; **P < 0.01.

Next, we evaluated aged-matched adjusted hazard ratio in subjects with ≥40 years old from the study population (Supplementary Table 2). An increased risk of total cancer (aHR 1.080; 95% CI, 1.040-1.122; $P < 0.0001$) was observed for the periodontitis group compared with the control group (Supplementary Table 2). By cancer site, significant associations for periodontitis groups were observed for bladder cancer (aHR, 1.307; 95% CI, 1.069-1.598; $P = 0.0091$), thyroid cancer (aHR, 1.123; 95% CI, 1.008-1.251; $P = 0.0349$), and leukemia (aHR, 1.407; 95% CI, 1.016-1.947; $P = 0.0396$), respectively.

Risks of secondary malignancy in patients with periodontitis

Finally, we evaluated whether periodontitis affects the occurrence of secondary cancer in cancer survivors. The incidence rates of secondary cancer were 0.89% (176 out of 19,778) and 1.03% (36 out of 3,503) in the control and periodontitis groups, respectively (Supplementary Table 3). Although the incidence rates of secondary malignancy are 0.14% higher in the periodontitis group, the difference was not statistically significant ($P = 0.428$). Collectively, there was no significant association between the development of secondary malignancy and the history of periodontitis in cancer survivors who were alive five years after they were diagnosed with the primary malignancy.

Discussion

In this study, we showed that patients with periodontal disease, including periodontitis, have an increased overall cancer incidence and an organ-specific cancer incidence compared to control individuals. Periodontitis was associated with increased risks of gastrointestinal cancers (such as stomach cancer, colon cancer), lung cancer, bladder cancer, thyroid cancer and leukemia. Even after controlling for confounding factors, such as sex, income, smoking history, BMI, and comorbidities, periodontitis was found to be a modest but obvious risk factor for cancer.

Previous observational reports and meta-analyses suggested that periodontal disease was associated with increased risks of several cancer types, including head and neck, lung, pancreatic, colorectal, kidney, and hematologic cancers (8, 12, 13). Our results are consistent and comparable with published data. For instance, previous studies showed periodontitis was positively correlated with an increased risk of lung cancer (HR, 2.33; 95% CI, 1.51 to 3.60) and colon cancer among never smokers (HR, 2.12; 95% CI, 1.00 to 4.47) (13). Our finding is consistent with the previous reports on colon cancer (adjusted HR, 1.129; P=0.011) and lung cancer (adjusted HR, 1.127; P=0.035). Although we did not observe a positive correlation between periodontitis and increased risk of head and neck cancers (adjusted HR, 0.914; P=0.573), the number of head and neck cases was small to draw a conclusion in this study. Future larger association study is warranted. Interestingly, our study showed a strong correlation with the development of inflammation-associated cancers, such as bladder cancer (adjusted HR, 1.307; P=0.008) and thyroid cancer (adjusted HR, 1.191; P<0.001) after correction of smoking history.

The potential relationship between periodontitis and cancer can be explained by the properties of local and systemic inflammation associated with bacteremia and increased myelopoietic activity (7). Periodontitis causes increased systemic inflammation because of increased bacterial infection, hematogenous dissemination of oral pathogenic bacteria, increased inflammatory mediators (such as interleukin [IL]-1, IL-6, and C-reactive protein and fibrinogen), and increased neutrophil number in the bloodstream (19–21). Chronic systemic inflammation causes cellular stress, including DNA damage through reactive oxygen species stress and reactive nitrogen species (22). Further, inflammatory mediators like NF-κB and STAT3 increase genetic instability. Additionally, repeated tissue damage and repair trigger chromosomal translocation. These mechanisms induce DNA damage and mutation. Inflammation and genetic instability have a sufficiently significant causal association for inflammation to be included as a hallmark of cancer (23, 24).

Another explanation for the relationship between periodontitis and cancer is oral bacteria (21). Frequent transient bacteremia of oral pathogens leading to sustained systemic inflammatory responses appears to be key to the mechanism of carcinogenesis in patients with chronic periodontitis (25). Periodontitis can also cause oral and gut dysbacteriosis. *Porphyronas gingivalis* (*P. gingivalis*) infection can alter the gut microbiota, enhance blood endotoxin levels, cause systemic inflammation, interfere with the host metabolism, and promote immune system evasion (21, 26–28). *P. gingivalis* has been shown to evade innate immune detection and enhance chronic inflammation of vascular structures through TLR-4 (21, 28). Patients with oral diseases such as gingivitis and periodontitis may be more likely to develop intestinal dysbiosis (29, 30). *P. gingivalis* is also found in patients with colorectal cancer, and human colon cells infected by *P. gingivalis* can develop into colorectal cancer (30). Moreover, oral bacteria such as *Gemella, Peptostreptococcus*, and *Fusobacterium* are strongly correlated with colorectal cancer (31). Many studies have proposed the association between colon cancer development by *Fusobacterium nucleatum (F. nucleatum)* (32–34). *F. nucleatum* binds to tumor cells *via* the virulent adhesin protein Fap2 and activates Wnt signaling pathway, leading to epithelial-mesenchymal transition (35). Furthermore, direct interaction between the FadA adhesin proteins and E-

