EClinicalMedicine 33 (2021) 100756

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research Paper

Estimating cancer survival and prevalence with the Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS): An empirical study in China

Hongrui Tian[a], Yanjun Hu[b], Qingxiang Li[c], Liang Lei[b], Zhen Liu[a], Mengfei Liu[a], Chuanhai Guo[a], Fangfang Liu[a], Ying Liu[a], Yaqi Pan[a], Isabel dos-Santos-Silva[d,*], Zhonghu He[a,*], Yang Ke[a,*]

[a] *Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Genetics, Peking University Cancer Hospital & Institute, Beijing, China*
[b] *Healthcare Security Administration of Hua County, Henan Province, China*
[c] *Center for Disease Control and Prevention of Hua County, Henan Province, China*
[d] *Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom*

ARTICLE INFO

Article History:
Received 24 November 2020
Revised 12 January 2021
Accepted 28 January 2021
Available online xxx

Keywords:
Cancer surveillance
Survival
Prevalence
Health-related big data
China

ABSTRACT

Background: We aimed to establish a new approach for surveillance of cancer prevalence and survival in China, based on the Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS).
Methods: We constructed a standard procedure for data collection, cleaning, processing, linkage, verification, analysis, and estimation of cancer prevalence and survival (including both actual observations and model estimates) by conjoint use of medical insurance claims data and all-cause death surveillance data. As a proof-of-principle study, we evaluated the performance of this surveillance approach by estimating the latest prevalence and survival for upper gastrointestinal cancers in Hua County, a high-risk region for oesophageal cancer in China.
Findings: In Hua County, the age-standardised relative 5-year survival was 39·2% (male: 36·8%; female: 43·6%) for oesophageal cancer and 33·3% (male: 29·6%; female: 43·4%) for stomach cancer. For oesophageal cancer, better survival was observed in patients of 45–64 years compared with national average estimates, and women of <75 years had better survival than men. The 5-year prevalence rate in Hua County was 99·8/100,000 (male: 105·9/100,000; female: 93·3/100,000) for oesophageal cancer and 41·5/100,000 (male: 57·4/100,000; female: 24·5/100,000) for stomach cancer. For both of these cancers, the prevalence burden peaked at 65–79 years. The model estimates for survival and prevalence were close to the observations in real investigation, with a relative difference of less than 4·5%.
Interpretation: This novel approach allows accurate estimation of cancer prevalence and survival with a short delay, which has great potential for regular use in general Chinese populations, especially those not covered by cancer registries.
Funding: The National Key R&D Program of China (2016YFC0901404), the National Science & Technology Fundamental Resources Investigation Program of China (2019FY101102), the National Natural Science Foundation of China (82073626), the Taikang Yicai Public Health and Epidemic Control Fund (TKYC-GW-2020), the Beijing-Tianjin-Hebei Basic Research Cooperation Project (J200016), and the Digestive Medical Coordinated Development Center of Beijing Hospitals Authority (XXZ0204).

1. Introduction

In addition to incidence and mortality, prevalence and survival figures are also important components in cancer surveillance. Prevalence measures the burden of cancer in a given population at a particular time, thus providing crucial evidence for medical resource allocation [1]. In general, estimation of cancer prevalence includes individuals diagnosed with cancer within the last five years who are still alive at the time of study. Those who were diagnosed at an earlier time are usually assumed to be “cured” as their death rates tend to be similar to those in the general population [2]. Population-level survival of cancer patients, which is commonly estimated at five years,

* Corresponding author.
E-mail addresses: zhonghuhe@foxmail.com (Z. He), keyang@bjmu.edu.cn (Y. Ke).

https://doi.org/10.1016/j.eclinm.2021.100756

Research in context

Evidence before this study Population-level cancer surveillance of prevalence and survival is crucial for establishing policies regarding medical resource allocation and strategies for prevention and control. Estimating cancer prevalence and survival currently relies on population-based cancer registries (PBCR). However, in China the population coverage of PBCRs is still low due to the vast territory of this country, its huge population, and limited resources (we searched PubMed and Google Scholar for articles published before Jun 1, 2020, without language restriction, using the terms "relative 5-year survival", "5-year relative survival", "prevalence", "cancer surveillance", etc.). We previously established a cancer incidence surveillance approach based on claims data from medical insurance systems, but whether this methodology can be extended to cancer prevalence and survival surveillance is as yet undetermined.

Added value of this study For the first time, we conducted an empirical study in a high-risk region for oesophageal cancer in China which is not covered by PBCRs, and evaluated the performance of cancer prevalence and survival surveillance based on medical insurance claims data together with death surveillance data. Results of this study demonstrated that these two kinds of health-related big data can be used jointly to estimate cancer prevalence and survival in a regular and timely manner in this area.

Implications of all the available evidence This new approach to surveillance of cancer prevalence and survival based on health-related big data has potential for adoption in broader areas not covered by PBCRs, and holds promise for reform of cancer surveillance methodology in China.

reflects the average prognosis of the cancer and may serve as a measure for evaluation of the performance of a health care system and preventive programmes [3]. In most cases, accurate and up-to-date estimation of prevalence and survival relies upon the availability of high-quality population-based cancer registries (PBCR) and death surveillance systems [3,4].

In China, cancer registration was launched in the 1950s, and a total of 574 cancer surveillance points had been established by 2019, which were located primarily in economically developed areas or in regions of high-risk for specific cancers [5–7]. Nevertheless, data for only 36 of these surveillance points covering 5% of the total national population, were considered to be of sufficient high quality for inclusion in *Cancer Incidence in Five Continents* [8,9]. In addition, the annual reports of the National Cancer Center in China provide only cancer incidence and mortality. For estimation of cancer deaths, data are obtained mainly through linkage with the death surveillance system, which was established in the 1950s and is currently operated by the Chinese Center for Disease Control and Prevention. This system contains 605 surveillance sites and covers 24·3% of the total population of China [10,11]. Thus far there is no surveillance system that regularly estimates and reports cancer survival and prevalence in China.

To date, only two studies reporting nationwide estimates of cancer survival have been published in China. Both included data from a network of 17 cancer registries located predominately in eastern regions of China, and covered approximately 2% of the national population [12,13]. The sole set of cancer prevalence statistics in China thus far was reported for the year 2011, and this has been estimated based on cancer incidence data from 177 cancer registries and survival data from 17 cancer registries [14].

In the mode of PBCRs, current surveillance of cancer prevalence and survival in China has encountered insurmountable barriers. First, population coverage cannot be increased rapidly as it cannot be carried out in areas lacking PBCRs. Second, it is difficult to provide cancer prevalence and survival estimates in a regular and timely manner, since individual-level ascertainment of cancer cases is overwhelmingly resource-intensive, and a 3-year delay is still incurred in the current mode of cancer incidence surveillance.

In China, basic medical insurance systems are administered and operated by the government, and cover over 95% of the national population. In a previous study, we established the Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS) and demonstrated that this approach is capable of complete and accurate identification of incident cancer cases, and allows reporting cancer incidence in a timely manner in two regions (six million residents altogether) which are not covered by PBCRs in China [15].

In this study, we aimed to further this work by developing a standard procedure for cancer prevalence and survival surveillance by bringing together data from MIS-CASS and the death surveillance system in Hua County, a high-risk area for oesophageal cancer in northern China. As "proof of principle" we assessed the performance of this novel approach by estimating the prevalence of, and survival from upper gastrointestinal (GI) cancers (oesophageal cancer, non-cardia gastric cancer, and cardia gastric cancer) in geographical areas covered by two on-going large-scale population-based cohorts [16,17].

2. Methods

2.1. Study site

Hua County is a predominantly rural area in northern Henan Province in China, which has 1·07 million permanent residents and had a per capita Gross Domestic Product of 24,585 Yuan (US $3715) in 2018 [18–20]. The New Rural Co-operative Medical Care Scheme (NCMS) is the sole medical insurance system for rural residents, and this system has had coverage of 99% since 2010 [15]. The incidence rate of oesophageal cancer ranked second among cancer types in Hua County in 2018, and this rate was 2·1 times as high as the national level [15,21]. Over the past decade, we have carried out a series of epidemiologic and etiologic studies of oesophageal cancer in Hua County [16,17,22,23]. Due to the lack of PBCRs, accurate and updated prevalence and survival data for upper GI cancers in this region which is a typical high-risk region, have thus far not been reported. However, such reporting is crucial for establishing medical resource allocation policies and prevention and control strategies.

2.2. Data collection and cleaning

2.2.1. Incidence data for upper GI cancers

In this study, we extracted NCMS claims data for 2012–2019 from the Healthcare Security Administration of Hua County in order to definitively identify incident cases of upper GI cancer diagnosed between 2014 and 2018 in the target villages of the Anyang Esophageal Cancer Cohort Study (AECCS) and Endoscopic Screening for Esophageal Cancer in China (ESECC) cohorts. These villages included 834·9 thousand rural residents and represented 66% of the total population of Hua County [16,17]. Ascertainment of incident cases based on the MIS-CASS has been described in detail in our previous study [15]. In brief, a 2-year time window (2012 and 2013) was adopted to washout the prevalent cases (diagnosed before January 1, 2014) as the incidence estimates were stable from the third year. Data from January 1 to June 30 of 2019 were also included which allowed a reimbursement delay of six months (the proportion of delays over six months was lower than 2·5‰) [15]. In this study, non-cardia gastric cancer and cardia gastric cancer were combined as stomach cancer.

2.2.2. Vital status data for upper GI cancer patients

The vital status of all incident upper GI cancer patients was determined by passive record linkage with death surveillance data and active follow-up.

We used all-cause death surveillance data between 2014 and 2019 from the Center for Disease Control and Prevention of Hua County to identify deaths that might have occurred up to the end of 2019 among all the incident upper GI cancer patients in the target villages of AECCS and ESECC cohorts newly-diagnosed in 2014–2018. The record linkage was performed using the unique personal identity number as the index variable. Name, gender, birthdate, township, and the village of origin were also used to assess the accuracy of the linkage.

In addition, we also conducted active follow-up for all the identified cancer patients in the target villages in December of 2019 and January of 2020. We first interviewed the village doctors in the target villages and public health directors in township health centers regarding all-cause death events among the cases of interest. We then merged the death list from active interviews with the death list from the death surveillance data. If no death was recorded in either list (n = 906; 40·2% of all cancer patients), or if it was recorded in one list (n = 328; 14·6% of all cancer patients), or recorded in both lists but with different dates (n = 543; 24·1% of all cancer patients), a home visit or telephone contact with the next of kin was made to confirm vital status and if the individual was dead, the date of death was checked. Finally, the "Verified vital status data set" was determined for the subjects of the present study. A total of 18 well-trained investigators in our team participated in the entire process of active follow-up (1·5 months), and they were required to regularly provide audio recordings of interviews for quality control. Informed consent from all the interviewees were obtained during the active follow-up.

2.2.3. Data cleaning and processing

The procedure for cleaning and processing of the cancer incidence and death data is depicted in Fig. 1. According to the MIS-CASS, a total of 2534 incident oesophageal cancers and 1023 incident stomach cancers occurred in Hua County from January 1, 2014 to December 31, 2018. Of these cases, 1623 oesophageal cancer cases and 649 stomach cancer cases from target villages were identified and included in both passive and active follow-up. 1605 (98·9%) of the oesophageal cancer cases and 647 (99·7%) of the stomach cancer cases had definite vital status information, and were included in the "Verified vital status data set". Survival surveillance and prevalence surveillance were then based on this data set.

2.3. Statistical analysis

2.3.1. Calculation of cancer survival

We chose median survival time, observed survival, and relative survival as the main survival indicators. Median survival time was defined as the shortest survival time for which the Kaplan-Meier survival function was no higher than 50%. Observed survival was calculated using the actuarial method. Expected survival was calculated according to the Ederer II method using life tables stratified by gender, age, and calendar year (Each investigated cancer case was matched 1:1 with a cancer-free individual from the general population in Hua County based on gender, age, and year of diagnosis) [24]. We divided observed survival by expected survival to obtain relative

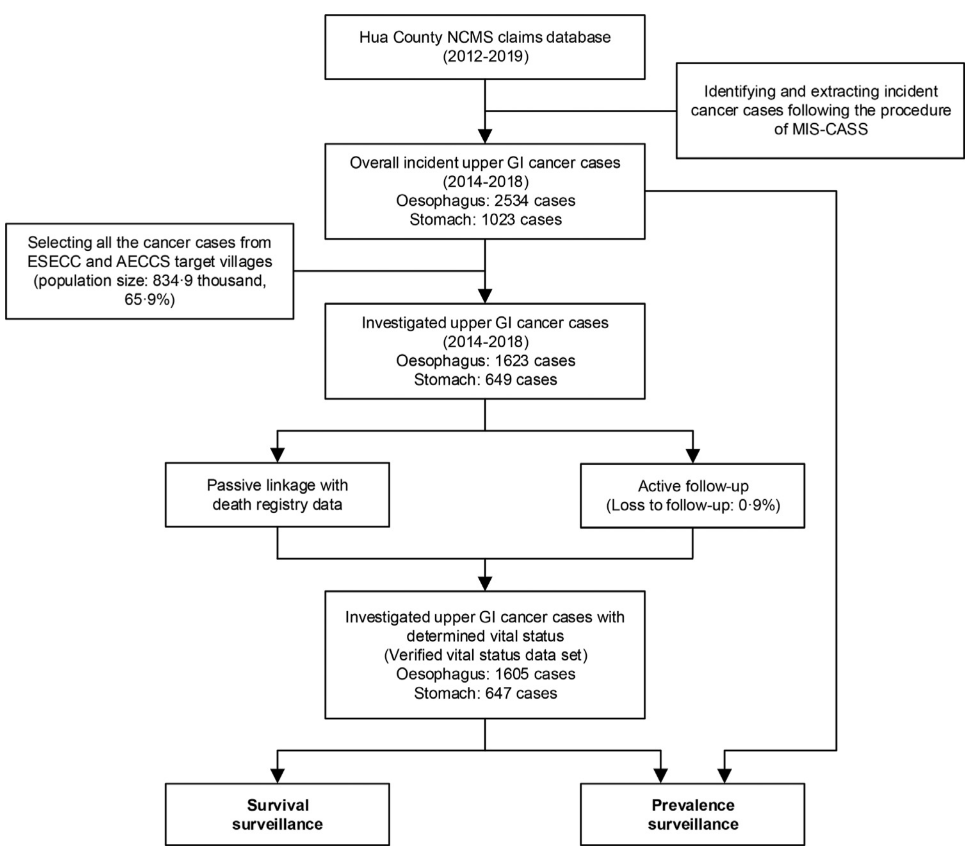

Fig. 1. Procedure for survival and prevalence surveillance of upper GI cancers based on the MIS-CASS and death surveillance data in Hua County, China.

NCMS=The New Rural Co-operative Medical Care Scheme. MIS-CASS=Medical-Insurance-System-based Cancer Surveillance System. GI=gastrointestinal. ESECC=Endoscopic Screening for Esophageal Cancer in China. AECCS=Anyang Esophageal Cancer Cohort Study.

survival. Compared to observed survival, relative survival better measures the excess mortality experienced by cancer patients [24]. To further consider comparability of survival among patient groups in different periods and regions, we categorized cancer cases into five age groups (0–44, 45–54, 55–64, 65–74, and 75–99 years), and calculated age-standardised relative survival according to the International Cancer Survival Standard weights: 0·07 (0–44 years), 0·12 (45–54 years), 0·23 (55–64 years), 0·29 (65–74 years), and 0·29 (75–99 years) [25]. Age standardization was carried out using the Brenner and Hakulinen approach to overcome the practical difficulties of age standardization due to lack of data in a single age group [26]. In addition, the Kaplan-Meier method was used to compare observed survival among groups, and Cox regression models were used to calculate hazard ratios (HR) as well as 95% confidence intervals.

2.3.2. *Calculation of cancer prevalence*

We estimated upper GI cancer prevalence in Hua County in 2018 by using two methods, namely the "indirect method" and the "direct method".

For the indirect method, the number of prevalent cases was estimated from survival estimates and year-specific incident cases based on the following formula:

$$P_k(n) = \sum_{i=1}^{n} IC_{k-i} * S_{k-i}(i - 0 \cdot 5)$$

where IC_k is the number of incident cancer cases at age k in Hua County, and $S_k(t)$ is the observed survival at time t after diagnosis for cases diagnosed at age k [2]. The prevalence rate was calculated as the count of 5-year prevalent cases in Hua County in 2018 divided by the total population of Hua County in 2018. The assumptions made for estimating prevalence based on the indirect method included: (a) annual incident cases occur at mid-year and (b) the curve of observed survival with a 6-month time interval follows the Weibull distribution [4].

In the direct method, the cancer prevalence rate in Hua County was represented by the prevalence estimates in the investigated AECCS and ESECC target villages (66% of the total population in Hua County). In brief, the prevalence rate was directly calculated as the count of investigated cases newly-diagnosed in 2014–2018 who were still alive at the end of 2018 divided by the corresponding population in the target villages in 2018.

2.3.3. *Imputation-based estimation of survival and prevalence for overall cancer and common cancers*

Through comparing the vital status information from passive linkage with death surveillance data and that from active follow-up, we observed that 21·7% cancer deaths were missed by the death surveillance system, and there was no significant discrepancy in male and female, young and old, or oesophageal cancer and stomach cancer (Supplementary Table 1, Supplementary Fig. 1). Assuming that under-reporting of death events occurred randomly, we used an imputation-based method to estimate survival and prevalence of overall cancer as well as common cancer types, based on county-wide data for cancer incidence (2014–2018) and death (2014–2019). In this regard, we imputed the under-reported dead cases through random selection of $\left(\frac{1}{1-21 \cdot 7\%} - 1\right) \times N$ "alive cases" (N is the number of patients with the cancer of interest whose death was actually recorded in the death surveillance system), using gender, cancer site and the year of diagnosis as stratification variables. This imputation-based approach was repeated 1000 times to estimate the survival and prevalence of cancer overall and five cancer types with the highest incidence in Hua County [15].

Stata 15.0 and R statistics version 3.5.0 were used for all analysis (strs in Stata was used for survival analysis).

2.4. *Ethics committee approval*

This study was approved by the Institutional Review Board of the Peking University School of Oncology, China. Informed consent was obtained from all the interviewees.

2.5. *Role of the funding source*

The funders of the study had no role in study design, collection, analysis, and interpretation of data, or writing of the report. HT, ZH, and YK had full access to all the data in the study, and the corresponding authors were responsible for the decision to submit for publication.

3. Results

3.1. *Survival in upper GI cancers*

In Hua County, the age-standardised relative 5-year survival was 39·2% (male: 36·8%; female: 43·6%) for oesophageal cancer and 33·3% (male: 29·6%; female: 43·4%) for stomach cancer. The observed survival estimates were lower than the corresponding age-standardised relative survival estimates for oesophageal cancer, but not for stomach cancer (Table 1).

Older groups in this population showed poorer survival. For oesophageal cancer, the combined relative 5-year survival for males and females was 48·2%, 51·7%, 35·5%, and 18·9% for patients diagnosed at ages 45–54, 55–64, 65–74, and 75–99 years respectively. Notably,

Table 1
Survival of upper GI cancer patients diagnosed in 2014–2018 in Hua County, China.

Site	Gender	Median survival time (month)	Observed survival (95% CI)			Age-standardised relative survival (95% CI)		
			1-year	3-year	5-year	1-year	3-year	5-year
Oesophagus								
	All	23·2	67·7% (65·3, 69·9)	40·6% (38·1, 43·2)	31·2% (28·4, 34·1)	70·9% (68·1, 73·6)	46·1% (42·8, 49·4)	39·2% (35·7, 42·8)
	Male	22·0	67·1% (63·9, 70·1)	39·2% (35·7, 42·6)	28·6% (24·9, 32·4)	69·6% (65·5, 73·3)	44·0% (39·4, 48·5)	36·8% (32·0, 41·7)
	Female	24·5	68·3% (64·8, 71·6)	42·4% (38·6, 46·3)	34·6% (30·3, 38·9)	74·0% (70·1, 77·4)	50·5% (45·3, 55·5)	43·6% (37·8, 49·2)
Stomach								
	All	23·2	69·4% (65·7, 72·8)	38·8% (34·6, 42·9)	32·8% (28·4, 37·4)	68·7% (64·4, 72·6)	38·3% (33·6, 42·9)	33·3% (28·3, 38·5)
	Male	22·6	69·8% (65·4, 73·7)	36·0% (31·2, 40·8)	28·8% (23·7, 34·2)	69·2% (63·9, 73·9)	35·5% (30·0, 41·1)	29·6% (23·9, 35·7)
	Female	26·3	68·4% (61·0, 74·7)	46·1% (38·1, 53·8)	43·4% (35·0, 51·5)	67·3% (59·3, 74·2)	45·6% (36·9, 53·9)	43·4% (34·1, 52·5)

GI=gastrointestinal. CI=confidence interval.

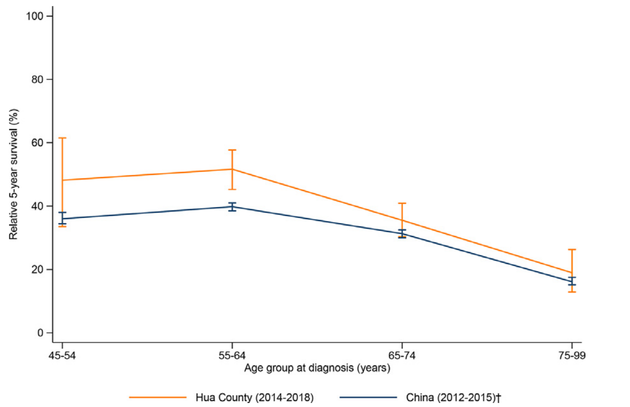

Fig. 2. Relative 5-year survival (with 95% CI) for oesophageal cancer by age group at diagnosis (45–99 years) in Hua County, China, 2014–2018.
CI=confidence interval.
†Source: Adapted from Zeng et al. [12].

survival at ages 45–54 and 55–64, but not at older ages was higher in Hua County than the reported national average estimates for patients diagnosed in 2012–2013 and followed up to 2015 (Fig. 2). For stomach cancer, relative 5-year survival was 47·9%, 49·4%, and 38·3% for patients of 45–54 years, 55–64 years, and 65–74 years (no estimate was available for patients older than 74 because the longest individual follow-up time after diagnosis was less than five years), similar to the national average estimates for patients diagnosed in 2012–2013 (Supplementary Fig. 2).

3.2. Age-gender interaction of survival of upper GI cancers

Overall, survival of oesophageal cancer was better in women than in men ($HR_{\text{adjusted for age}}$=0·81, 95% CI: 0·71–0·92). We further stratified patients into three age groups at diagnosis, and the adjusted female-to-male HRs were 0·65 (95% CI: 0·48–0·87), 0·73 (95% CI: 0·60–0·90), and 0·96 (95% CI: 0·78–1·18) for age groups 0–64, 65–74, and 75–99 years respectively ($p_{interaction}$=0·005) (Fig. 3). A similar trend was also found in stomach cancer patients, although the *p* value for interaction was not significant (Supplementary Fig. 3).

3.3. Prevalence of upper GI cancers

According to the indirect method, the 5-year prevalence rate for oesophageal cancer was 99·8/100,000 (male: 105·9/100,000; female: 93·3/100,000; male-to-female ratio: 1·1). For stomach cancer, the 5-year prevalence rate was 41·5/100,000, with a larger gender discrepancy than oesophageal cancer (male: 57·4/100,000; female: 24·5/100,000; male-to-female ratio: 2·3). The prevalence/incidence ratio was 2·7 for oesophageal cancer and 2·4 for stomach cancer (Table 2) [15]. For both oesophageal cancer and stomach cancer, the absolute number of prevalent cases peaked at 65–74 years, while the age-specific prevalence rate peaked at 70–79 years (Fig. 4). Among all 5-year prevalent cases of upper GI cancers in 2018, patients who had been diagnosed 2–3 years earlier constituted the largest group (Supplementary Fig. 4). Similar results were observed with the direct method (Supplementary Table 2).

3.4. Imputation-based estimations of survival and prevalence for overall cancer and common site-specific cancers

Based on the county-wide dataset of cancer incidence and death, as well as a 21·7% under-reporting rate of death surveillance system, we re-estimated relative 5-year survival and prevalence rates for oesophageal cancer and stomach cancer using an imputation-based method, and observed results similar to the above real-world results (oesophageal cancer survival: 39·6% vs. 39·2%; stomach cancer survival: 34·8% vs. 33·3%; oesophageal cancer prevalence rate: 103·1/100,000 vs. 99·8/100,000; stomach cancer prevalence rate: 40·9/100,000 vs. 41·5/100,000). Based on this, we further estimated survival and prevalence for other cancers in Hua County. The overall

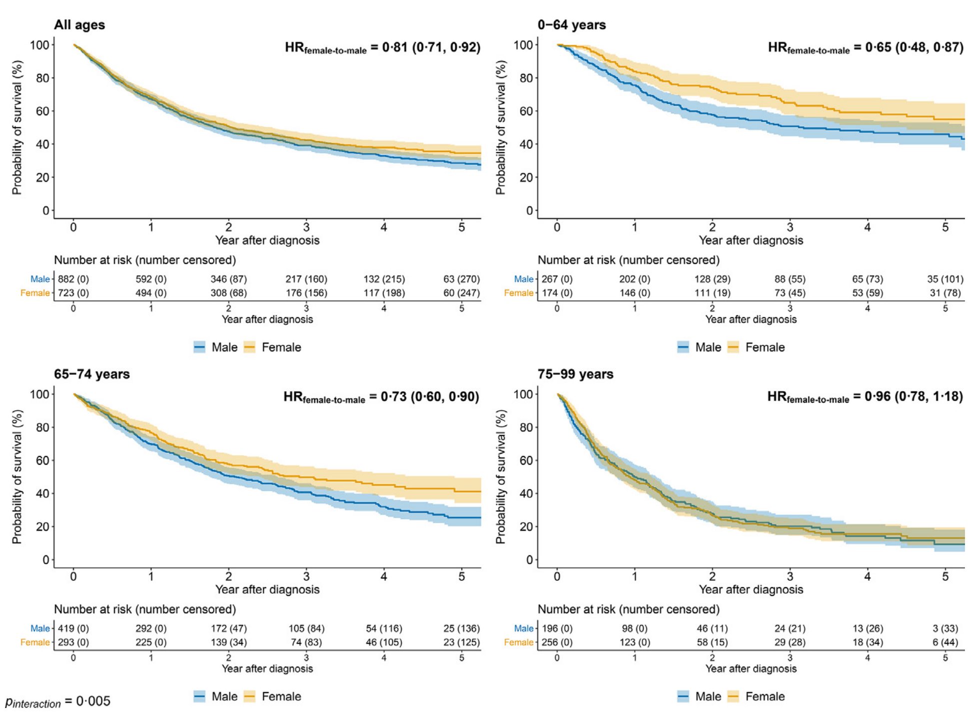

Fig. 3. Kaplan-Meier survival curves (with 95% CI) for oesophageal cancer by age group at diagnosis in Hua County, China, 2014–2018.
HR=hazard ratio (adjusted for age). CI=confidence interval.

Table 2
Prevalence of upper GI cancers in Hua County, China, 2018 (Indirect Method).

Site	Gender	Prevalent cases in Hua County 1 year	3 years	5 years	5-year prevalence rate [1/10^5 (95% CI)]	New cases in Hua County in 2018	Incidence rate (1/10^5)	P:I ratio
Oesophagus								
	All	379	885	1264	99·8 (94·4, 105·5)	474	37·4	2·7
	Male	219	490	694	105·9 (98·2, 114·1)	277	42·3	2·5
	Female	160	395	570	93·3 (85·8, 101·3)	197	32·2	2·9
Stomach								
	All	180	396	525	41·5 (38·0, 45·2)	218	17·2	2·4
	Male	137	291	376	57·4 (51·7, 63·5)	167	25·5	2·2
	Female	43	105	150	24·5 (20·8, 28·8)	51	8·3	2·9

GI=gastrointestinal. P:I = 5-year cancer prevalence to incidence. CI=confidence interval.

relative 5-year survival for cancer patients in Hua County was estimated as 44·5% (95% CI: 42·9–46·0), and the overall 5-year prevalence rate for cancer patients in Hua County was estimated as 585·1/100,000. The relative 5-year survival of common cancers in Hua County ranged from 17·5% (liver cancer in men) to 83·9% (thyroid cancer in women). The prevalence rate of common cancers in Hua County ranged from 22·1/100,000 (liver cancer in men) to 175·8/100,000 (female breast cancer). (Supplementary Table 3)

4. Discussion

Comprehensive surveillance of cancer incidence, prevalence and survival is critical for optimizing strategies for cancer prevention and control, and for efficiently allocating medical resources, as well as for evaluating the performance of clinical treatment. However, under the current situation in China, it is impossible to obtain estimates of cancer prevalence and survival in areas not covered by PBCRs. In fact, even in areas covered by PBCRs, it is difficult to produce up-to-date estimates in a regular and timely manner. In this study, we have established a new approach to surveillance for cancer prevalence and survival based on population-level medical insurance claims data and death surveillance data. For the first time, we have reported the survival and prevalence of upper GI cancers in an area which is high-risk for oesophageal cancer that is not covered by PBCRs in China, and in addition we have estimated survival and prevalence of overall cancer and common cancer types using an imputation-based method. This study is one extension of our MIS-CASS which was primarily applicable to cancer incidence surveillance, and this approach can be regularly carried out with broader population coverage in China, given that basic medical insurance covers 95% of the national population, and incident cancer cases can be ascertained using MIS-CASS with only a short time lag (~ six months) between a cancer diagnosis and its registration in the insurance database [15].

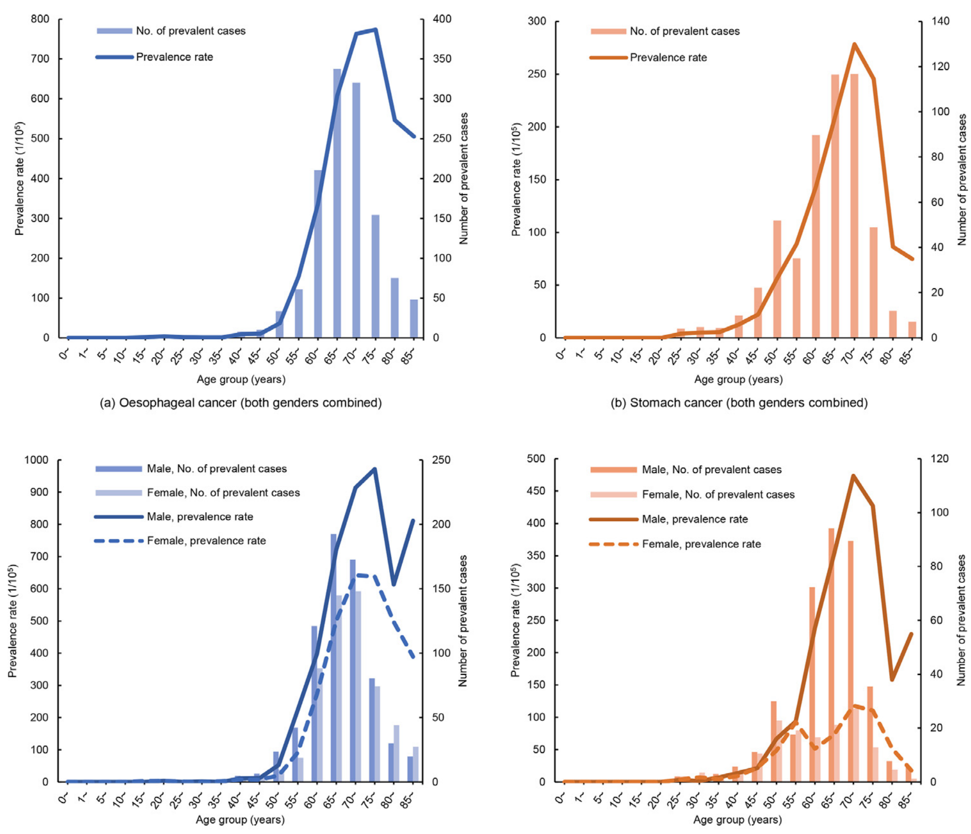

Fig. 4. Age-specific 5-year prevalence for upper GI cancers in Hua County, China, 2018. GI=gastrointestinal.

Hua County is a region which is well-known to be at high-risk for oesophageal cancer, however no cancer survival statistics have been reported locally due to the lack of cancer registries [15–17]. According to our estimation in this study, the age-standardised relative 5-year survival of oesophageal cancer in Hua County is 39·2%, which is notably higher than the average national level (30·3%), especially in the 45–64 age group (relative 5-year survival ratio: 1·3) [12]. This may be explained in part by a series of screening programmes conducted in this local area. In 2006, we initiated the AECCS and over 8000 residents aged 25–65 years from Hua County underwent endoscopic screening for upper GI cancers [16,27]. In 2012, we initiated the ESECC randomised controlled trial in Hua County and a total of 15,188 local residents aged 45–69 years were screened endoscopically from 2012 to 2016 [17]. In addition, after 2011, Hua County has been covered by the Screening, Early Diagnosis, and Early Treatment program for upper GI cancers (1500–2000 enrollees annually, 40–69 years), which is subsidized by the central government of China [28,29]. These cancer screening programmes promote detection, diagnosis, and treatment of early oesophageal cancer cases, and this may have in part contributed to the notably better survival of oesophageal cancer in the corresponding age group locally. However, it must be noted that whether early screening efforts do indeed improve the prognosis of oesophageal cancer cannot be answered by observing the survival of cancer patients per se due to lead-time and length-time bias, and population-based randomised controlled trials are still warranted [17,27,30]. Unlike oesophageal cancer, the survival of stomach cancer in Hua County (33·3%) is similar to the average level for China (35·1%) [12]. Given that stomach cancer is not a high-risk cancer in Hua County, the protective effect of upper GI tract endoscopy on stomach cancer might not be as obvious as that for oesophageal cancer at the population level.

Overall, upper GI cancer cases in males had a higher mortality risk than those in females, which is consistent with the general pattern in China [12]. However, stratified analysis by age demonstrated notable heterogeneity in the association of gender and mortality risk for upper GI cancers across age groups at diagnosis, especially for oesophageal cancer (Fig. 3, Supplementary Fig. 3). For oesophageal cancer patients diagnosed before the age of 75, the mortality risk in males was 1·4–1·5 times as high as that in females, but no appreciable discrepancy in mortality risk in male versus female patients was observed in patients diagnosed after 75. This indicates male patients and female patients are comparably susceptible to mortality in older groups, where age rather than gender serves as the determinant of cancer prognosis.

The oesophageal cancer prevalence in Hua County is 2·6 times as high as the national level (Table 2) [14]. This may largely be explained by the fact that Hua County is a high-risk area for oesophageal cancer, with an incidence 2·1 times as high as national level [15,17]. For both oesophageal cancer and stomach cancer, the prevalence rate in this study was extremely low in individuals of less than 50 years, but increased with age and peaked in the age group 70–79 years. This pattern of age-specific prevalence rates in upper GI cancers is consistent with the overall trend of age-specific incidence rates (Supplementary Fig. 5) [15]. For both the prevalence count and prevalence rate, the burden of upper GI cancers was the greatest in individuals aged 65–79 years, who generally suffer from more health inequalities in rural China [31]. This suggests that adjustments should be made in terms of medical resource allocation and medical insurance policies to meet the needs of patients in this age group.

As one of the important preconditions for this new approach to surveillance, the evaluation of quality of local death surveillance data was conducted through comparing the performance of passive and active follow-up for upper GI cancers. We observed that the Hua County death registry reported death events with moderate sensitivity (78·3%) and high specificity (99·1%), and no significant discrepancy in the rate of under-reporting was found by gender, age group, or cancer site (Supplementary Table 1, Supplementary Fig. 1). As such, we further conducted a simulation analysis (repeated 1000 times) to estimate the survival and prevalence of overall cancer and five major cancers in Hua County, by randomly imputing the under-reported deaths from the "alive" patients. The 1000 imputation-based estimates for each common cancer were within a relatively narrow range (mostly <10%), and the median estimates for both oesophageal cancer and stomach cancer were very close to the estimates which resulted from real investigation, with a relative difference of less than 4·5% (Table 1, 2, Supplementary Table 2, 3, Supplementary Fig. 6). However, there is a caveat in applying the imputation-based approach since under-reporting rates of death may vary in different cancer types. Given that under-reporting is a common issue for death surveillance system in China (12%–17% at the national level), we strongly recommend assessing the data quality of a given local death surveillance system before using this method [11,32]. Active follow-up of a representative sample of cancer patients would allow estimation of the level of under-reporting in the local death surveillance system, and the extent to which it varies by critical factors such as cancer site and patient characteristics such as gender, age, and area of residence.

Our study showed that the proposed MIS-CASS approach for surveillance of cancer prevalence and survival has three main strengths. First, its coverage is high, as both the medical insurance system (e.g. NCMS) and the death surveillance system which are the two data sources used to determine the occurrence and endpoint of a cancer case, covered the whole population under study [15]. Over 99% (668/674) of the villages we investigated in this study were target villages in the ESECC trial which were randomly selected from the ~1000 villages in Hua County, and this further ensures the representativeness of the study sample [17]. Second, with use of the MIS-CASS approach, where incident cancer cases can be identified with a delay of only six months, up-to-date prevalence and survival estimates can be produced and reported in almost real-time. This would greatly facilitate timely assessment of cancer burden and cancer-specific prognosis, as well as performance of health care programmes by the government. Third, the prevalence and survival estimates for upper GI cancers in this study are robust, given that similar results were observed with both the indirect and direct methods (e.g. the relative difference in prevalence rates was less than 10%), and similar results were also found in both the real-world data-based and imputation-based approaches (shown in Results). Based on these, this new approach to cancer prevalence and survival surveillance has potential for regular use in wider administrative areas including those not covered by PBCRs.

However, this study also has limitations. Although rural Hua County has a population of 1·27 million and the villages we investigated included as many as 835 thousand individuals, this is nonetheless a single-center study conducted at the county level. It is necessary to conduct comparative studies in areas covered by high-quality PBCRs in China to further evaluate the performance of this novel approach. Given that cancer is still a relatively rare disease even in a high-risk area, estimates of the prevalence and survival of low incidence cancer types (such as nasopharyngeal cancer and melanoma of the skin) may not be statistically robust [15]. This concern could be addressed by using this method in a larger area with a larger population. Moreover, caution must be exercised when applying this approach to surveillance in areas with different levels of socioeconomic development, and evaluation of the accuracy and completeness of local death surveillance data is needed.

The advent of the "big data" era and the increasing availability of electronic health records, has resulted in a promise of a real-time multi-dimensional cancer surveillance system with a wider population coverage. This study, together with our previous report, describe a novel comprehensive approach to surveillance of cancer in China which incorporates three critical dimensions – incidence, prevalence, and survival – by bringing together multiple "big" routine data sources including medical insurance claims data and death surveillance data [15]. The latter covers only 24% of the national population, but the Healthy China Programme (2019–2030) has recently emphasized the urgent need to further expand death surveillance in China with the ultimate goal of establishing a nation-wide surveillance system. In short, this newly proposed cancer surveillance approach – which provides valid and up-to-date cancer incidence, prevalence, and survival estimates in a timely way – offers great potential for being scaled up for use in much wider geographical areas, particularly in those that are not currently covered by PBCRs at a relatively low cost by relying on existing data sources.

Declaration of Competing Interest

The authors have no conflict of interest for the publication of this study.

Funding

This work was supported by the National Key R&D Program of China (2016YFC0901404), the National Science & Technology Fundamental Resources Investigation Program of China (2019FY101102), the National Natural Science Foundation of China (82073626), the Taikang Yicai Public Health and Epidemic Control Fund (TKYC-GW-2020), the Beijing-Tianjin-Hebei Basic Research Cooperation Project (J200016), and the Digestive Medical Coordinated Development Center of Beijing Hospitals Authority (XXZ0204) .

Contributors

YK, ZH, and HT contributed to the conception and design of the study. YH, QL, LL, CG, and HT contributed to the acquisition of data. ZH and HT contributed to data analysis. YK, ZH, Id-S-S, HT, ZL, ML, FL, YL, and YP contributed to interpretation of data and checking of the results. HT, ZH, YK, and Id-S-S contributed to drafting of the manuscript, which was reviewed and approved by all coauthors.

Data sharing statement

Data used in this study are available from the corresponding authors upon reasonable request.

Acknowledgements

We gratefully acknowledge all the following team members and collaborators for their contributions to the fieldwork done for this study: Hui Wang, Linlin Liang, and Zifan Qi from Peking University Cancer Hospital & Institute; 18 employed investigators in Hua County including Lijuan An, Jiaqi Duan, Xiangyun Fang, Ruifang Li, Junying Lv, Sumin Qi, Shaorui Geng, Lu Liu, Yixin Bi, Panyue Chen, Saijiao Duan, Ruru Guo, Ke Li, Ruixiang Liu, Junmeng Shi, Jiaqi Wang, Xiaopeng Wang, and Yinghui Zheng; 674 village doctors in the target villages of the ESECC and AECCS cohorts in Hua County; 22 township leaders in Hua County. We would also like to thank the People's Government of Hua County, the Health Commission of Hua County, and Hua County People's Hospital.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2021.100756.

References

[1] Capocaccia R, Colonna M, Corazziari I, et al. Measuring cancer prevalence in Europe: the EUROPREVAL Project. Ann Oncol 2002;13:831–9.
[2] Pisani P, Bray F, Parkin DM. Estimates of the world-wide prevalence of cancer for 25 sites in the adult population. Int J Cancer 2002;97:72–81.
[3] Sankaranarayanan R, Swaminathan R, Lucas E. Cancer survival in Africa, Asia, the Caribbean and Central America (SurvCan). Lyon, France: International Agency for Research on Cancer; 2011.
[4] Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer 2013;132:1133–45.
[5] National Cancer Center. Chinese guideline for cancer registration. Beijing, China: People's Medical Publishing House; 2016.
[6] He J. China cancer registry annual report 2018. Beijing, China: People's Medical Publishing House; 2019.
[7] Wei W, Zeng H, Zheng R, et al. Cancer registration in China and its role in cancer prevention and control. Lancet Oncol 2020;21:e342–e9.
[8] Bray F, Colombet M, Mery L, et al. Cancer incidence in five continents, XI. Lyon, France: International Agency for Research on Cancer; 2017.
[9] National Bureau of Statistics of China. China statistical yearbook 2018. [in Chinese] http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm [Accessed 10 Jan 2020].
[10] Yang G, Hu J, Rao KQ, Ma J, Rao C, Lopez AD. Mortality registration and surveillance in China: history, current situation and challenges. Popul Health Metr 2005;3:3.
[11] Liu S, Wu X, Lopez AD, et al. An integrated national mortality surveillance system for death registration and mortality surveillance, China. Bull World Health Organ 2016;94:46–57.
[12] Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. The Lancet Global health 2018;6:e555–e67.
[13] Zeng H, Zheng R, Guo Y, et al. Cancer survival in China, 2003–2005: a population-based study. Int J Cancer 2015;136:1921–30.
[14] Zheng R, Zeng H, Zhang S, Chen T, Chen W. National estimates of cancer prevalence in China, 2011. Cancer Lett 2016;370:33–8.
[15] Tian H, Yang W, Hu Y, et al. Estimating cancer incidence based on claims data from medical insurance systems in two areas lacking cancer registries in China. EClinicalMedicine 2020;20:100312.
[16] Liu F, Guo F, Zhou Y, et al. The anyang esophageal cancer cohort study: study design, implementation of fieldwork, and use of computer-aided survey system. PLoS ONE 2012;7:e31602.
[17] He Z, Liu Z, Liu M, et al. Efficacy of endoscopic screening for esophageal cancer in China (ESECC): design and preliminary results of a population-based randomised controlled trial. Gut 2019;68:198–206.
[18] National Bureau of Statistics of China. Statistical Communiqué of the People's Republic of China on the 2018 National Economic and Social Development. 2019. http://www.stats.gov.cn/english/PressRelease/201902/t20190228_1651335.html [Accessed 17 Jan 2020].
[19] Hua County Bureau of Statistics. Statistical bulletin of national economy and social development of Hua County in 2017 [in Chinese]. 2018. http://www.ha.stats.gov.cn/2018/06-19/1367984.html [Accessed 17 Dec 2018].
[20] Hua County Bureau of Statistics. Statistical bulletin of national economy and social development of Hua County in 2018 [in Chinese]. 2019. http://www.ha.stats.gov.cn/2019/05-13/1367989.html [Accessed 17 Jan 2020].
[21] Chen W, Sun K, Zheng R, et al. Cancer incidence and mortality in China, 2014. Chin J Cancer Res 2018;30:1–12.
[22] Liu M, Liu Z, Cai H, et al. A model to identify individuals at high risk for esophageal squamous cell carcinoma and precancerous lesions in regions of high prevalence in China. Clin Gastroenterol Hepatol 2017;15:1538–46 .e7.
[23] Liu M, Liu Z, Liu F, et al. Absence of iodine staining associates with progression of esophageal lesions in a prospective endoscopic surveillance study in China. Clin Gastroenterol Hepatol 2020;18:1626–35 .e7.
[24] Cho H, Howlader N, Mariotto AB, Cronin KA. Estimating relative survival for cancer patients from the Seer program using expected rates based on Ederer I versus ederer II method. Surveillance Research Program, National Cancer Institute; 2011 https://surveillance.cancer.gov/reports [Accessed 18 Mar 2020].
[25] Corazziari I, Quinn M, Capocaccia R. Standard cancer patient population for age standardising survival ratios. Eur J Cancer 2004;40:2307–16.
[26] Brenner H, Hakulinen T. Age adjustment of cancer survival rates: methods, point estimates and standard errors. Br J Cancer 2005;93:372–5.
[27] Liu M, He Z, Guo C, et al. Effectiveness of intensive endoscopic screening for esophageal cancer in china: a community-based study. Am J Epidemiol 2019;188:776–84.
[28] Zou XN. Epidemic trend, screening, and early detection and treatment of cancer in Chinese population. Cancer Biol Med 2017;14:50–9.
[29] Hua County People's Hospital. The expert group of the National Health and Family Planning Commission visited Hua County People's Hospital to supervise the early diagnosis and early treatment programme of upper gastrointestinal cancer [in Chinese]. 2017. http://hxrmyy.hnhx.gov.cn/info/1004/2271.htm [Accessed 24 Jun 2020].
[30] Yang S, Wu S, Huang Y, et al. Screening for oesophageal cancer. Cochrane Database Syst Rev 2012;12:Cd007883.
[31] Pan C, Fan Q, Yang J, Deng D. Health inequality among the elderly in rural china and influencing factors: evidence from the Chinese longitudinal healthy longevity survey. Int J Environ Res Public Health 2019;16:4018.
[32] Guo K, Yin P, Wang L, et al. Propensity score weighting for addressing underreporting in mortality surveillance: a proof-of-concept study using the nationally representative mortality data in China. Popul Health Metr 2015;13:16.