



Research Paper

Estimating cancer incidence based on claims data from medical insurance systems in two areas lacking cancer registries in China

Hongrui Tian^{a,1}, Wei Yang^{b,1}, Yanjun Hu^{c,1}, Zhen Liu^{a,1}, Lei Chen^b, Liang Lei^c, Fan Zhang^b, Fen Cai^b, Huawen Xu^b, Mengfei Liu^a, Chuanhai Guo^a, Yun Chen^d, Ping Xiao^d, Junhui Chen^d, Ping Ji^d, Zhengyu Fang^d, Fangfang Liu^a, Ying Liu^a, Yaqi Pan^a, Isabel dos-Santos-Silva^e, 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 Cancer Hospital of Shantou University Medical College, Guangdong Province, China

^c Healthcare Security Administration of Hua County, Henan Province, China

^d Peking University Shenzhen Hospital, Guangdong Province, China

^e Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

ARTICLE INFO

Article History:

Received 23 January 2020

Revised 19 February 2020

Accepted 24 February 2020

Available online xxx

Key laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Genetics, Peking University Cancer Hospital & Institute, Beijing, PR China (H Tian BCh, Z Liu PhD, Prof M Liu PhD, C Guo BCh, Prof F Liu PhD, Y Liu PhD, Y Pan BCh, Prof Z He PhD, Prof Y Ke MD); Cancer Hospital of Shantou University Medical College, Guangdong Province, PR China (Prof W Yang MM, Prof L Chen BCh, F Zhang MM, F Cai MM, H Xu BCh); Healthcare Security Administration of Hua County, Henan Province, PR China (Y Hu BCh, L Lei BCh); Peking University Shenzhen Hospital, Guangdong Province, PR China (Prof Y Chen MD, Prof P Xiao MD, Prof J Chen MD, Prof P Ji PhD, Prof Z Fang PhD); Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK (Prof I dos-Santos-Silva PhD)

Keywords:

Medical insurance system

Claims data

Cancer surveillance

Cancer incidence

China

SUMMARY

Background: We aimed to establish a Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS) in China and evaluate the completeness and timeliness of this system through reporting cancer incidence rates using claims data in two regions in northern and southern China.

Methods: We extracted claims data from medical insurance systems in Hua County of Henan Province, and Shantou City in Guangdong Province in China from Jan 1, 2012 to Jun 30, 2019. These two regions have been considered to be high risk regions for oesophageal cancer. We developed a rigorous procedure to establish the MIS-CASS, which includes data extraction, cleaning, processing, case ascertainment, privacy protection, etc. Text-based diagnosis in conjunction with ICD-10 codes were used to determine cancer diagnosis.

Findings: In 2018, the overall age-standardised (Segi population) incidence rates (ASR World) of cancer in Hua County and Shantou City were 167.39/100,000 and 159.78/100,000 respectively. In both of these areas, lung cancer and breast cancer were the most common cancers in males and females respectively. Hua County is a high-risk region for oesophageal cancer (ASR World: 25.95/100,000), whereas Shantou City is not a high-risk region for oesophageal cancer (ASR World: 11.43/100,000). However, Nanao island had the highest incidence of oesophageal cancer among all districts and counties in Shantou (ASR World: 36.39/100,000). The age-standardised male-to-female ratio for oesophageal cancer was lower in Hua County than in Shantou (1.69 vs. 4.02). A six-month lag time was needed to report these cancer incidences for the MIS-CASS.

Interpretation: MIS-CASS efficiently reflects cancer burden in real-time, and has the potential to provide insight for improvement of cancer surveillance in China.

Funding: The National Key R&D Program of China (2016YFC0901404), the Digestive Medical Coordinated Development Center of Beijing Municipal Administration of Hospitals (XXZ0204), the Sanming Project of Shenzhen (SZSM201612061), and the Shantou Science and Technology Bureau (190829105556145, 180918114960704).

© 2020 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

* Corresponding authors.

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

¹ These authors contributed equally to this paper.

Research in context

Evidence before this study

Surveillance-site-based cancer registration has been the only approach for estimating the national incidence of cancer in China. However due to the vast territory, huge population, and limited resources of China, it is almost impossible to establish a nationwide population-level cancer surveillance system with a short time lag using this traditional method. Claims data in health insurance systems have been used to identify incident cancer cases in western countries (We searched PubMed, Google Scholar, and China National Knowledge Infrastructure for studies published before Sep 1, 2019, with no language restriction, using the terms “cancer incidence”, “health insurance system”, “claims data”, etc.). Moreover, we have already conducted two comparative studies which demonstrate the feasibility and accuracy of identifying cancer cases using claims data extracted from health insurance systems in rural China. However, whether the Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS) can be applied to urban as well as rural areas on a large scale in China has as yet not been determined.

Added value of this study

We, for the first time, established a MIS-CASS and evaluated its performance in both urban and rural China in this study. We demonstrated that claims data from medical insurance systems have great potential for application in identifying newly diagnosed cancer cases, and showed that the MIS-CASS is feasible, representative, accurate, timely and inexpensive for monitoring the current status and trends of all-site cancer incidence in China.

Implications of all the available evidence

We recommend adoption and further evaluation of the MIS-CASS on a larger scale, which will provide new insight into the reform and development of national level cancer surveillance in China.

Introduction

Cancer is the second most frequent cause of death among non-communicable diseases. In 2018, there were 18.1 million new cancer cases worldwide and 9.6 million persons died of cancer [1]. In China, there were 4.3 million new cancer cases in 2018, and 6.7 million cases are projected to occur in 2040 [2]. Continuous and accurate surveillance of cancer incidence is crucial for determination of actual cancer burden, and for efficient allocation of medical resources, as well as for timely launching of preventive strategies.

Cancer registration is defined as “the process of continuing, systematic collection of data on the occurrence and characteristics of reportable neoplasms” [3]. Population-based cancer registration (PBCR) is acknowledged to be the gold standard method for reporting cancer incidence in a defined population [4]. Currently, national-level cancer registration has been established in some developed countries such as Austria, Belgium, Canada, Finland, Ireland, and South Korea [5]. In China however, it is unrealistic to establish real nation-wide PBCR due to the vast territory of China and its huge population and limited resources. Sentinel surveillance, that is, establishing surveillance sites in certain areas to estimate national level cancer incidence, has to date been the only approach used in China. As many as 449 cancer surveillance sites had been established in China up to 2014. However, only 36 of these surveillance sites which were mostly in areas of high-risk for specific cancers or economically developed regions were of high

quality and were included in the latest volume of Cancer Incidence in Five Continents, accounting for 5% of the national Chinese population [6–8]. Moreover, there has been a characteristic three-year lag in reporting cancer statistics in China due to the time and resource consuming process of data collection, quality control, and data analysis [9]. For example, the latest annual cancer report (published in Jan 2019) was based on data for the year of 2015 [10].