cadherin on the surface of colonic epithelial cells increased E-cadherin/β-catenin-modulated transcription factors, leading to DNA damage, epithelial cell proliferation, and acquisition of cancer stemness.

This study has several limitations. First, previous studies suggest a positive correlation between periodontitis and breast and genitourinary cancers (36–39). Unfortunately, the National Health Insurance Database does not classify reproductive organ carcinomas such as breast and genitourinary cancers but categorizes them as "other" cancers. Although we could not evaluate the association between periodontitis and breast cancer and genitourinary incidence, we included them in the overall cancer incidence calculation. Second, because the NHIS database does not classify the severity of periodontitis (such as the number of teeth affected), we could not evaluate the association of periodontitis severity or treatment history with cancer risk (13). Third, the lack of circulating markers or bacterial levels in the NHIS database does not allow further analysis to identify the role of specific oral microbiota in cancer development. Forth, the heterogeneity nature of the ICD-10 diagnosis code for periodontitis and cancer diagnosis may lead to selection bias and underestimate the association in this study. Of note, this study based on ICD-10 codes does not reflect recently updated periodontitis classification criteria (3). A prospective cohort using the updated periodontal disease classification criteria study will increase the accuracy of the analysis. Fifth, a large portion of missing information in smoking history and body mass index is a potential bias for adjusting for confounding factors. We collected smoking history and body mass index [BMI] based on the patient-reported healthy examination questionnaire in the patient cohort. However, many subjects in the patient cohort had missing information on the healthy examination questionnaire. In addition, although we confirmed the primary tumor site, the pathological findings were not available from the National Health Insurance Database. Therefore, it was not possible to assess the association between the pathological characteristics of periodontitis and cancer.

In conclusion, periodontal disease, including periodontitis, was associated with increased risk of cancer, which persisted after controlling for confounding factors. Further prospective research is warranted to establish a causal relationship.

Data availability statement

Data access to the NHIS database enables via National Health Insurance Sharing Service (https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do). The application form, a research proposal, and an approval document from the appropriate IRB be submitted to and reviewed by the NHIS inquiry committee for research support.

Ethics statement

The studies involving human participants were reviewed and approved by Severance Hospital Human Research Protection Center (Institutional Review Board No. 4-2019-0616). Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

HK, IJ, EK, and ML contributed to the conception, design, and the interpretation of the study. EHK collected and analyzed the data. HK, SN, EK, and CP reviewed, interpreted, and drafted the manuscript. YK, JA, SS, YP, HJ, and B-IK reviewed and edited the manuscript. All the authors read and approved the final manuscript for submission and take responsibility for the data presented in this manuscript.

Funding

This study was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C4001879 to H.S.K.), the Bio & Medical Technology Development Program of the National Research Foundation (NRF) & funded by the Korean government (MSIT) (No. 2022M3A9F3016364 to H.S.K.), the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. HI21C0974 to H.S.K.), and a faculty research grant of Yonsei University College of Medicine (6-2020-0218 to H.S.K.).

Acknowledgments

The content of this manuscript has been presented in part at the European Society for Medical Oncology (ESMO) in 2019.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.901098/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Cumulative incidence in primary cancer of the stomach, colon, lung, bladder, thyroid, and leukemia in subjects with periodontitis (n=53,075) and without periodontitis (n=660,126).

References

1. Nazir MA. Prevalence of periodontal disease, its association with systemic diseases and prevention. *Int J Health Sci (Qassim)* (2017) 11(2):72–80.

2. Sanz M, D'Aiuto F, Deanfield J, Fernandez-Avilés F. European Workshop in periodontal health and cardiovascular disease–scientific evidence on the association between periodontal and cardiovascular diseases: a review of the literature. *Eur Heart J Suppl* (2010) 12(suppl_B):B3–B12. doi: 10.1093/eurheartj/suq003

3. Caton JG, Armitage G, Berglundh T, Chapple ILC, Jepsen S, Kornman KS, et al. A new classification scheme for periodontal and peri-implant diseases and conditions - introduction and key changes from the 1999 classification. *J Clin Periodontol* (2018) 45 Suppl 20:S1–8. doi: 10.1111/jcpe.12935

4. Peres MA, Macpherson LMD, Weyant RJ, Daly B, Venturelli R, Mathur MR, et al. Oral diseases: a global public health challenge. *Lancet* (2019) 394(10194):249–60. doi: 10.1016/S0140-6736(19)31146-8

5. Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJ, Marcenes W. Global burden of severe periodontitis in 1990-2010: a systematic review and meta-regression. *J Dent Res* (2014) 93(11):1045–53. doi: 10.1177/0022034514552491

6. Collaborators GBDOD, Bernabe E, Marcenes W, Hernandez CR, Bailey J, Abreu LG, et al. Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: A systematic analysis for the global burden of disease 2017 study. *J Dent Res* (2020) 99(4):362–73. doi: 10.1177/0022034520908533

7. Hajishengallis G, Chavakis T. Local and systemic mechanisms linking periodontal disease and inflammatory comorbidities. *Nat Rev Immunol* (2021) 21(7):426–40. doi: 10.1038/s41577-020-00488-6

8. Michaud DS, Liu Y, Meyer M, Giovannucci E, Joshipura K. Periodontal disease, tooth loss, and cancer risk in male health professionals: a prospective cohort study. *Lancet Oncol* (2008) 9(6):550–8. doi: 10.1016/s1470-2045(08)70106-2

9. Arora M, Weuve J, Fall K, Pedersen NL, Mucci LA. An exploration of shared genetic risk factors between periodontal disease and cancers: a prospective co-twin study. *Am J Epidemiol* (2010) 171(2):253–9. doi: 10.1093/aje/kwp340

10. Nwizu NN, Marshall JR, Moysich K, Genco RJ, Hovey KM, Mai X, et al. Periodontal disease and incident cancer risk among postmenopausal women: Results from the women's health initiative observational cohort. *Cancer Epidemiol Biomarkers Prev* (2017) 26(8):1255–65. doi: 10.1158/1055-9965.Epi-17-0212

11. Michaud DS, Kelsey KT, Papathanasiou E, Genco CA, Giovannucci E. Periodontal disease and risk of all cancers among male never smokers: an updated analysis of the health professionals follow-up study. *Ann Oncol* (2016) 27(5):941–7. doi: 10.1093/annonc/mdw028

12. Michaud DS, Fu Z, Shi J, Chung M. Periodontal disease, tooth loss, and cancer risk. *Epidemiol Rev* (2017) 39(1):49–58. doi: 10.1093/epirev/mxx006

13. Michaud DS, Lu J, Peacock-Villada AY, Barber JR, Joshu CE, Prizment AE, et al. Periodontal disease assessed using clinical dental measurements and cancer risk in the ARIC study. *J Natl Cancer Inst* (2018) 110(8):843–54. doi: 10.1093/jnci/djx278

14. Park SY, Kim SH, Kang SH, Yoon CH, Lee HJ, Yun PY, et al. Improved oral hygiene care attenuates the cardiovascular risk of oral health disease: a population-based study from Korea. *Eur Heart J* (2019) 40(14):1138–45. doi: 10.1093/eurheartj/ehy836

15. Lee JH, Choi JK, Jeong SN, Choi SH. Charlson comorbidity index as a predictor of periodontal disease in elderly participants. *J Periodontal Implant Sci* (2018) 48(2):92–102. doi: 10.5051/jpis.2018.48.2.92

16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* (1987) 40(5):373–83. doi: 10.1016/0021-9681(87)90171-8

17. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD-10 diagnostic coding used to assess charlson comorbidity index conditions in the population-based Danish national registry of patients. *BMC Med Res Methodol* (2011) 11:83. doi: 10.1186/1471-2288-11-83

18. Bannay A, Chaignot C, Blotiere PO, Basson M, Weill A, Ricordeau P, et al. The best use of the charlson comorbidity index with electronic health care database to predict mortality. *Med Care* (2016) 54(2):188–94. doi: 10.1097/MLR.0000000000000471

19. Hajishengallis G. The inflammophilic character of the periodontitis-associated microbiota. *Mol Oral Microbiol* (2014) 29(6):248–57. doi: 10.1111/omi.12065

20. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. *Nat Rev Immunol* (2015) 15(1):30–44. doi: 10.1038/nri3785