Claims data in health insurance systems have been applied to identify incident cancer cases in some western countries, and the performance of this method has been previously validated [11,12]. In China, basic medical insurance systems are completely administered and operated by the central and local governments. Urban Employee Basic Medical Insurance (UEBMI, initiated in 1998), the New Rural Co-operative Medical Care Scheme (NCMS, initiated in 2003), and Urban Residents Basic Medical Insurance (URBMI, initiated in 2007) constitute China’s basic medical insurance systems, and cover over 95% of the national population [13]. In 2018, the establishment of the National Healthcare Security Administration directly under the State Council brought about integration of China’s basic medical insurance systems, which has facilitated the standardised management of medical claims data in China [14].

Accuracy (validity), completeness, comparability and timeliness are the crucial principles for cancer surveillance [15,16]. The quality of claims data per se determines the accuracy and completeness of collecting cancer cases, and two of our preceding studies have addressed this issue. In one study we compared active follow-up (door-to-door interviews) with passive linkage with NCMS claims data for identification of incident cancer cases in a large-scale randomized controlled trial involving 33,948 subjects from 2012 to 2017, and found that the overall sensitivity of passive linkage with NCMS claims data was significantly higher than that of active follow-up (95.6% vs 54.9%, $p < 0.001$) [17]. In another study, we further verified claims-based diagnoses with corresponding electronic medical records at an individual level in the general population in Hua County in 2017. We found that use of claims data identified cancer patients with high sensitivity (92.0–95.2%) and positive predictive value (89.0–97.8%) [18]. Both of these two comparative studies demonstrated that claims data extracted from health insurance systems is an ideal and accurate data source for identification of cancer cases. At the same time, claims data are generated and submitted to corresponding medical insurance systems in real time, which notably lowers the cost and the time lag involved in data collection and quality control [19]. Based on these findings, we assumed that establishing a population-level cancer surveillance system based on medical insurance system claims data in China would be feasible, representative, accurate, timely, and inexpensive.

Hua County (Taihang Mountain area) in Henan Province and Shantou City (Chaoshan area) in Guangdong Province are known as areas of high oesophageal cancer incidence (Supplementary Figure 1) [20,21]. However, no cancer registries have been established in either of these two regions, leaving uncertainty about the actual patterns of cancer incidence and oesophageal cancer burden there.

In this study, we established a Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS) and evaluated the completeness and timeliness of the MIS-CASS by reporting all-site cancer incidence in 2018 together with temporal trends of common cancer incidence from 2014 to 2018 in Hua County and Shantou City.

Method

Medical insurance systems in Hua County and Shantou City

Hua County is located in northern Henan Province and covers an area of 700 square miles (Supplementary Figure 1). This area is predominantly rural and 90% of the total population is involved in agriculture. In 2018, Hua County was home to 1.07 million permanent

residents, with a Gross Domestic Product (GDP) per capita of 24,585 Yuan (US \$3715) [22–24]. The sole medical insurance system in rural Hua County is the NCMS, which is a government-run, voluntary, community-based, cost-sharing medical insurance programme. This programme has covered 99% of the local rural residents since 2010 [25].

Shantou City is located in eastern Guangdong Province and covers 797 square miles (Supplementary Figure 1). It consists of six districts (Chaonan, Chaoyang, Jinping, Longhu, Haojiang, and Chenghai) and one county (Nanao). In 2018, Shantou City was home to 5.62 million permanent residents, with a per capita GDP of 44,672 Yuan (US \$6751) [22,26]. The medical insurance systems in Shantou include UEBMI, URBMI, and NCMS, and the overall rate of enrollment has exceeded 90% in recent years.

Source of claims data

We extracted the medical claims data from Jan 1, 2012 to Jun 30, 2019 from the NCMS system in Hua County and the UEBMI, URBMI, and NCMS systems in Shantou City. The key variables included but were not limited to gender, year of birth, admission date, name of the hospital where the diagnosis was established, text-based diagnosis, International Classification of Diseases (tenth revision, ICD-10) code, and reimbursement date. All of the hospitals of record were classified as “grade 3”, “grade 2”, or “grade 1 or unrated” in descending order of their qualification for cancer diagnosis, according to the official database of the National Health Commission of the People’s Republic of China (zgcx.nhc.gov.cn:9090/unit/index).

Data security and confidentiality

Data security and confidentiality are crucial preconditions for the use of big data in healthcare. We established rigorous procedures in this study to guarantee security of the data over the course of data extraction, storage, processing, and statistical analysis. To secure protection of privacy, we masked individual-level identifiable personal information including name, identity (ID) number, medical insurance card number, telephone number, and home address. We then generated a unique, encrypted hexadecimal identification code for each patient as a label to further ensure elimination of duplicate diagnostic records.

Identification of incident cancer cases

A window of at least two years was set to exclude prevalent cancer cases by washing out repeated diagnoses emerging in the following years, as the cancer incidence rates in 2018 remained stable when the time window was reset from two years (2016–2017) to six years (2012–2017) (Supplementary Figure 2). We also used the records from the first six months of 2019 to avoid missing claims records due to reimbursement delay, based on the fact that the proportion of reimbursement delay over six months has been extremely low (<2.5%) in this study (Supplementary Figure 3).

We defined an incident cancer case on the basis of the first definite cancer-related claims record in each given case, with the corresponding date of admission as the incidence date, according to the recommendation of the European Network of Cancer Registries [15]. Text-based diagnosis and ICD-10 codes were used to ascertain cancer diagnoses.

The procedure for identifying incident cancer cases in Hua County and Shantou City is shown in Fig. 1. After data masking, the claims records which represented cancer-related diagnoses were exported from the medical insurance systems in these two regions once they met any of the following criteria: a. ICD-10 codes ranged from C00 to C97; b. Text-based diagnoses contained cancer-related key words, including “cancer”, “carcinoma”, “sarcoma”, “malignant tumour”, “leukaemia”, “lymphoma”, “Franklin disease”, “Alpha heavy chain

disease”, etc. (We reviewed the Chinese version of the ICD-10 and constructed a dictionary which included all the key words indicating a cancer diagnosis). Of all these extracted records, those diagnosed in grade 1 or unrated hospitals (eg, township-level health centres) were excluded due to insufficient diagnostic qualification, which is consistent with the standard procedure of the National Central Cancer Registry of China. Cancer diagnoses were then determined using both ICD-10 codes and text-based diagnoses. In cases of records where any of the ICD-10 code or text-based diagnosis was missing, or the cancer site determined by ICD-10 code disagreed with that determined by text-based diagnosis, active manual evaluation was performed with group discussion among at least three trained investigators with a background in oncology and epidemiology. Of all the qualified claims records, ~3.5% needed to be confirmed manually, and in only ~0.3% did disagreement arise among the investigators, which were resolved by following the majority rule in group discussion. For all claims records of each patient, only the first definite cancer diagnosis at a specific site during the study period was retained. Data from the years 2012 and 2013 were removed in consideration of prevalent cases, and data within the period Jan 1, 2014 to Dec 31, 2018 were defined as incident cancer diagnoses in this study. Multiple primary cancer cases were determined after excluding extension, recurrence or metastasis, which conforms to the international rules for multiple primary cancers [27].

Statistical analysis

For cancers from all sites, crude incidence rates (CIR), age-specific incidence rates, and age-standardised incidence rates (ASR) were calculated by gender in Hua County and Shantou City in 2018 using the corresponding insured population (the list of insured individuals was updated annually) as the denominator (Fig. 2, Supplementary Figure 4). Calculation of ASRs was based on the China standard population in 2000 (ASR China) and the Segi standard population (ASR World). Stata (version 15.0) was used for data cleaning, cancer case extraction, and calculation of incidence rates.

(1) Crude incidence rate per 100,000=

$$\frac{\text{New cancer cases in a given year}}{\text{Insured population in the same year}} \times 100,000$$

(2) Age-specific incidence rate per 100,000=

$$\frac{\text{New cancer cases in a given age group}}{\text{Insured population in the same age group}} \times 100,000$$

(3) Age-standardised incidence rate per 100,000=

$$\frac{\sum \text{Age-specific rate} \times \text{Standard population in corresponding age group}}{\sum \text{Standard population}} \times 100,000$$

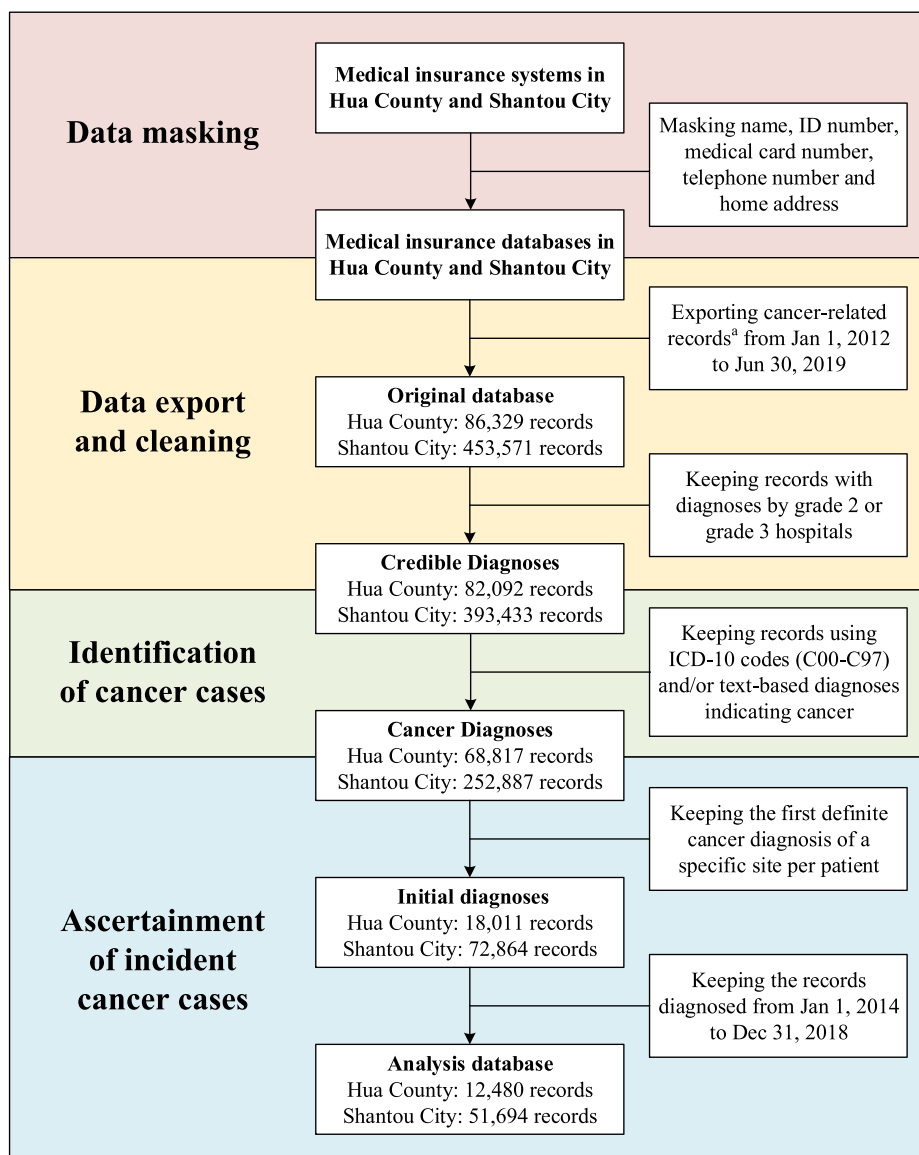
Ethics committee approval

This study was approved by the Institutional Review Board of the Peking University School of Oncology, China.

Result

Cancer patterns

In 2018, the age-standardised (Segi population) incidence rates for cancer in Hua County and Shantou City were 167.39/100,000 (164.79/100,000 in males, 176.72/100,000 in females) and 159.78/100,000 (179.87/100,000 in males, 142.05/100,000 in females) respectively. In Hua County, the most commonly diagnosed cancer in males was lung cancer, followed by oesophageal cancer, stomach cancer, liver cancer, and colorectal cancer. The most commonly diagnosed cancer in females was breast cancer, followed by lung cancer, cervical cancer, oesophageal cancer, and thyroid cancer. In Shantou



^aRecords were identified as cancer-related and exported if they met any of the following criteria: a. ICD-10 codes ranged from C00 to C97; b. text-based diagnoses contained cancer-related key words such as “cancer”, “carcinoma”, “sarcoma”, “malignant tumour”, “leukaemia”, “lymphoma”, “Franklin disease”, “Alpha heavy chain disease”, etc.

Fig. 1. The procedure for capturing new cancer cases from medical insurance systems in Hua County, Henan Province and Shantou City, Guangdong Province.

^aRecords were identified as cancer-related and exported if they met any of the following criteria: a. ICD-10 codes ranged from C00 to C97; b. text-based diagnoses contained cancer-related key words such as “cancer”, “carcinoma”, “sarcoma”, “malignant tumour”, “leukaemia”, “lymphoma”, “Franklin disease”, “Alpha heavy chain disease”, etc.

City, the five most common sites of cancer in males were lung, colorectum, oesophagus, liver, and stomach; and breast, colorectum, lung, cervix, and thyroid, in females (Table 1). In both Hua County and Shantou City, the age distribution of incidence rates for common cancers could in general be characterized as “left skewed distribution” which peaks in the elderly (such as oesophageal cancer, lung cancer, stomach cancer, and colorectal cancer), and “symmetric peak distribution” which peaks in middle age (such as breast cancer and cervical cancer) (Supplementary Figure 5, 6).

In Hua County, incidence rates from 2014 to 2018 decreased for oesophageal cancer in men, and increased for cancer from all sites in men, and for breast cancer, cervical cancer, thyroid cancer, and cancer from all sites in women (Supplementary Figure 7). In particular, we observed a sharp increase in the incidence of female breast cancer from 2015 to 2016 (Supplementary Figure 7). In Shantou City, a trend of increase in incidence rates was seen for lung cancer, colorectal

cancer, and liver cancer in men as well as colorectal cancer in women, whereas a decreasing trend was observed for oesophageal cancer and stomach cancer in men (Supplementary Figure 8).

Oesophageal cancer

In Hua County, the incidence of oesophageal cancer ranked first in 2014–2015 and third in 2016–2018 in both sexes combined (Supplementary Table 1a). The ASR world for oesophageal cancer in 2018 was 25.95/100,000 (34.78/100,000 in males, ranked second; 18.72/100,000 in females, ranked fourth) (Table 1a). By contrast, in Shantou City oesophageal cancer ranked fifth throughout the study period in both sexes combined except 2017 (when it ranked fourth), and the ASR world for oesophageal cancer was 11.43/100,000 (18.50/100,000 in males, ranked third; 4.79/100,000 in females, ranked eighth) in 2018 (Table 1b, Supplementary Table 1a).

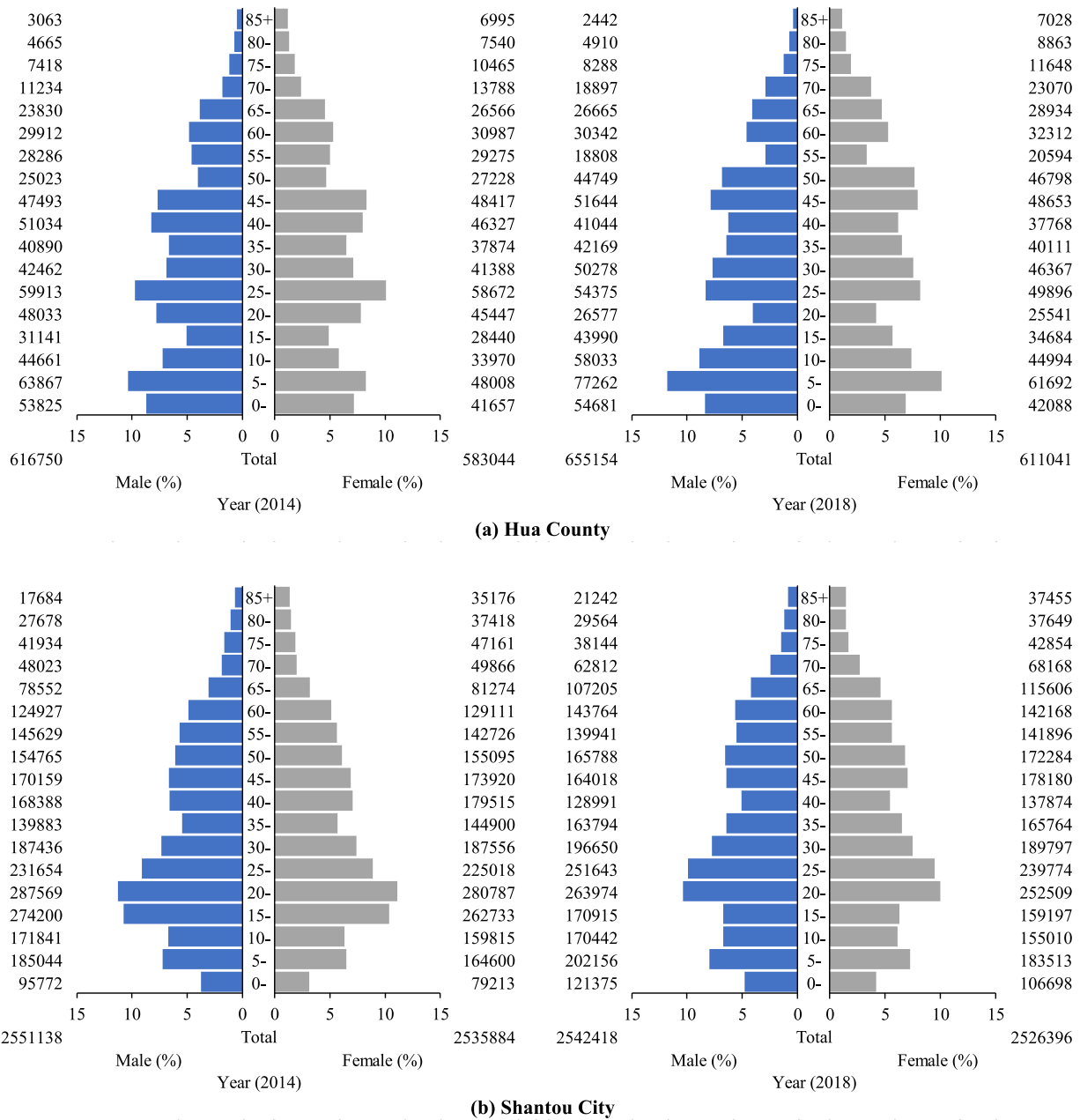


Fig. 2. Population pyramid of the insured population in Hua County, Henan Province and Shantou City, Guangdong Province, China, 2014 and 2018.

The ASR World of oesophageal cancer in Shantou City was only 44% of that in Hua County. However, distinct disparity in geographical pattern in the incidence of oesophageal cancer within Shantou City was observed. In 2018, Nanao, which is a part of Shantou City, was home to only 1.4% of the population in Shantou City, but had the highest incidence of oesophageal cancer (ASR World 36.39/100,000). This was more than twice as high as that in adjacent Chenghai, and six times as high as that in Chaonan, which is the district most distant from Nanao, and was even higher than that in Hua County (Fig. 3, Supplementary Table 2).

We further compared the age and gender distribution of the incidence of oesophageal cancer in Hua County and Shantou City. The incidence rate was relatively low in individuals of less than 50 years, and increased in ages over 50 in both of these two regions. However, differences in gender-specific risk of incident oesophageal cancer were observed in these two regions. In Hua County, the male-to-female ASR

World ratio for oesophageal cancer was only 1.69, which was much lower than that in Shantou City (ASR World ratio 4.02) (Fig. 4).

Discussion

Over the past 20 years, sentinel-based cancer surveillance has made a tremendous contribution in depicting the cancer burden in China, and has provided a crucial basis for subsequent cancer-related research and public health strategies. It has been an optimal strategy for maintaining a cancer registration system taking into consideration huge population and resource availability in China. However, based on full population coverage by basic medical insurance systems and the national-level integration of these systems in China, coupled with the advent of the big-data era, it is time to reconsider cancer surveillance strategies in China. In this study, we established a standard operation procedure for MIS-CASS, which includes privacy

Table 1a
Cancer incidence in Hua County, Henan Province, China, 2018.

Site	ICD-10	Male and Female					Male				Female			
		Cases	Proportion (%)	CIR (1/10 ⁵)	ASR China (1/10 ⁵)	ASR World (1/10 ⁵)	Cases	CIR (1/10 ⁵)	ASR China (1/10 ⁵)	ASR World (1/10 ⁵)	Cases	CIR (1/10 ⁵)	ASR China (1/10 ⁵)	ASR World (1/10 ⁵)
Lip, oral cavity, & pharynx ^a	C00-C10, C12-C14	26	0.93	2.05	1.52	1.52	12	1.83	1.49	1.56	14	2.29	1.55	1.47
Nasopharynx	C11	7	0.25	0.55	0.49	0.42	4	0.61	0.46	0.47	3	0.49	0.54	0.38
Oesophagus	C15	474	17.04	37.43	25.58	25.95	277	42.28	32.01	34.78	197	32.24	19.92	18.72
Stomach	C16	218	7.84	17.22	12.45	12.34	167	25.49	19.25	20.54	51	8.35	6.32	5.89
Colorectum	C18-C21	181	6.51	14.29	10.39	10.14	86	13.13	10.76	10.77	95	15.55	10.01	9.47
Liver	C22	160	5.75	12.64	9.37	9.51	109	16.64	13.11	14.04	51	8.35	5.93	5.72
Gallbladder	C23-C24	55	1.98	4.34	2.93	2.80	28	4.27	3.14	3.14	27	4.42	2.75	2.56
Pancreas	C25	56	2.01	4.42	3.16	3.17	30	4.58	3.63	3.77	26	4.26	2.74	2.61
Larynx	C32	17	0.61	1.34	0.94	0.91	16	2.44	1.86	1.89	1	0.16	0.09	0.09
Lung	C33-C34	500	17.98	39.49	28.51	28.58	307	46.86	36.14	38.59	193	31.59	21.68	20.57
Other thoracic organs	C37-C38	9	0.32	0.71	0.52	0.52	4	0.61	0.51	0.52	5	0.82	0.52	0.49
Bone	C40-C41	17	0.61	1.34	1.24	1.25	9	1.37	1.39	1.46	8	1.31	1.09	1.03
Melanoma of the skin	C43	11	0.40	0.87	0.76	0.77	4	0.61	0.61	0.61	7	1.15	0.91	0.93
Breast	C50	296	10.64	23.38	22.44	20.05	1	0.15	0.10	0.11	295	48.28	45.35	40.09
Cervix	C53	148	5.32	11.69	10.53	9.64	NA	NA	NA	NA	148	24.22	21.01	18.99
Uterus	C54-C55	63	2.27	4.98	4.30	4.26	NA	NA	NA	NA	63	10.31	8.47	8.40
Ovary	C56	60	2.16	4.74	4.23	3.91	NA	NA	NA	NA	60	9.82	8.55	7.81
Prostate	C61	30	1.08	2.37	1.64	1.58	30	4.58	3.42	3.69	NA	NA	NA	NA
Testis	C62	4	0.14	0.32	0.34	0.32	4	0.61	0.65	0.61	NA	NA	NA	NA
Kidney	C64-C66, C68	35	1.26	2.76	2.18	2.13	27	4.12	3.26	3.25	8	1.31	1.12	1.13
Bladder	C67	24	0.86	1.90	1.41	1.40	17	2.59	2.07	2.35	7	1.15	0.85	0.77
Brain, CNS	C70-C72	21	0.76	1.66	1.44	1.41	7	1.07	1.14	1.14	14	2.29	1.73	1.63
Thyroid	C73	143	5.14	11.29	10.11	9.41	26	3.97	4.02	3.73	117	19.15	16.23	15.04
Lymphoma	C81-C85, C88, C90, C96	82	2.95	6.48	5.18	5.34	52	7.94	6.53	6.95	30	4.91	3.88	3.83
Leukaemia	C91-C95	67	2.41	5.29	4.90	5.01	37	5.65	5.42	5.74	30	4.91	4.34	4.12
All other sites	A_0	77	2.77	6.08	4.70	5.08	30	4.58	4.19	5.09	47	7.69	5.24	4.97
All sites	ALL	2781	100.00	219.63	171.25	167.39	1284	195.98	155.14	164.79	1497	244.99	190.83	176.72

^a Nasopharyngeal cancer is excluded. ICD-10=International Classification of Diseases, tenth revision. CIR=crude incidence rate. ASR=age-standardised incidence rate. NA=not applicable.

Table 1b
Cancer incidence in Shantou City, Guangdong Province, China, 2018.

Site	ICD-10	Male and Female					Male				Female			
		Cases	Proportion (%)	CIR (1/10 ⁵)	ASR China (1/10 ⁵)	ASR World (1/10 ⁵)	Cases	CIR (1/10 ⁵)	ASR China (1/10 ⁵)	ASR World (1/10 ⁵)	Cases	CIR (1/10 ⁵)	ASR China (1/10 ⁵)	ASR World (1/10 ⁵)
Lip, oral cavity, & pharynx ^a	C00-C10, C12-C14	214	1.91	4.22	3.13	3.04	154	6.06	4.49	4.52	60	2.37	1.78	1.65
Nasopharynx	C11	301	2.68	5.94	4.94	4.65	222	8.73	7.39	7.05	79	3.13	2.56	2.36
Oesophagus	C15	855	7.62	16.87	11.16	11.43	660	25.96	17.24	18.50	195	7.72	5.07	4.79
Stomach	C16	872	7.77	17.20	12.01	11.84	626	24.62	16.79	17.61	246	9.74	7.15	6.48
Colorectum	C18-C21	1441	12.85	28.43	20.23	19.70	824	32.41	22.89	23.53	617	24.42	17.54	16.17
Liver	C22	776	6.92	15.31	11.39	11.06	610	23.99	18.06	18.07	166	6.57	4.75	4.48
Gallbladder	C23-C24	136	1.21	2.68	1.94	1.88	67	2.64	1.87	1.92	69	2.73	2.00	1.86
Pancreas	C25	149	1.33	2.94	2.05	2.02	77	3.03	2.13	2.21	72	2.85	2.00	1.83
Larynx	C32	104	0.93	2.05	1.43	1.44	100	3.93	2.71	2.85	4	0.16	0.13	0.11
Lung	C33-C34	2133	19.02	42.08	28.53	28.84	1583	62.26	41.46	44.26	550	21.77	15.28	14.54
Other thoracic organs	C37-C38	41	0.37	0.81	0.68	0.62	26	1.02	0.85	0.78	15	0.59	0.51	0.46
Bone	C40-C41	40	0.36	0.79	0.66	0.63	23	0.90	0.79	0.80	17	0.67	0.55	0.46
Melanoma of the skin	C43	19	0.17	0.37	0.32	0.29	7	0.28	0.20	0.21	12	0.47	0.44	0.36
Breast	C50	1179	10.51	23.26	19.68	18.58	10	0.39	0.26	0.28	1169	46.27	38.45	36.03
Cervix	C53	429	3.82	8.46	7.23	6.76	NA	NA	NA	NA	429	16.98	14.21	13.24
Uterus	C54-C55	256	2.28	5.05	4.02	3.91	NA	NA	NA	NA	256	10.13	7.93	7.65
Ovary	C56	138	1.23	2.72	2.21	2.08	NA	NA	NA	NA	138	5.46	4.40	4.07
Prostate	C61	169	1.51	3.33	2.07	1.97	169	6.65	3.95	4.29	NA	NA	NA	NA
Testis	C62	4	0.04	0.08	0.08	0.07	4	0.16	0.16	0.13	NA	NA	NA	NA
Kidney	C64-C66, C68	137	1.22	2.70	1.88	1.83	92	3.62	2.56	2.63	45	1.78	1.19	1.09
Bladder	C67	266	2.37	5.25	3.55	3.59	226	8.89	6.02	6.41	40	1.58	1.04	1.01
Brain, CNS	C70-C72	158	1.41	3.12	2.55	2.56	101	3.97	3.12	3.31	57	2.26	1.99	1.82
Thyroid	C73	334	2.98	6.59	5.96	5.39	75	2.95	2.72	2.41	259	10.25	9.10	8.26
Lymphoma	C81-C85, C88, C90, C96	335	2.99	6.61	5.10	4.95	206	8.10	6.18	6.24	129	5.11	4.02	3.76
Leukaemia	C91-C95	212	1.89	4.18	3.45	3.60	109	4.29	3.51	3.83	103	4.08	3.39	3.36
All other sites	A_O	519	4.63	10.24	7.24	7.05	286	11.25	7.67	8.03	233	9.22	6.78	6.20
All sites	ALL	11,217	100.00	221.29	163.47	159.78	6257	246.10	173.04	179.87	4960	196.33	152.24	142.05

^a Nasopharyngeal cancer is excluded. ICD-10=International Classification of Diseases, tenth revision. CIR=crude incidence rate. ASR=age-standardised incidence rate. NA=not applicable.



Fig. 3. Geographical pattern of oesophageal cancer age-standardised incidence rates (World standard, per 100,000) in Shantou City, Guangdong Province, China, 2018. ASR=age-standardised incidence rate.

protection, data extraction and cleaning, quality control, incident cancer case identification, statistical analysis, and reporting. Our study has demonstrated that the MIS-CASS can efficiently reflect local cancer burden, and thus has great potential to be utilized for cancer surveillance in China in the future.

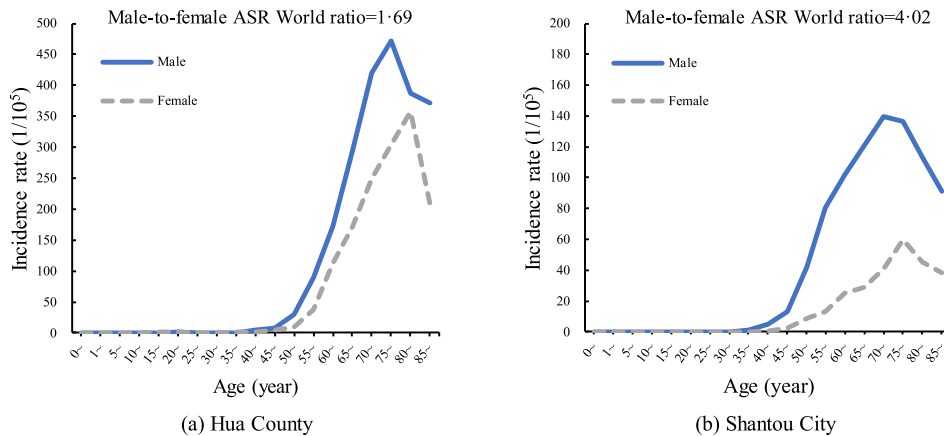
Over the past 10 years, we have carried out a series of population-level epidemiologic and etiologic studies on oesophageal cancer in China [20]. We selected Hua County and Shantou City for this study, because these areas have both previously been recognised as high-risk regions of oesophageal cancer, but have not been covered by cancer registries. In addition, Hua County in northern China is rural-dominant and Shantou City in southern China consists of both urban and rural residents. These two areas together thus represent an ideal sampling of the true current situation in China as regards health insurance systems.

Our MIS-CASS was based on individual-level hospitalization medical claims records. The first problem we faced was distinguishing incident cancer cases from the prevalent cases. For this study we set a washout time window of at least two years prior to the starting year of formal statistical analysis of cancer incidence, and all cancer patients who had claims records within this period were classified as prevalent cases and excluded from analysis. According to our

estimates, two years should be a sufficiently long period of time, as the overall cancer incidence rates in these two regions of interest were stable when the washout time window was set starting from 1 year (Supplementary Figure 2). However, we nevertheless recommend that the time window be as long as possible to eliminate the impact of prevalent cases to as great an extent as possible. In fact, in this study, we adopted a washout period of up to six years (2012–2017) to estimate the latest cancer specific incidence in these two areas for the year of 2018, which would make the impact of “prevalent cases” almost negligible.

In addition, cancer surveillance using claims data also has potential to miss incident cases due to reimbursement delay. We thus assessed the time length between admission date and reimbursement date for all the cancer patients, and found that the proportion of cases with a delay of over six months was extremely low (2.4%) (Supplementary Figure 3). This indicates that the MIS-CASS “yields results in real time” with only a six-month delay, which is far shorter than the delay in the current cancer surveillance system in China (three years).

We also evaluated the internal and external validity of the MIS-CASS. From 2014 to 2018, in Hua County and in Shantou City, the age distribution of the entire insured population across the study years



ASR=age-standardised incidence rate.

Fig. 4. Age-specific incidence rates of oesophageal cancer by gender in Hua County, Henan Province and Shantou City, Guangdong Province, China, 2014–2018. ASR=age-standardised incidence rate.

moved rightward stably as a series of “birth-cohorts” reflecting the extreme robustness of the “denominator” (Supplementary Figure 4). The year-by-year percentage changes in cancer incidence rates were mostly within 10%, denoting the stability of cancer incidence statistics over time (Supplementary Table 3) [5,6]. In terms of external validity, the profiles of the common cancer types observed in both of these two regions were nearly consistent with those in China overall [10]. At the same time, region-specific cancer patterns were also clearly observed, such as for oesophageal cancer in Hua County and colorectal cancer in Shantou City (Supplementary Table 4).

In addition to reporting routine cancer specific incidence statistics, there were also some interesting findings from an epidemiologic perspective based on the MIS-CASS.

First, over the course of many past decades, Hua County and Shantou City have been “understood” to be “high-risk” regions for oesophageal cancer. In this study, we demonstrated for the first time that Hua County is indeed a high-risk region for oesophageal cancer as expected. Despite the overall downward trend of incidence, the incidence of oesophageal cancer in Hua County ranked first before the year of 2015, and remained third from 2016 to 2018 (Supplementary Table 1). Moreover, the incidence was 2.1 times as high as the average national level according to the latest estimate for 2018 [28]. However, Shantou City as a whole was unexpectedly found not to be a high-risk region for oesophageal cancer, as the incidence rate ranked only as fifth, and the ASR world was very close to the national level for both men and women [28]. Nevertheless, further stratification analysis revealed distinct geographical pattern of oesophageal cancer incidence in Shantou City. Among its seven districts, Nanao County, which is an island located at the very east of Shantou City that is encircled by the sea, had the highest incidence of oesophageal cancer, with an ASR World 3.0 times as high as national rates which is consistent with previous studies [28–30], and 6.1 times as high as that of Chaonan district which has the lowest oesophageal cancer incidence. These findings would may provide crucial epidemiologic evidence for establishing prevention and control programmes for oesophageal cancer in local areas.

Second, we also observed oesophageal cancer demonstrates heterogeneity in incidence vis-a-vis gender in these two regions. In Shantou City, the risk of oesophageal cancer in men was notably higher than that in women (male-to-female ASR World ratio 4.02), and the ratio was higher than that at the national level (male-to-female ASR World ratio 2.81) [28]. However, in Hua County which is a high risk area in the famous “Taihang Mountain Region”, the gender difference was much smaller with a male-to-female ASR World ratio of 1.69. This is consistent with previous studies in Hua County and Linzhou, which is another well-known high-risk area for oesophageal cancer that is approximately 50 miles from Hua County [20,31]. This finding suggested that unlike Shantou City and most other regions in China, common risk factors for oesophageal cancer are shared by both men and women. That is, non-gender related environmental exposure and/or genetic susceptibility and so on were likely to have played a crucial role in the occurrence of oesophageal cancer in the Taihang Mountain high-risk areas.

Third, we observed a sharp increase in the all-site cancer incidence rate for females in Hua County in 2016, which mainly resulted from the contribution of population level screening programmes for breast cancer and cervical cancer (Supplementary Table 1b, 3, Supplementary Figure 7). After further investigation, we found that a series of countywide female breast and cervical cancer screening programmes had been initiated by the Hua County Maternal and Child Health Hospital together with other health institutions in 2015–2016, and these programmes endured over the remainder of the study years. This resulted in a 63% and 107% increase in breast cancer and cervical cancer incidence from 2015 to 2018.

Basic medical insurance systems in China are administered and operated by the central and local government, so that the MIS-CASS by its nature has great strengths. First, the MIS-CASS is a real

population level cancer surveillance system with >95% coverage on a national scale, which is distinct from health insurance systems in some western countries such as Medicare and the Ontario Health Insurance Plan [32,33]. Second, the original claims data of MIS-CASS is of high quality [17], as the diagnostic information is entered by professionals in qualified medical institutions and verified by officials from medical insurance managing departments. Third, the proportion of cancer cases missed by the MIS-CASS is low, based on our two previous individual level comparative evaluation studies [16–18]. This is presumed to be the result of economic interest and severity of cancer. Fourth, the MIS-CASS greatly shortens the reporting delay of cancer incidence from three years to six months due to management policies and “real-time” claims settlement in health insurance systems in China [5,34]. Fifth, the high quality and high standardization level of the original claims data in the MIS-CASS results in low cost for analysing and reporting cancer incidence in a large population. This system is therefore obviously less labor intensive and requires less time resource input than the current cancer registries in China.

However, some limitations of the MIS-CASS should also be noted. First, non-inpatient claims records for cancer cases where the patient was never hospitalized would not be included in the MIS-CASS. Despite the fact that almost all cancer patients in China are diagnosed and treated in the form of “hospitalization”, we nonetheless propose addition of outpatient data and even emergency data to reduce possible underestimation as much as possible. Secondly, cancer-related information contained in the claims database is relatively limited at present. Integration of more diagnosis-related information (such as ICD-O-3 codes including morphology and topography) into the database would further increase the accuracy of this system for ascertaining cancer cases, and expand its applicability.

As indicated in the 13th Five-Year Plan (2016–2020) in China and “Healthy China 2030” Planning Outline, great efforts will be made to promote the integration, sharing, mining, and application of big data in different health-related areas. In 2017, the National Health and Family Planning Commission proposed accelerating the informatization of surveillance and management of major chronic diseases at the national level. This study has established the value of MIS-CASS for applying big data in medical insurance systems to cancer surveillance, demonstrating that the MIS-CASS can efficiently reflect local cancer burden and guarantee data security and privacy. We therefore recommend that the MIS-CASS be adopted on a larger scale, especially in areas with high quality China cancer registries, so that parallel comparative studies could be conducted to further validate the completeness and accuracy of the MIS-CASS. This will provide new insights into the reform and development of cancer surveillance strategy in China.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the National Key R&D Program of China (2016YFC0901404), the Digestive Medical Coordinated Development Center of Beijing Municipal Administration of Hospitals (XXZ0204), the Sanming Project of Shenzhen (SZSM201612061), and the Shantou Science and Technology Bureau (190829105556145, 180918114960704).

We gratefully acknowledge all the following experts and collaborators for their assistance in this study: Fu Gao from Chinese Center for Disease Control and Prevention; Dongfeng Gu from Fuwai Hospital, Chinese Academy of Medical Sciences; Qimin Zhan from Peking University/Peking University Health Science Center; Lingzhi Kong and Liangyou Wu from Bureau of Disease Prevention and Control, National Health Commission of the People's Republic of China;

Qingyue Meng and Siyan Zhan from School of Public Health, Peking University; Kaifeng Pan from Peking University Cancer Hospital and Institute; Hanran Chen from the Health Bureau of Shantou, Guangdong Province; Junhui Bian and Fen Yao from Shantou University Medical College; Jinwu Zheng from the People's Government of Hua County, Henan Province; Xiangdong Liu from the Health Commission of Hua County, Henan Province; Xiaopeng Zhao and Min Chen from the Social Insurance Fund Administration of Shantou, Guangdong Province; Fenglei Li and Shaojiang Lv from Hua County People's Hospital, Henan Province.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.eclinm.2020.100312](https://doi.org/10.1016/j.eclinm.2020.100312).

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 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.
- International Agency for Research on Cancer. Estimated number of incident cases from 2018 to 2040, all cancers, both sexes, all ages. https://gco.iarc.fr/tomorrow/graphic-isotype?type=0&population=900&mode=population&sex=0&cancer=39&age_group=value&apc_male=0&apc_female=0#collapse-group-0-3 [Accessed 31 October 2019].
- Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. Cancer registration: principles and methods. Lyon: International Agency for Research on Cancer; 1991.
- Bray F, Znaor A, Cueva P, et al. Planning and developing population-based cancer registration in low-and middle-income settings. Lyon: International Agency for Research on Cancer; 2014.
- Forman D, Bray F, Brewster DH, et al. Cancer Incidence in Five Continents, Vol. X. Lyon: International Agency for Research on Cancer; 2014.
- He J, Chen W. China Cancer Registry Annual Report 2017. Beijing: People's Medical Publishing House; 2018.
- Bray F, Colombet M, Mery L, et al. Cancer Incidence in Five Continents, Vol. XI (electronic version). Lyon: International Agency for Research on Cancer; 2017 <http://ci5.iarc.fr> [Accessed 10 January 2020].
- National Bureau of Statistics of China. China Statistical Yearbook 2018. China Statistics Press; 2018 <http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm> [Accessed 10 January 2019].
- Chen W, Zheng R, Zhang S, et al. Cancer incidence and mortality in China, 2013. *Cancer Lett* 2017;401:63–71.
- Zheng R, Sun K, Zhang S, et al. Report of cancer epidemiology in China, 2015 [in Chinese]. *Chin J Oncol* 2019;41:19–28.
- Nattinger AB, Laud PW, Bajorunaite R, Sparapani RA, Freeman JL. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. *Health Serv Res* 2004;39:1733–49.
- Cooper GS, Yuan Z, Jethva RN, Rimm AA. Use of Medicare claims data to measure county-level variation in breast carcinoma incidence and mammography rates. *Cancer Detect Prev* 2002;26:197–202.
- National Health and Family Planning Commission. China's basic medical insurance covers 1.35 bln people. 2018. http://www.china.org.cn/china/Off_the_Wire/2018-02/12/content_50500993.htm [Accessed 29 November 2018].
- State Council. China establishes state medical insurance administration. 2018. http://english.gov.cn/state_council/state_councilors/2018/06/01/content_281476167596608.htm [Accessed 4 January 2019].
- Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer (Oxf Engl: 1990)* 2009;45:747–55.
- Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *Eur J Cancer (Oxf Engl: 1990)* 2009;45:756–64.
- Shi C, Liu M, Liu Z, et al. Using health insurance reimbursement data to identify incident cancer cases. *J Clin Epidemiol* 2019;114:141–9.
- Tian H, Xu R, Li F, et al. Identification of cancer patients using claims data from health insurance systems: A real-world comparative study. *Chin J Cancer Res* 2019;31:699–706.
- Seo HJ, Oh IH, Yoon SJ. A comparison of the cancer incidence rates between the national cancer registry and insurance claims data in Korea. *Asian Pac J Cancer Prev* 2012;13:6163–8.
- 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.
- Liu X, Zhang M, Ying S, et al. Genetic alterations in esophageal tissues from squamous dysplasia to carcinoma. *Gastroenterology* 2017;153:166–77.
- 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 January 2020].
- Hua County Bureau of Statistics. Statistical Bulletin of Economic and Social Development of Hua County in 2017 [in Chinese]. 2018. http://www.ha.stats.gov.cn/sitesources/hntj/page_pc/tjfw/tjgb/szgxgb/article677fc6a201e54965b8ece848a48f125c.html [Accessed 17 December 2018].
- Hua County Bureau of Statistics. Statistical Bulletin of Economic and Social Development of Hua County in 2018 [in Chinese]. 2019. http://www.ha.stats.gov.cn/sitesources/hntj/page_pc/tjfw/tjgb/szgxgb/article9ab7a3f8c4874c34a307dc3d4e6c83a.html [Accessed 17 January 2020].
- Wang J, Zhou H, Lei Y, Wang X. Financial protection under the New Rural Cooperative Medical Schemes in China. *Med Care* 2012;50:700–4.
- Shantou City Bureau of Statistics. Shantou statistical yearbook. Shantou Statistical Yearbook 2019 [in Chinese]. Shantou City Bureau of Statistics; 2019. [in Chinese]. 2019. <https://www.shantou.gov.cn/tjj/2019nj/201912/9ba321a2b98a4f32ba2881fb968497b3/files/bd87016243f24c14bd107b7b2dbc0e0b.pdf> [Accessed 17 January 2020].
- Demaret E, Ferlay J, Parkin M, Tyczynski J, Whelan S. International rules for multiple primary cancers. *Asian Pac J Cancer Prev* 2005;6:104–6.
- Chen W, Sun K, Zheng R, et al. Cancer incidence and mortality in China, 2014. *Chin J Cancer Res* 2018;30:1–12.
- Ke L. Mortality and incidence trends from esophagus cancer in selected geographic areas of China circa 1970–90. *Int J Cancer* 2002;102:271–4.
- Su M, Liu M, Tian DP, et al. Temporal trends of esophageal cancer during 1995–2004 in Nanao island, an extremely high-risk area in China. *Eur J Epidemiol* 2007;22:43–8.
- Tran GD, Sun X, Abnet CC, et al. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int J Cancer* 2005;113:456–63.
- Diop M, Strumpf EC, Datta GD. Measuring colorectal cancer incidence: the performance of an algorithm using administrative health data. *BMC Med Res Methodol* 2018;18:38.
- Rabeneck L, Paszat LF. A population-based estimate of the extent of colorectal cancer screening in Ontario. *Am J Gastroenterol* 2004;99:1141.
- Barber SL, Yao L. Health insurance systems in China: a briefing note. World Health Report (2010) Background Paper, No. 37. Geneva: World Health Organization; 2010.