21. Martínez-García M, Hernández-Lemus E. Periodontal inflammation and systemic diseases: An overview. *Front Physiol* (2021) 12:709438. doi: 10.3389/fphys.2021.709438

22. Lamont RJ, Koo H, Hajishengallis G. The oral microbiota: dynamic communities and host interactions. *Nat Rev Microbiol* (2018) 16(12):745–59. doi: 10.1038/s41579-018-0089-x

23. Trinchieri G. Cancer and inflammation: an old intuition with rapidly evolving new concepts. *Annu Rev Immunol* (2012) 30:677–706. doi: 10.1146/annurev-immunol-020711-075008

24. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis* (2009) 30(7):1073–81. doi: 10.1093/carcin/bgp127

25. Meurman JH. Oral microbiota and cancer. *J Oral Microbiol* (2010) 2 (1):5195. doi: 10.3402/jom.v2i0.5195

26. Bui FQ, Almeida-da-Silva CLC, Huynh B, Trinh A, Liu J, Woodward J, et al. Association between periodontal pathogens and systemic disease. *BioMed J* (2019) 42(1):27–35. doi: 10.1016/j.bj.2018.12.001

27. Binder Gallimidi A, Fischman S, Revach B, Bulvik R, Maliutina A, Rubinstein AM, et al. Periodontal pathogens porphyromonas gingivalis and fusobacterium nucleatum promote tumor progression in an oral-specific chemical carcinogenesis model. *Oncotarget* (2015) 6(26):22613–23. doi: 10.18632/oncotarget.4209

28. Xu W, Zhou W, Wang H, Liang S. Roles of porphyromonas gingivalis and its virulence factors in periodontitis. *Adv Protein Chem Struct Biol* (2020) 120:45–84. doi: 10.1016/bs.apcsb.2019.12.001

29. Lourenvarsigmao TGB, Spencer SJ, Alm EJ, Colombo APV. Defining the gut microbiota in individuals with periodontal diseases: an exploratory study. *J Oral Microbiol* (2018) 10(1):1487741. doi: 10.1080/20002297.2018.1487741

30. Mu W, Jia Y, Chen X, Li H, Wang Z, Cheng B. Intracellular porphyromonas gingivalis promotes the proliferation of colorectal cancer cells *via* the MAPK/ERK signaling pathway. *Front Cell Infect Microbiol* (2020) 10:584798. doi: 10.3389/fcimb.2020.584798

31. Koliarakis I, Messaritakis I, Nikolouzakis TK, Hamilos G, Souglakos J, Tsiaoussis J. Oral bacteria and intestinal dysbiosis in colorectal cancer. *Int J Mol Sci* (2019) 20(17):4146. doi: 10.3390/ijms20174146

32. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 protein of fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* (2015) 42(2):344–55. doi: 10.1016/j.immuni.2015.01.010

33. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates

the tumor-immune microenvironment. *Cell Host Microbe* (2013) 14(2):207–15. doi: 10.1016/j.chom.2013.07.007

34. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/beta-catenin signaling *via* its FadA adhesin. *Cell Host Microbe* (2013) 14(2):195–206. doi: 10.1016/j.chom.2013.07.012

35. Bi D, Zhu Y, Gao Y, Li H, Zhu X, Wei R, et al. Profiling fusobacterium infection at high taxonomic resolution reveals lineage-specific correlations in colorectal cancer. *Nat Commun* (2022) 13(1):3336. doi: 10.1038/s41467-022-30957-6

36. Sfreddo CS, Maier J, De David SC, Susin C, Moreira CHC. Periodontitis and breast cancer: A case-control study. *Community Dent Oral Epidemiol* (2017) 45 (6):545–51. doi: 10.1111/cdoe.12318

37. Guo Z, Gu C, Li S, Gan S, Li Y, Xiang S, et al. Periodontal disease and the risk of prostate cancer: a meta-analysis of cohort studies. *Int Braz J Urol* (2021) 47 (6):1120–30. doi: 10.1590/S1677-5538.IBJU.2020.0333

38. Yuan S, Fang C, Leng WD, Wu L, Li BH, Wang XH, et al. Oral microbiota in the oral-genitourinary axis: identifying periodontitis as a potential risk of genitourinary cancers. *Mil Med Res* (2021) 8(1):54. doi: 10.1186/s40779-021-00344-1

39. Wei Y, Zhong Y, Wang Y, Huang R. Association between periodontal disease and prostate cancer: a systematic review and meta-analysis. *Med Oral Patol Oral Cir Bucal* (2021) 26(4):e459–e65. doi: 10.4317/medoral.24308

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership