From Department of Molecular Medicine and Surgery
Karolinska Institutet, Stockholm, Sweden

ESOPHAGEAL SQUAMOUS CELL CARCINOMA - OPPORTUNITIES FOR PREVENTION

Qiaoli Wang, M.D.

王巧丽

Stockholm 2020

All previously published papers were reproduced with permission from the publisher.
Cover design “The first female physician-scientist Merit Ptah” by the author of the book
Illustration by Yi Zheng
Published by Karolinska Institutet
Printed by E-print AB 2020

ISBN 978-91-7831-644-1

Esophageal Squamous Cell Carcinoma - Opportunities for Prevention

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Qiaoli Wang

Principal Supervisor:
Professor Jesper Lagergren
Karolinska Institutet
Department of Molecular Medicine and Surgery
Upper Gastrointestinal Surgery

Co-supervisor:
Assistant Professor Shaohua Xie
Karolinska Institutet
Department of Molecular Medicine and Surgery
Upper Gastrointestinal Surgery

Opponent:
Associate Professor Deirdre Cronin Fenton
Aarhus University
Department of Clinical Epidemiology

Examination Board:
Professor Kamila Czene
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Associate Professor Jakob Hedberg
Uppsala University
Department of Surgical Sciences
Upper Abdominal Surgery

Associate Professor Simon Ekman
Karolinska Institutet
Department of Oncology and Pathology

“The only purpose of science is to ease the hardship of human existence.”

- Galileo Galilei

To my beloved family

致我深爱的家人

ABSTRACT

Esophageal squamous cell carcinoma (ESCC) is the predominant histological subtype of esophageal cancer, a highly fatal malignant neoplasm. Most ESCC patients are diagnosed at a late stage when tumors are unresectable or have metastasized. The median survival is less than one year, highlighting a great need for early diagnosis and preventive measures. The overall aim of the thesis is to provide a better knowledge of how ESCC can be prevented.

Study I is an incidence study based on the data collected directly from 30 cancer registries in 20 countries for 1970-2015. Cross-sectional analyses of the year 2012 showed that the highest incidence rate of ESCC was in Japan (9.7/100,000 person-years). The incidence had decreased continuously in men globally but slightly increased in women from Japan, the Netherlands, New Zealand, Norway, and Switzerland. Age-period-cohort analyses revealed that birth-cohort effects were strong determinants for the incidence trends.

Study II is a systematic review and meta-analysis assessing tobacco smoking cessation and risk of ESCC. We found 41 relevant studies from 15,009 publications. The random-effects model was applied to estimate pooled risk ratios (RRs) with 95% confidence intervals (CIs). Compared with current smokers, those who stopped smoking 5-9 years earlier had a decreased risk of ESCC (RR 0.59, 95% CI 0.47-0.75), and the risk reduction was stronger in those who had stopped smoking 10-20 years earlier and reached almost the level of nonsmokers in those who had stopped smoking >20 years ago (RR 0.34, 95% CI 0.25-0.47). Thus, smoking cessation seems to reduce the risk of ESCC strongly.

Study III is a Swedish nationwide population-based cohort study in 2005-2015. Among 8.4 million participants, we identified 411,603 metformin users for the study who were compared with ten times as many age- and sex-matched nonusers of metformin. Hazard ratios (HRs) were estimated using multivariable cause-specific proportional hazards modeling. The ESCC incidence rate was 3.5/100,000 person-years in metformin users and 5.3/100,000 person-years in nonusers. Compared with nonusers, ever-users of metformin had an HR of 0.68 (95% CI 0.54-0.85) and new metformin users had an HR of 0.44 (95% CI 0.28-0.64). Thus, metformin use may prevent ESCC.

Study IV is a Swedish nationwide case-control study in 1995-1997, including 167 ESCC cases and 820 randomly selected control participants who were all personally interviewed. A risk prediction model was developed based on the predictors: age, sex, smoking, alcohol use, education, duration of the partnership, and childhood residence. The area under the receiver operating characteristic curve was 0.81 (95% CI 0.77-0.84). With these predictors, an individual's absolute risk of ESCC within the next five years can be predicted.

In summary, this thesis indicates that ESCC remains common cancer globally, that prevention of this tumor may be possible by smoking cessation and metformin use, and those high-risk individuals can be identified by a risk prediction model, which may enable earlier tumor detection.

LIST OF SCIENTIFIC PAPERS

This thesis is based on the following four papers, which are referred to in the text by their Roman numerals (I-IV).

I. Wang QL, Xie SH, Wahlin K, Lagergren J.
Global time trends in the incidence of esophageal squamous cell carcinoma.
Clinical Epidemiology 2018;10:717-728.

II. Wang QL, Xie SH, Li WT, Lagergren J.
Smoking cessation and risk of esophageal cancer by histological type: systematic review and meta-analysis.
Journal of the National Cancer Institute 2017;109(12).

III. Wang QL, Santoni G, Ness-Jensen E, Lagergren J, Xie SH.
Association between metformin use and risk of esophageal squamous cell carcinoma in a population-based cohort study.
American Journal of Gastroenterology 2020;115(1):73-78.

IV. Wang QL, Lagergren J, Xie SH.
Prediction of individuals at high absolute risk of esophageal squamous cell carcinoma.
Gastrointestinal Endoscopy 2019;89(4):726-732.

CONTENTS

LIST OF ABBREVIATIONS

AAPC	Average Annual Percentage Change
APC	Annual Percentage Change
ASR	Age Standardized Rate
ATC	Anatomical Therapeutic Chemical
AUC	Area Under the Curve
CI	Confidence Interval
DDD	Defined Daily Dose
EAC	Esophageal Adenocarcinoma
EMR	Endoscopic Mucosal Resection
ESCC	Esophageal Squamous Cell Carcinoma
ESD	Endoscopic Submucosal Dissection
HR	Hazard Ratio
HUNT	Nord-Trøndelag Health Study
IARC	International Agency for Research on Cancer
ICD	International Classification of Diseases
ICD-O-3	3rd version of International Classification of Disease for Oncology
NSAIDS	Non-Steroidal Anti-Inflammatory Drugs
OR	Odds Ratio
RR	Risk Ratio
SD	Standard Deviation
SECC	Swedish Esophageal and Cardia Cancer study
SEER	The Surveillance, Epidemiology and End Results
SPREDH	Swedish Prescribed Drugs and Health cohort
WHO	World Health Organization

1 INTRODUCTION

Esophageal cancer has an aggressive nature and poor prognosis. In 2018, esophageal cancer was the sixth leading cause of cancer-related deaths and the seventh most common cancer worldwide. More than 572,000 new esophageal cancer cases and 508,000 deaths were estimated that year. Esophageal squamous cell carcinoma (ESCC) is the predominant (87%) histological type of esophageal cancer globally. Early-stage ESCC patients are usually asymptomatic, and more than half of the patients first present with unresectable or metastatic disease when the prognosis is dismal. Better outcomes in ESCC patients are strongly related to diagnosis at an early stage, and in tumor stage I, a five-year survival rate of 80-90% is expected. Despite many efforts to advance the treatments during the last decades, the prognosis is still poor, with only a 10-20% overall five-year survival rate.

This thesis contains four studies, focusing on the possibilities of prevention and early detection of ESCC. In the first study, we examined the ESCC incidence worldwide and analyzed its time trends using the age-period-cohort method. In the second study, we assessed the influence of smoking cessation on the risk of ESCC. The third study investigated how metformin use (chemoprevention) influences the risk of ESCC. Finally, in the fourth study, we developed a risk prediction model for ESCC to help estimate an individual's absolute risk and facilitate early tumor detection.

2 BACKGROUND

2.1 ESOPHAGUS ANATOMY AND ESOPHAGEAL CANCER

2.1.1 Structure and function of the esophagus

The adult human esophagus is an 18-25cm long and relatively straight muscular tube through which food passes from the pharynx to the stomach. Anatomically, starting from the pharyngoesophageal junction in the neck (C5-6 vertebral level), it descends posteriorly to the trachea and anteriorly to the spinal column through the mediastinum.[1] The esophagus further traverses the diaphragm at the hiatus (T10 vertebral level) and extends through the gastroesophageal junction to the cardia of the stomach at the T11 vertebral level. Esophageal sphincters are located at both the upper and lower ends of the esophagus and prevent food backflow. Peristalsis is maintained by contractions of the muscles in the esophageal wall, which is composed of four layers: mucosa, submucosa, muscularis propria, and adventitia (but without serosa layer).[2] The esophageal lumen is normally lined with squamous epithelium and shows as a smooth and pale pink tube under endoscopy, with visible submucosal blood vessels.

2.1.2 Esophageal cancer

2.1.2.1 History

Although without specific mention of esophageal cancer, the earliest description of this disease could trace back to around 3000 BC, as "a gaping wound of the throat penetrating the gullet" from the *Smith Surgical Papyrus* in Egypt.[3] Over 2000 years ago, a clear description of esophageal cancer appeared in China, referred to as "Ye Ge", which means dysphagia and belching. In the ancient Chinese medical literature, this tumor was believed to be a distinct disease caused by "heavy indulgence of heated liquors", more commonly seen in the elderly with dysphagia and having a poor prognosis with less than one-year survival after diagnosis.[4] Centuries later, the Greek physician Galen noticed fleshy growths obstructing the esophagus that caused cachexia and death.[5] However, it was not until the 11th century that the Arab physician Avicenna described the esophageal tumor as one of the causes of dysphagia.[5]

During the late renaissance, physicians started documenting esophageal cancer cases to collect more medical knowledge. In 1543, the Flemish anatomist Vesalius published the first influential anatomy book *De humani corporis fabrica,* clearly describing the anatomy of the esophagus. In the following centuries, the invention of microscopy and the advent of pathology prompted a deeper understanding of esophageal cancer. In the 19th century, several studies started to link esophageal cancer with risk factors, such as heavy alcohol drinking, not the least of absinthe, the most popular alcoholic beverage during that time.[6 7] In 1868, the development and application of the first esophagoscopy by Adolf Kussmaul enabled direct observation of a living esophagus and pathological diagnostics. The following decades witnessed a great progress in the surgical treatments of esophageal cancer, including the first resection of the cervical esophagus in 1877, the first esophagectomy in 1913, and

esophagectomy with intrathoracic esophagogastric anastomosis in 1929.[5] Surgery was considered as the only curative treatment option for esophageal cancer patients until mid-1980s when the concept of "multimodality treatment" was introduced by Vincenz Czerny, adding radiotherapy and chemotherapy to the surgical treatment. Nowadays, new therapeutic methods are practiced in clinics, such as minimal invasive esophagectomy, targeted therapy, and immunotherapy.

2.1.2.2 Histopathology

There are a few rare histological types of malignant esophageal tumors, e.g., mucoepidermoid carcinoma, endocrine tumors, gastrointestinal stromal cell tumors, small cell carcinoma, lymphoma, and melanoma, but the two dominating histological subtypes are ESCC and esophageal adenocarcinoma (EAC). These two are different entities, considering disease distribution, demographics, etiology, pathogenesis, treatment, and prognosis.[8] ESCC occurs anywhere in the esophagus but is more commonly seen in the distal and middle third of the esophagus, while EAC occurs in the distal segment. ESCC develops from the native squamous epithelial cells lining the esophagus that might result from local injury and inflammation, hyperplasia, and dysplasia, while EAC arises through the replacement of squamous epithelium to columnar epithelium by intestinal metaplasia (entitled Barrett's esophagus).[8] Well-differentiated ESCC usually contains keratinocytes, accompanied by intercellular bridges and keratin, whereas poorly-differentiated ESCC is characterized by the presence of intraepithelial neoplasia, *in situ* lesions in adjacent squamous mucosa, and squamous-oriented infiltration. Differentiation grades of EAC are mainly determined by the amount of gland formation and the nuclear atypia.[8]

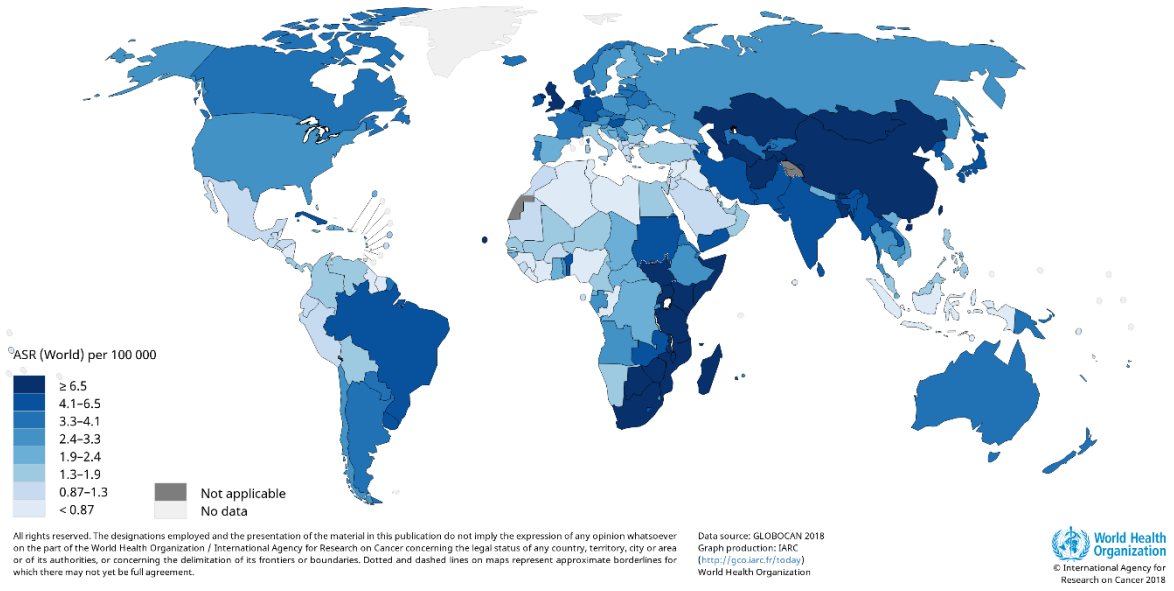


Figure 1. Estimated age-standardized incidence rate of esophageal cancer worldwide in 2018
Reproduced with permission from the International Agency for Research on Cancer (IARC).[9]

An overview of the incidence of esophageal cancer worldwide in 2018 is shown in Figure 1. Although a shift of in the incidence of ESCC has been observed in recent decades with an

increasing incidence of EAC and a decreasing incidence of ESCC in several developed countries, ESCC remains the predominant histological type worldwide, accounting for 87% of all esophageal cancer cases.[10] This thesis focuses on ESCC.

2.2 CLINICAL ASPECT OF ESOPHAGEAL SQUAMOUS CELL CARCINOMA

2.2.1 Diagnosis

2.2.1.1 Clinical symptoms

Typical clinical symptoms of ESCC usually present at a late stage when the tumor has already invaded more than 60% of the esophageal circumference and is clearly visible at endoscopy.[11 12] The most common symptoms when first diagnosed are progressive dysphagia occurring among 74% of cases, followed by progressive and involuntary weight loss among 57% of ESCC patients, and odynophagia (pain on swallowing) among 17% of cases.[13] Dysphagia could also be accompanied by radiated pain of the chest or back, and aspiration pneumonia. Typically, progressive and involuntary weight loss is reported and is an independent predictor of poor prognosis.[14] Other less common symptoms include coughing, hoarseness, dyspnea, and retrosternal pain, indicating the presence of locally advanced tumor growth and invasion of surrounding tissues and organs.[15] Some patients develop tracheoesophageal fistulas and hypercalcemia without osseous metastases.[11] Advanced ESCC usually metastasizes to the supraclavicular lymph nodes (including Virchow's node), liver, lungs, pleura, and skeleton.

2.2.1.2 Diagnosis and staging

Esophageal squamous dysplasia is regarded as the precursor of ESCC.[16 17] ESCC develops through multiple pathological alterations from the normal esophagus or low-grade intraepithelial neoplasia (mainly esophagitis or basal cell hyperplasia), to middle- or high-grade intraepithelial neoplasia (i.e., squamous dysplasia or cancer *in situ* without lamina propria invasion), and finally invasive carcinoma.[18] Depending on the individual ESCC risk pattern, patients with low-grade intraepithelial neoplasia are recommended regular follow-ups with endoscopy, while high-grade intraepithelial neoplasia is often treated with endoscopic mucosal resection (EMR) or endoscopic submucosal dissection (ESD).[19] Individuals with high-grade squamous dysplasia and cancer *in situ* have significantly higher risks of developing ESCC compared with low-grade intraepithelial neoplasia, indicating that screening for precursors of ESCC might be justified in high-risk populations.[16]

The standard diagnostic method of ESCC is white light esophagogastroduodenoscopy with biopsies for histopathologic examination and confirmation.[14 20] Recent research has found that the sensitivity of finding early lesions can be increased from 55% using white light endoscopy to 92% using the chromoendoscopy with Lugol's iodine, and the narrow-band imaging system could further improve the sensitivity to 97%.[21 22] For poorly differentiated

tumors, immunehistochemical staining is recommended by the World Health Organization (WHO) to distinguish ESCC from EAC and other histological types of esophageal malignancies or secondary tumors.[23] Following pathological diagnosis, accurate staging is crucial for selecting the most appropriate treatment, and for adequate assessment of the prognosis. Staging includes the use of separate or combined diagnostic methods such as contrast-enhanced computerized tomography, positron emission tomography, endoscopic ultrasound, and sometimes also diagnostic laparoscopy or thoracoscopy.[14 24] Specific meticulous examinations of the oral cavity, oropharynx, and hypopharynx, and trachea-bronchoscopy investigation are recommended for some ESCC patients depending on the results of the primary assessment.[23] The TNM staging system is commonly applied for ESCC staging worldwide, taking into account the depth of the primary lesion (T0-4), the involvement of lymph nodes (N0-3) and distant metastasis (M0-1).[14] The latest version of TNM staging (version 8) is shown in Supplementary Table 1, and the illustration of T and N category is shown below in Figure 2.

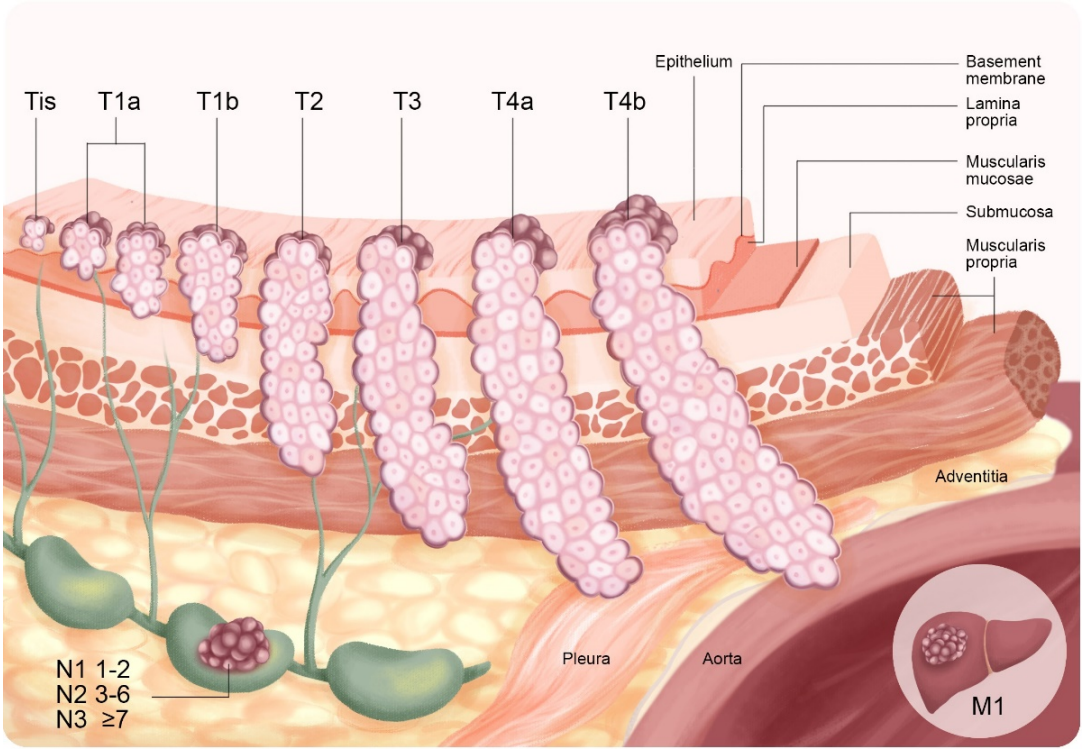


Figure 2. Illustration of TNM staging of esophageal cancer by Yi Zheng

2.2.2 Treatment

Various approaches have been used to treat ESCC, and the treatment differs mainly depending on the tumor stage and fitness of the patient.[19] Common therapies include endoscopic treatment for very early lesions, surgery (esophagectomy) with or without neoadjuvant and adjuvant therapy, definitive chemotherapy and/or radiotherapy, as well as palliative therapy (e.g., stenting, brachytherapy, chemotherapy, external radiotherapy). Novel therapies using immune checkpoint inhibitors are attempted in several ongoing clinical trials, showing some promising early results in ESCC patients.[25] A multidisciplinary assessment of

disease status and treatment options has been proven to improve the clinical decision-making and is therefore widely recommended.[23 26]

2.2.2.1 Endoscopic treatment

Endoscopic treatment is applicable in patients with early-stage ESCC (Tis and T1a) with no evidence of lymph node metastasis.[23] EMR or ESD, radiofrequency ablation or cryoablation therapy, and photodynamic therapy are the most commonly used methods.[27 28] Combination of EMR and radiofrequency ablation seems to be particularly effective in the prevention of cancer progression in patients with dysplasia.[27] ESD has a higher en bloc resection rate (97% *vs.* 49%) and a higher curative resection rate (92% *vs.* 53%) than EMR.[29 30] Endoscopic surveillance following endoscopic treatment is needed due to potential local tumor recurrence, metachronous ESCC, or associated head and neck squamous cell carcinoma.[31]

2.2.2.2 Curative surgery and neoadjuvant therapy

For most T1b tumors, esophagectomy is recommended because of the higher rates of lymph node involvement than T1a tumors, resulting in a higher risk of local or distant recurrence.[28] Moreover, esophagectomy is widely used in patients with T2N0 tumors and also in patients with early-stage tumors whenever endoscopic treatment has failed. Surgery is usually combined with neoadjuvant therapy for more locally advanced tumors.[32] Common surgical procedures are the thoracoabdominal or transhiatal approaches with open or minimally invasive techniques.[33] Lymphadenectomy is recommended for ESCC patients because of the high frequency of lymph node metastasis. However, evidence from large cohort studies indicates that a tailored and moderate lymphadenectomy is sufficient.[34 35] Neoadjuvant chemotherapy or neoadjuvant chemoradiotherapy is a standard therapy for locally advanced ESCC.[36] A landmark prospective randomized clinical trial (the JCOG9907 trial) focused specifically on ESCC and found a significantly decreased hazard ratio (HR) of five-year mortality by 0.73 (95% CI 0.54-0.99) in the neoadjuvant chemotherapy group than in the adjuvant chemotherapy group in patients with stage II/III ESCC.[37]

2.2.2.3 Chemotherapy and radiotherapy

More and more studies indicate limited or no survival benefits of additional surgery to treat ESCC patients with a complete response to chemotherapy or chemoradiotherapy.[38 39] A Cochrane review concluded that there is sufficient and high-quality evidence to show no significant improvement in survival by adding surgery.[40] Chemoradiotherapy (cisplatin+5-fluorouracil +50Gy) has been reported to be superior to radiotherapy alone (64Gy) in locally advanced ESCC,[41] and oxaliplatin-based and cisplatin-based definitive chemoradiotherapy has similarly positive effects.[42] Studies examining the intensification of radiotherapy dosing for ESCC are ongoing (NCT02741856, NCT02551458). Definitive chemoradiotherapy could be a nonsurgical option for locally advanced ESCC patients who are unfit for surgery. For unresectable or metastatic ESCC patients, it is also a treatment option for down-staging to make the tumor resectable. However, more than 40% of patients fail after definitive chemoradiotherapy, and these patients might need salvage surgery.[43]

2.2.3 Prognosis

More than half of ESCC patients have a late presentation to healthcare and are diagnosed at an advanced stage. The global overall five-year survival rate of ESCC was lower than 10-20% during the recent decade, ranging from 95% for stage 0 disease (cancer *in situ*) to 10-15% for stage III disease and less than 5% for stage IV.[15 44] The population-based five-year overall survival rate of ESCC patients in Sweden is 10.3%.[45] Independent prognostic predictors of ESCC are tumor stage, sub-site location, weight loss (>10%), patients' performance status, health-related quality of life, and comorbidity.[15 32 46 47]

2.3 EPIDEMIOLOGICAL ASPECT OF ESOPHAGEAL SQUAMOUS CELL CARCINOMA

2.3.1 Incidence

In 2018, approximately 572,000 new cases of esophageal cancer occurred worldwide, making it the seventh most common cancer.[48] As stated earlier, ESCC accounts for 87% of all esophageal cancer cases.[10] Its incidence varies greatly in different geographic regions, with a more than tenfold difference between some countries.[10] The highest ESCC incidence areas include southern Europe, Eastern and Southern African, and the so-called "esophageal cancer belt", which runs from Northeastern Iran through Central Asia to North-Central China.[32 49] More than half of all ESCC cases (53%) globally occur in China.[10] There are also marked differences in ESCC incidence within the same country. For example, in China, the ESCC incidence in Cixian is 18-fold higher than that in Shanghai.[50] The ESCC incidence has been reported to be as high as over 100/100,000 person-years in North Central Taihang Mountain area, while the Chinese national average rate is around 13/100,000 person-years.[50 51]

ESCC is more common in men than women (overall global male-to-female ratio of 2.7).[10] The highest male-to-female ratio is in Eastern Europe at 7.8, and the lowest is in Northern Africa and Western Asia at 1.2.[10] The sex difference in ESCC is at least partly explained by the distribution of etiological factors. Tobacco smoking and heavy alcohol consumption greatly increase the risk of ESCC in Western populations,[52 53] and the prevalence of smoking and alcohol overconsumption is much higher in men than women in most countries.[54] With the converging smoking behavior between the sexes, the male-to-female ratio decreased from 2.93 to 2.25 in the United States between 1975 and 2004.[55] However, in some high-risk areas (e.g., China), smoking and alcohol overconsumption are more similarly distributed among men and women and cannot readily explain the sex ratio difference. Thus, other risk factors could play a stronger role in these countries.

The racial disparities are substantial in ESCC. In 2013, the ESCC incidence was 3.3-fold higher in black men and 1.9-fold higher in Asian/Pacific Islander men compared to non-Hispanic white men in the United States.[56] Compared with native Asian populations (e.g.,

Korean), the incidence of ESCC is much lower in Asian immigrants in the United States, indicating a strong influence of environmental risk factors in ESCC development.[57]

Over the last few decades, alterations in the incidence of esophageal cancer subtypes have been witnessed. In some Western countries, e.g., the United States, Australia, and some countries in North-Western Europe, the incidence of ESCC has declined, whereas the incidence of EAC has increased and now exceeds the incidence of ESCC, especially in men.[10 58] ESCC is still predominant among women in these countries, and an increasing trend has even been predicted in some countries, e.g., Australia, Denmark, and Switzerland.[10 59] In Southern and Eastern Europe, Asian and African countries, ESCC represents more than 80% of all esophageal cancer cases, and in China and Japan, ESCC is almost the exclusive histological type of esophageal cancer.[51] In Japan, the incidence of ESCC has increased during the last decades.[54] ESCC is usually easy to distinguish from EAC histologically. Thus, misclassification of these tumors is unlikely to explain the reported changes in incidence.

2.3.2 Etiology

The etiological factors of ESCC vary among populations. In Western countries, the strongest risk factors for ESCC are tobacco smoking and heavy alcohol consumption. These two factors together can explain 78% of all ESCC cases in men, and a strong positive synergistic effect of tobacco smoking on alcohol use has been identified.[60] However, in high-incidence areas, e.g., the "esophageal cancer belt" countries, these exposures play less important roles in ESCC etiology. Several other risk factors have been identified, which include low intake of vegetables and fruits, consumption of red/processed meat and pickled vegetables, nutritional deficiencies, intake of very hot food/drinks, and low socioeconomic status.[61 62] Other possible risk factors of ESCC are poor oral hygiene, human papillomavirus infections, and exposure to environmental polycyclic aromatic hydrocarbons.[61] More information on risk factors is presented in Table 1. Yet, the knowledge of ESCC etiology is still limited, especially in high incidence areas.

2.3.2.1 Tobacco smoking

Tobacco smoking has a strong association with increased risk of ESCC, particularly in Western countries. Smoking alone can contribute to 65% of all ESCC cases in white men in the United States, but only 18% of cases in Chinese men.[63 64] Several case-control studies have reported three to five times higher risk of ESCC among current smokers than among never smokers.[65 66] These findings have been supported by cohort studies.[67] Dose-response associations between pack-years of tobacco smoking and ESCC have also been established.[67 68] Interestingly, longer duration of low-intensity tobacco smoking has been found to be worse than shorter duration with more intensive tobacco smoking.[65 67] An increased ESCC risk has also been found among passive smokers.[69] Nitrosamines and polycyclic aromatic hydrocarbons are believed to be the main carcinogenic agents in the contents of tobacco smoke, regardless of tobacco smoking patterns.[70]

Table 1. Risk factors for ESCC

Risk factor	Evidence	Comments
Tobacco smoking	++	Types of tobacco, linear dose-dependence
Alcohol consumption	++	Types of alcohol, J-shape dose-dependence
Low fruits and vegetable consumption	+	The effect might be modest
Red/processed meat	+	The effect might be modest and confounded by other risk factors
Picked vegetables	+	Moderate evidence, only found in Asian countries
Hot food/drinks	+	The effect differed with different temperatures and populations
Low socioeconomic status	+	Consistent risk factor but is complicated, may represent different perspectives of other factors
Nutritional deficiencies	+	Could be confounded by socioeconomic status
Betel quid	+	Confirmed risk factor by IARC, positive synergistic effects with tobacco smoking
Tylosis	+	Genetic abnormality at chromosome 17q25, familial clusters
Family history of cancer	+	Family history of esophageal and head and neck cancer have a significantly higher risk
Polycyclic aromatic hydrocarbons	+	Come from foods, beverages, and cooking or heating methods, the main carcinogenesis in some high-risk populations
Poor oral hygiene	(+)	Limited evidence, involved in inflammation and microbiota
Achalasia	(+)	High relative risk but a low absolute risk
HPV infection	(+)	No strong evidence, cases caused by HPV are low
H. pylori infection	(+)	Limited and inconsistent results
History of thoracic radiation	(+)	Dose-related, a higher risk with higher radiation, but a low absolute risk
Caustic injury	(+)	Inflammation might have distant effects
Fanconi anemia/Plummer Vinson Syndrome	(+)	High risk with low incidence rate and low absolute cases
Reproductive factors	(+)	Lack of evidence and might be protective by exposure to estrogen
Gastric atrophy	(+)	The effect varies in different populations
Opium	(+)	Not established
Low microbiome	(+)	The effect varies by composition and richness of microbiota
	++	*Strong evidence*
	+	*Moderately strong evidence*
	(+)	*Some evidence*

2.3.2.2 Alcohol overconsumption

Heavy alcohol consumption increases the risk of ESCC, and the first metabolite acetaldehyde has been recognized as a carcinogenic factor. Mutations in two central enzymes, alcohol dehydrogenase (ADH1B) and acetaldehyde dehydrogenase (ALDH2), could affect the production of acetaldehyde associated with alcohol overconsumption.[50] Exposure to alcohol alone may contribute to 72% of all ESCC cases in the United States, but only 11% of cases in China.[71 72] Studies have shown two to five folds increased risk of ESCC among alcohol consumers in high-incidence regions, e.g., Asia and South Africa. The risk is six-fold increased in European and nine-fold increased in North American alcohol consumers, compared with individuals not using alcohol.[36] The drinking patterns in these countries could be different. A possible J-shaped curve of association has been identified between the amount of alcohol consumption and risk of ESCC, with a low risk associated with low-dose of alcohol use and substantially higher risk associated with heavy alcohol consumption.[63 73 74]

2.3.3 Chemoprevention

Experimental and clinical studies have indicated preventive effects of ESCC by some groups of medications, including aspirin, statins, and metformin.[75]

2.3.3.1 Aspirin

Aspirin is an anti-inflammatory and analgesic medication. Evidence from observational studies and clinical trials have revealed that aspirin may reduce cancer risk, notably, cancer of the stomach, colon, lung, and breast.[76 77] No clinical trial has examined aspirin use in relation to ESCC risk, but observational studies have reported a protective effect and a meta-analysis study found a 40% decreased risk of ESCC among aspirin users.[78 79] A proposed anti-cancer mechanism is that aspirin inhibits cyclooxygenase-2 and consequently prohibits the synthesis of prostaglandin.[76 80]

2.3.3.2 Statin

Statin is a molecular inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase and is used to reduce serum cholesterol levels and thus prevent cardiovascular disease. Recent studies have found a potential chemoprevention effect of statins on the risk of developing cancer of the stomach, EAC, pancreas, prostate, and colon rectum,[75 81 82] but the only study examining the effect of statins on ESCC found no inverse association.[83] The biologic mechanisms for the anti-carcinogenic effects are not clear, but statin acts as a rate-limiting enzyme in the mevalonate pathway, resulting in disruption of post-translational modification of small guanosine triphosphate proteins, which are crucial for cell growth signal transduction, cellular proliferation, and cell death.[83]

2.3.3.3 Metformin

Metformin is widely used in the treatment of type 2 diabetes. It has a good safety profile and low risk of side effects. The use of metformin has been reported to counteract colorectal-,

breast-, and gastric cancer.[84] The anticancer properties could be due to the following effects of metformin: 1) activation of liver kinase B1(LKB1)/AMP-activated protein kinase (AMPK) pathway, 2) inhibition of cell division and/or promotion of apoptosis, 3) promotion of autophagy, 4) down-regulation of circulating insulin, and 5) activation of the immune system.[85] The few studies that have investigated the influence of metformin on esophageal cancer risk have found contradictory results. Cohort studies from Taiwan and the Netherland suggested a preventive effect of esophageal cancer by metformin use, while a British case-control study and another cohort study from Taiwan did not find any association between metformin use and esophageal cancer risk.[86-89] Unfortunately, none of these studies separated ESCC from EAC despite the substantial etiological differences between these histological types of esophageal cancer.[86-89] Large and prospective studies with long and complete follow-up are needed to clarify the effect of metformin on the risk of developing ESCC.

2.3.4 Early detection

ESCC is a highly lethal disease with a five-year survival rate of less than 20% in developed countries and less than 5% in developing countries, despite all new therapeutic technologies developed in recent decades.[17] Notwithstanding, 80% to 90% of early-stage ESCC patients survive after five years, indicating the great survival benefit of early detection of this disease. For endemic areas in China, population-based screening programs have been implemented and studies have proven their cost-effectiveness in reducing ESCC-related mortality.[90] But given the low absolute incidence of ESCC in the general population of the United States, a population-based screening program has not been advocated by the guidelines from the American Gastroenterology Association.[91] It is believed that stratifying population risk using a valid prediction model and implementing tailored screening programs only for high-risk individuals could be a more feasible method.

2.3.4.1 Screening

Endoscopic screening for the ESCC precursor esophageal squamous dysplasia and early-stage ESCC is regarded as a secondary prevention of ESCC.[14] To improve both sensitivity and specificity of such screening, several novel endoscopic screening modalities other than white-light endoscopy have been developed, i.e., Lugol chromoendoscopy, narrow-band imaging system, Fuji intelligent chromoendoscopy, transnasal endoscopy, endocytoscopy, and high-resolution microendoscopy. However, widespread endoscopic screening is not always feasible because of low cost-effectiveness, low tolerance by individuals, and physician-dependent diagnostic accuracy.

Simpler and less expensive non-endoscopic screening modalities have been investigated during the last few years, such as esophageal cytology specimens using balloon or sponges with or without tissue collection, breath tests using volatile organic compounds, and blood markers using autoantibodies (anti-p53), circulating microRNAs, and methylated DNA markers.[91] However, because of its relatively low sensitivity and specificity, the clinical application of these tools is currently uncertain.[17]

2.3.4.2 Prediction model

An alternative to broad screening programs is risk stratification using prediction models. It might identify a limited group at the high absolute risk of ESCC who may benefit from endoscopic screening or surveillance, with the aim of detecting ESCC at a pre-invasive or curable stage.[92-95] A few prediction models have been developed for high incidence countries. A study from China tried to identify subjects with esophageal squamous dysplasia based on environmental exposures and physical and dental examinations, but the area under the curve (AUC) was low (0.58).[94] By pooling environmental risk factors with genetic risk factors, a moderately accurate prediction model reported an AUC of 0.71.[93] A study from Iran included all known risk factors in that region and developed a prediction model with a higher AUC of 0.77.[95] All these three studies are hospital-based, and population-based studies are warranted. A recently published study used a population-based design in a rural county in China and evaluated ESCC risk factors and found an AUC of 0.80 in individuals aged less than 60 years and an AUC of 0.68 in those older than 60 years.[92] No risk prediction model of ESCC has thus far been developed for Western populations.

3 AIMS

The overarching aim of this thesis is to provide knowledge that can contribute to identifying better opportunities for the prevention of ESCC by revealing new insights into the etiology, chemoprevention, and prediction of this tumor.

To reach this overarching aim, four studies were conducted with the following specific aims:

- To assess temporal changes in incidence rates of ESCC globally in order to reveal the burden of this cancer and provide clues for its etiology (Study I);
- To clarify how tobacco smoking cessation influences the risk of ESCC (Study II);
- To elucidate whether and to which extent that metformin use prevents ESCC (Study III);
- To develop a risk prediction model of ESCC to identify high-risk individuals in a Western population who may benefit from endoscopic screening or surveillance to facilitate the detection of premalignant lesions and early-stage ESCC (Study IV).

4 METHODS

4.1 OVERVIEW

Table 2. Methods overview of the included studies

	Study I	Study II	Study III	Study IV
Short title	Global incidence trends of ESCC	Smoking cessation and ESCC risk	Metformin use and ESCC risk	Risk prediction model of ESCC
Design	Descriptive study	Systematic review and meta-analysis	Population-based cohort study	Population-based case-control study
Study period	1970-2015	-2016	2005-2015	1995-1997
Data sources	30 cancer registries in 20 countries	Medline, Embase, Web of Science, Cochrane Library, ClinicalTrials.gov	Swedish Prescribed Drugs and Health cohort (SPREDH)	Swedish Esophageal and Cardia Cancer study (SECC)
Participants	General populations	From selected studies of smoking and ESCC risk	Users of commonly prescribed medications	Incident ESCC patients and cancer-free control participants
Exposure	Time (calendar year)	Smoking status and smoking cessation	Metformin use	Age, sex, education, partnership, childhood residence, smoking, alcohol, fruits/vegetables, family cancer history
Confounders	Age, sex	As assessed in the included studies	Age, sex, calendar year, smoking, alcohol, residence, use of non-steroidal anti-inflammatory drug (NSAID) or aspirin, use of statin	Not applicable
Outcome	ESCC incidence	ESCC	ESCC	ESCC
Statistical analysis	Age-standardized incidence rates; Joinpoint analysis; Age-period-cohort analysis	Random-effects meta-analysis; Cochran's Q test and I^2 statistic	Cumulative incidence competing risk curve; Cause-specific proportional hazards model	Unconditional logistic regression; Prediction modeling

4.2 DATA SOURCES

4.2.1 Cancer registries in 20 countries

In Study I, the following 30 well-established cancer registries from 20 countries were included (in alphabetical order): Asturias Cancer Registry (Spain), Australian Capital Territory Cancer Registry (Australia), Cancer Institute New South Wales (Australia), Cancer Registry of Norway, Cancer Registry of Republic of Slovenia, Chiang Mai Cancer Registry (Thailand), Croatian National Cancer Registry, Finnish Cancer Registry, Granada Cancer Registry (Spain), Hong Kong Cancer Registry (China), Icelandic Cancer Registry, Kuwait Cancer Registry, Miyagi Prefectural Cancer Registry (Japan), Nagasaki Prefectural Cancer Registry (Japan), National Cancer Registry (Ireland), National Institute for Cancer Epidemiology and Registration (Switzerland), Netherlands Cancer Registry, New Zealand Cancer Registry, Public Health England (United Kingdom), Public Health Wales (United Kingdom), Registre des Tumeurs Digestives du Calvados (France), Scottish Cancer Registry, South Australian Cancer Registry (Australia), Surveillance Epidemiology and End Results Program (United States), Swedish Cancer Registry, Tarragona Cancer Registry (Spain), Tasmanian Cancer Registry (Australia), Tumor and Tissue Registry Office Hiroshima (Japan), Veneto Tumor Registry (Italy), Western Australian Cancer Registry (Australia).

The included cancer registries have all shown high completeness and good data quality.[96] We also assessed the completeness of the histological classification of esophageal cancer in these registries and found that unspecified histology was <20% in all included cancer registries, except for higher rates in Granada Cancer Registry in Spain (22%) and Chiang Mai Cancer Registry in Thailand (30%).

4.2.2 Publicly available databases

In Study II, we searched in publicly available databases for relevant published literature, i.e., from Medline, Embase, Web of Science, Cochran Library database, and ClinicalTrials.gov. We also reviewed the IARC monographs on “Smokeless tobacco and some tobacco-specific N-nitrosamines” and “Tobacco smoking and involuntary smoking” to identify additional relevant studies.[97 98]

4.2.3 The Swedish Prescribed Drugs and Health cohort (SPREDH)

Study III used data from the SPREDH. It is a registry-based database that started on July 1, 2005, which contains 8.4 million people with records of selected commonly-used drugs.[99] SPREDH includes data from the following four national registries and participants are linked between registries by the personal identity number, a unique identifier in all Swedish residents.

4.2.3.1 The Swedish Prescribed Drug Registry

The Swedish Prescribed Drug Registry is the primary registry in SPREDH, and it started on July 1, 2005, and contains all prescribed and dispensed drugs in all pharmacies in Sweden.

The registry attributes to about 84% of total drug prescriptions nationwide. The remaining 16% is over-the-counter medicines and drugs dispensed in the hospitals. Data on drugs in pharmacies are transferred and updated monthly to the Swedish National Board of Health and Welfare. The registry records the following information in outpatient care: age, sex, personal identity number, place of residence, dispensed medication, data of prescription and dispensation, and the prescriber's profession and affiliated practice clinics or center.[100] Information regarding each dispensed medication includes substance, brand name, formulation, package size, amount, dosage, expenditure, and reimbursement. The Anatomical Therapeutic Chemical (ATC) classification system is used to classify all medications, with a measurement unit of each prescription. Defined Daily Dose (DDD) per package are also defined.

4.2.3.2 The Swedish Patient Registry

The Swedish Patient Registry covers almost 100% of the Swedish inpatient healthcare since 1987 and almost 100% of specialized outpatient care given by public caregivers since 2001, including day surgery and psychiatric care.[99] The overall completeness of the registry is around 80% due to missing data from private health caregivers, while the completeness is 100% for specific diseases (e.g., ESCC) which are not handled in private care.[101] The registry contains information on age, sex, personal identity number, date of admission and discharge, diagnoses, and surgical procedures. The Swedish version of the International Classification of Diseases (ICD) system is used to code the diagnoses.

4.2.3.3 The Swedish Cancer Registry

The Swedish Cancer Registry was nationwide from its start in 1958 and records data on newly diagnosed malignancies among residents in Sweden. The overall completeness for all cancer types is about 96%.[102] For cancer of the esophagus and cardia, the completeness is 98%, with a 100% histological confirmation rate,[103] and the site-specific completeness is 91% for ESCC. The Cancer Registry holds information on age, sex, personal identity number, place of residence, basis of diagnosis, anatomic site, histology, stage of the tumor, date of diagnosis, and the reporting hospital and department. From 1958, the Swedish version of ICD-7 codes and WHO/HS/CAN/24.1 have been used for the anatomic site of the tumor and histological type, respectively. From 2005, the Swedish version of ICD-10 was introduced for coding of the site of the tumor, and the 3rd version of the International Classification of Disease for Oncology (ICD-O-3) was used to code histological types.

4.2.3.4 The Swedish Cause of Death Registry

The Swedish Cause of Death Registry has been available for research since the year 1952, and it records the date and causes of deaths in Sweden on an annual basis. Before 2011, deaths of Swedish residents (who died in or outside Sweden) were recorded, while from 2012, all deaths in Sweden, including those of non-Swedish residents, are included. Overall, the registry has 100% completeness for the date of death and 96% completeness for the underlying cause of death.[104] The registry includes information on age, sex, personal identity

number, place of residence, and underlying and contributing causes of death. ICD-6 was first introduced for the registry in 1951, and new versions have been introduced in 1958 (ICD-7), 1969 (ICD-8), 1987 (ICD-9), and 1997 (ICD-10).

4.2.4 The Swedish Esophageal and Cardia Cancer study (SECC)

Study IV used data from SECC, a nationwide population-based case-control study in Sweden from December 1, 1994, to December 31, 1997. Eligible participants were those younger than 80 years old, born in Sweden, and living in Sweden during the study period. The sources of ascertainment of control subjects, case patients, and data needed are presented below.

4.2.4.1 The Registry of the Total Population

The Registry of the Total Population was used to randomly select population-based control subjects. It was founded in 1968 and records data on life events such as date of birth, date of death, marital status, family relationships, migration in Sweden, and immigration from and emigration to other countries. The registry data is updated daily by the Swedish Tax Agency.[105] The coverage is virtually 100% of the general population in Sweden.

4.2.4.2 Case ascertainment

New esophageal or gastric cardia cancer cases were included by local contact persons at all 195 hospital departments and six regional cancer centers in Sweden. Half of all ESCC (born on even-numbered dates) were eligible. Uniform clinical routines were implemented in all these departments to ensure accurate information of the tumor site and histology. Endoscopic examination was performed for all patients and 97% of patients' biopsy samples or surgical specimens were re-reviewed by a single experienced pathologist.

4.2.4.3 Personal interview

Both control subjects and case patients underwent computer-aided face-to-face interviews in their own homes by professional interviewers from Statistics Sweden, who were unaware of the study hypotheses and educated to treat all participants equally. Most interviews in case patients were conducted within a few weeks after the cancer diagnosis. The average length of the interview was 80 minutes with a total number of questions between 169 and 553, depending on the answers. The interviews provided information about age, sex, tobacco smoking, alcohol consumption, education, fruit and vegetable intake, physical activity, and many other variables.

4.3 STUDY DESIGN

4.3.1 Study I

4.3.1.1 Design

Study I was a global registry-based descriptive study that explored changes in the incidence of ESCC over time using joinpoint analysis and age-period-cohort analysis.

4.3.1.2 Data collection

We collected data directly from 30 well-established cancer registries in 20 countries in Europe, Northern America, Australia, and Asia. The data covered the period from January 1, 1970, to December 31, 2015 (varying across registries). The requested data from each cancer registry included the number of newly diagnosed ESCC by the calendar year (one-year increment), age group (0-80 in five-year increments, and 85+), sex, and the size of the corresponding background population.

For classification of ESCC, most cancer registries used the code C15 in ICD-10 for defining esophagus as the site of the tumor, alternatively, the code 150 in ICD-7, ICD-8, or ICD-9 was used. The histological codes 8050-8078 or 8083-8084 were used to define squamous cell carcinoma in ICD-O-3.

4.3.1.3 Statistical analysis

Age-standardized annual incidence rates (ASR) with the gamma distribution confidence intervals (CIs) were calculated using the direct method with the WHO World Standard Population (2000) as reference.[106] We used the joinpoint regression program developed by the National Cancer Institute in the United States to identify points overtime where changes in the incidence occurred and to estimate the annual percentage change (APC) for each time segment before and after the joinpoints.[44] The APC in the incidence was computed on a relative scale (log-linear) under the assumption that the rate changed at a constant percentage linearly on a log scale within a specific period. With the joinpoint model, we calculated a weighted average of annual percent change (AAPC) with weights the same as the length of the APC interval during the whole observation period. We also performed age-period-cohort regression for each sex, to estimate the influence of age, calendar periods, and birth cohorts on the observed ESCC incidence rates. Relative rates in any given calendar period (or birth cohort), adjusted for age and non-linear cohort (or period) effects were compared with a reference period (or birth cohort).

4.3.2 Study II

4.3.2.1 Design

Study II was a systematic literature review and meta-analysis following the PRISMA statements and MOOSE guidelines.[107 108] This study aimed to estimate the influence of

tobacco smoking cessation on the risk of esophageal cancer, with a focus on ESCC, across time latencies and geographic regions.

4.3.2.2 Data collection

Relevant publications were systematically searched on the publicly available databases listed in Section 4.2.2 for studies reporting the association between tobacco smoking and risk of esophageal cancer, with three themes of Medical Subject Headings terms and related extended versions: “smoking or tobacco”, “esophageal or oesophageal”, and “cancer,

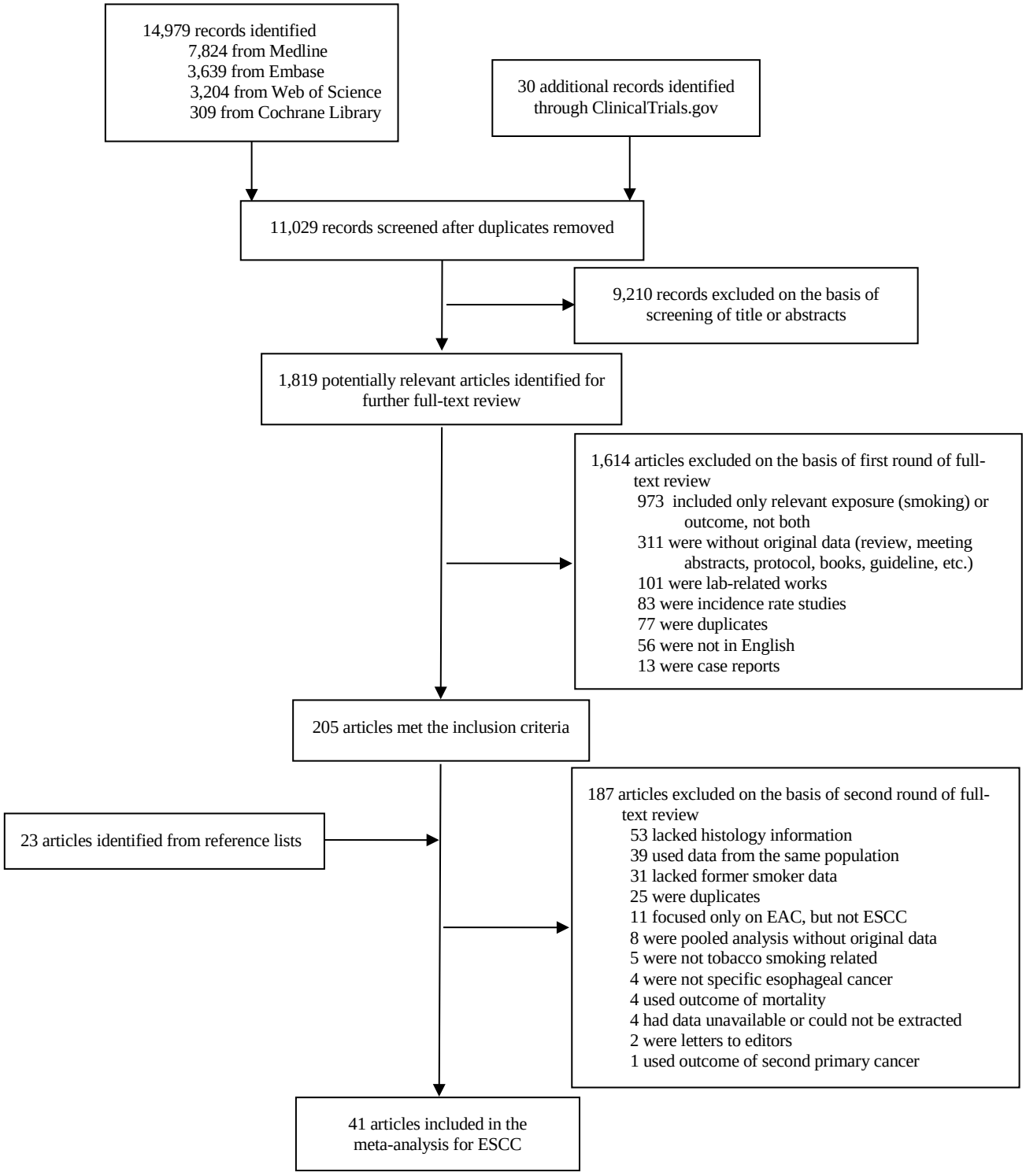


Figure 3. Flow chart of study selection

squamous cell carcinoma or adenocarcinoma". These three themes were combined with a Boolean operator "AND". Relevant references were also found in original articles, review articles, and systematic reports. There were no specific restrictions on the initial screening.

Inclusion criteria were: 1) studies where relative risk of ESCC associated with smoking status can be estimated, 2) observational studies (case-control, cohort, or cross-sectional) or interventional studies (randomized clinical trials), and 3) original and independent studies with the full texts available. We only included the most recent or most informative publication if there were more than one publication from the same population. A detailed flow chart of the study selection is shown in Figure 3.

4.3.2.3 Quality assessment

Two authors independently evaluated the quality of the identified studies, using a maximal ten-points scale including the nine-item Newcastle-Ottawa Scale and an additional item for the main exposure (smoking).[109 110] The final quality assessment items for case-control studies were: case definition, representativeness of cases, selection of controls, definition of controls, adjustment for alcohol, adjustment for diet, ascertainment of the exposure, method of ascertainment, non-response rate, and smoking as the main exposure. For cohort studies, the quality assessment items were: representativeness of the exposed cohort, selection of non-exposed cohort, ascertainment of the exposure, exclusion of outcome at baseline, adjustment for alcohol, adjustment for diet, assessment of the outcome, follow-up for at least ten years, less than 20% loss-to-follow-up, and smoking as the main exposure. The study quality score ranged from 0 to 10 points, where higher scores represent higher quality. Discrepancies between the two evaluators were reviewed together to reach consensus or further assessed by a third author.

4.3.2.4 Statistical analysis

Risk ratio (RR) was the main measure for the association between smoking cessation (or smoking status) and ESCC risk across studies. For studies using HRs and odds ratios (ORs), these measures were used as approximate estimates of the RRs value given the low incidence rate of ESCC.[111] Considering heterogeneity, pooled RR results were computed in the random-effects model using Der-Simonian and Laird's method.[112]

Statistical heterogeneity was evaluated by Cochrane's Q statistic with a significant P-value defined as <0.1 and I^2 statistic, which provided an estimate of the amount of variance across studies derived from heterogeneity rather than chance. Stratified analyses and exploratory meta-regression were carried out to examine potential sources of study heterogeneity. Sensitivity analysis by removing one study at a time was conducted to evaluate the robustness of the main results. Publication bias was examined by Begg's and Egger's tests as well as funnel plots.

4.3.3 Study III

4.3.3.1 Design

Study III was a nationwide population-based cohort study during the study period from July 1, 2005, to December 31, 2015. The objective of this study was to test whether use of metformin prevents ESCC.

4.3.3.2 Study cohort

All residents in Sweden included in the SPREDH cohort were considered for inclusion. The exposed group included those who had been dispensed metformin during the study period. The entry of the cohort was the date of the first dispensation of metformin. New metformin users were defined as those without records of metformin dispensation between July 1, 2005, and June 30, 2006 (one-year observation window), but with records of metformin use after this year. The unexposed group was those who were randomly sampled from the general population in SPREDH and had no metformin dispensation record by the time when the matched exposed participants entered the cohort. The unexposed participants were ten times as many as the exposed participants and were frequency-matched by age (within one year) and sex. Inclusion criteria were those aged 18 years or above with no previous cancer history (except for non-melanoma skin cancer) at cohort entry. Participants with metformin dispensation records before the age of 18 or with an incongruous date of death were excluded. The flowchart of participants' enrolment is shown in Figure 4.

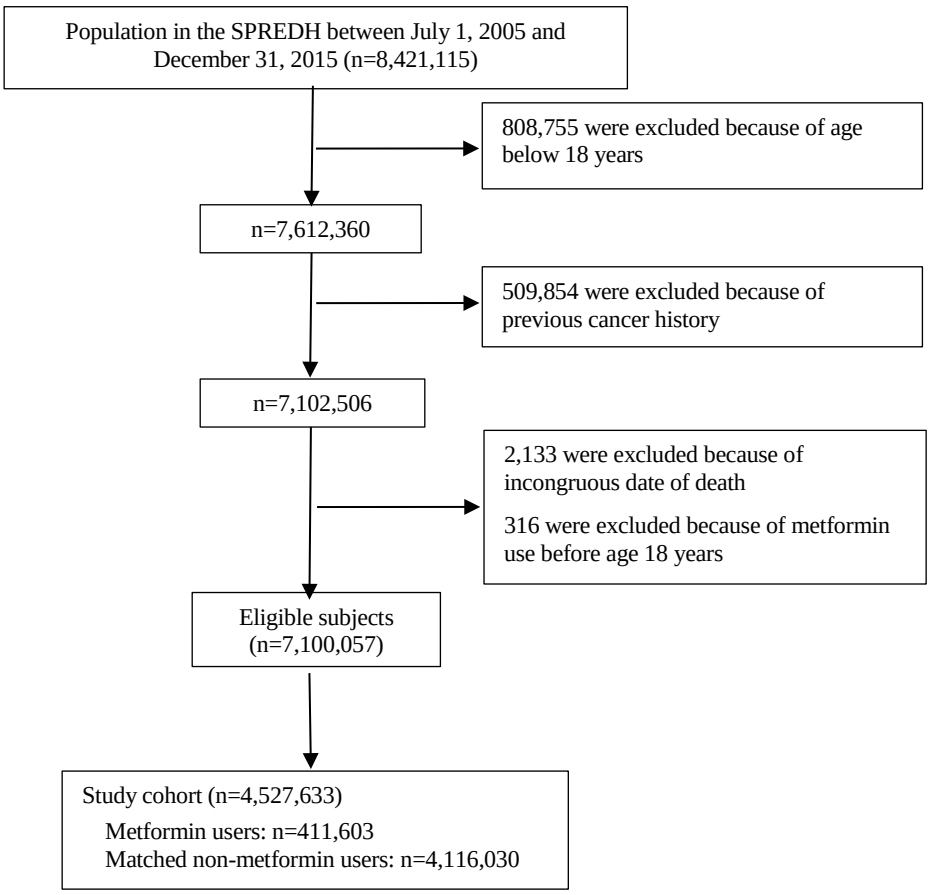


Figure 4. Flow chart of participants' enrolment

4.3.3.3 Covariates

Covariates were age, sex, calendar year of study entry, place of residence, smoking-related diagnosis (within ten years before study entry), alcohol use-related diagnosis (within ten years before study entry), use of NSAIDS or aspirin (within the first year after study entry), and use of statin (within the first year after study entry).

4.3.3.4 Outcome

The outcome of interest was a new diagnosis of ESCC. All participants were followed up until their first diagnosis of ESCC, first metformin dispensation (non-exposed group only), death, or end of the study (December 31, 2015), whichever occurred first.

4.3.3.5 Statistical analysis

Cumulative incidence competing risk curves with K-sample test were computed for the metformin user group and nonuser group. A multivariable cause-specific proportional hazards model was used to calculate HRs, treating metformin use as a time-dependent covariate. The proportionality assumption was tested using the scaled Schoenfeld residuals. Stratified analyses were performed by age at study entry, sex, and calendar period. We categorized metformin users into three groups according to the total DDD used during the first year of metformin dispensation. A dose-response analysis of metformin use in association with the risk of ESCC was conducted. Sensitivity analyses were performed by 1) censoring participants who developed any cancer (except the non-melanoma skin cancer) before ESCC and 2) excluding participants who were followed up for less than one year, regardless of exposure status and reasons for the end of follow-up. To explore unmeasured or residual confounding, a rule-out approach analysis was conducted.[113]

4.3.4 Study IV

4.3.4.1 Design

Study IV provided a risk prediction model based on a nationwide population-based case-control study (SECC) in Sweden. The aim of the study was to identify individuals of high absolute risk of ESCC who may benefit from endoscopic screening or surveillance.

4.3.4.2 Candidate predictors

We selected candidate predictors based on a literature review and associations found in analyses of data in SECC. Candidate predictors were: age, sex, years of formal education, years of living with a partner, area of residence during childhood, tobacco smoking, alcohol overconsumption, intake of fruits and vegetables, and history of cancer of the esophagus or head and neck in first-degree relatives.

4.3.4.3 Statistical analysis

Two models were developed. In a “full model”, well-established risk factors for ESCC were included directly, i.e., age, sex, tobacco smoking, alcohol overconsumption, while for other

predictors, a stepwise backward selection approach was conducted. Using the likelihood ratio test, we re-entered eliminated predictors back to the final model one by one to make sure no predictor significantly improved the goodness of fit. In a "simple model", we only included the above four well-established risk factors.

To test the performance of the models, we assessed its discriminative ability using the AUC and Somers' D statistics.[114] To consider the over-fitting, we also assessed the model performance with leave-one-out cross-validation.

An individual's five-year absolute risk of ESCC was calculated using the relative risk, population attributable risk, age- and sex-specific incidence rate of ESCC in the general population, and age- and sex-specific mortality rate from competing causes.

4.4 ETHICAL CONSIDERATIONS

Rigorous ethical considerations have been made for each of the studies included in this thesis, including individual consent, confidentiality, security regarding storage, and analysis of sensitive data.

In study I, we only collected tabulated data from different cancer registries without using any individual data. In study II, we conducted a systematic review and meta-analysis by pooling published results and analyzing the association of smoking cessation and esophageal cancer without requesting information of individuals. Therefore, ethical permissions were exempted for these two studies.

In study III, individual information was collected from national registries maintained by the governmental agency Swedish National Board of Health and Welfare. In Sweden, individual consent is exempt for registry-based research.[115] The Swedish National Board of Health and Welfare did the linkages between the registries and substituted the personal identity numbers with unidentifiable codes before delivering the data. The electronic dataset is stored on a university-based secure server, which is backed-up everyday. The derived results were interpreted and presented only at the group level, which further ensures anonymization. The Regional Ethical Review Board in Stockholm approved this study and the National Board of Health and Welfare approved to leave out the data.

In study IV, individual information was obtained from interviews and medical records, which required informed and written consents of the participants. Individual data handled by researchers were treated with the utmost confidentiality. Paper copies of medical records were stored in safes on locked university premises. All regional ethical review boards in Sweden approved the study.

5 RESULTS

5.1 STUDY I

Study I included 180,395 incident ESCC cases from the 30 registries. Of these, 61% were men. The median age at ESCC diagnosis was 67.5 years (interquartile range 62.5-77.5 years). During the study period, year 2012 witnessed the highest incidence rate of ESCC in men in Japan (9.7/100,000 person-years) and in women in Scotland (2.7/100,000 person-years). The male-to-female incidence ratio varied between 9:1 and 1:1 depending on the geographic area.

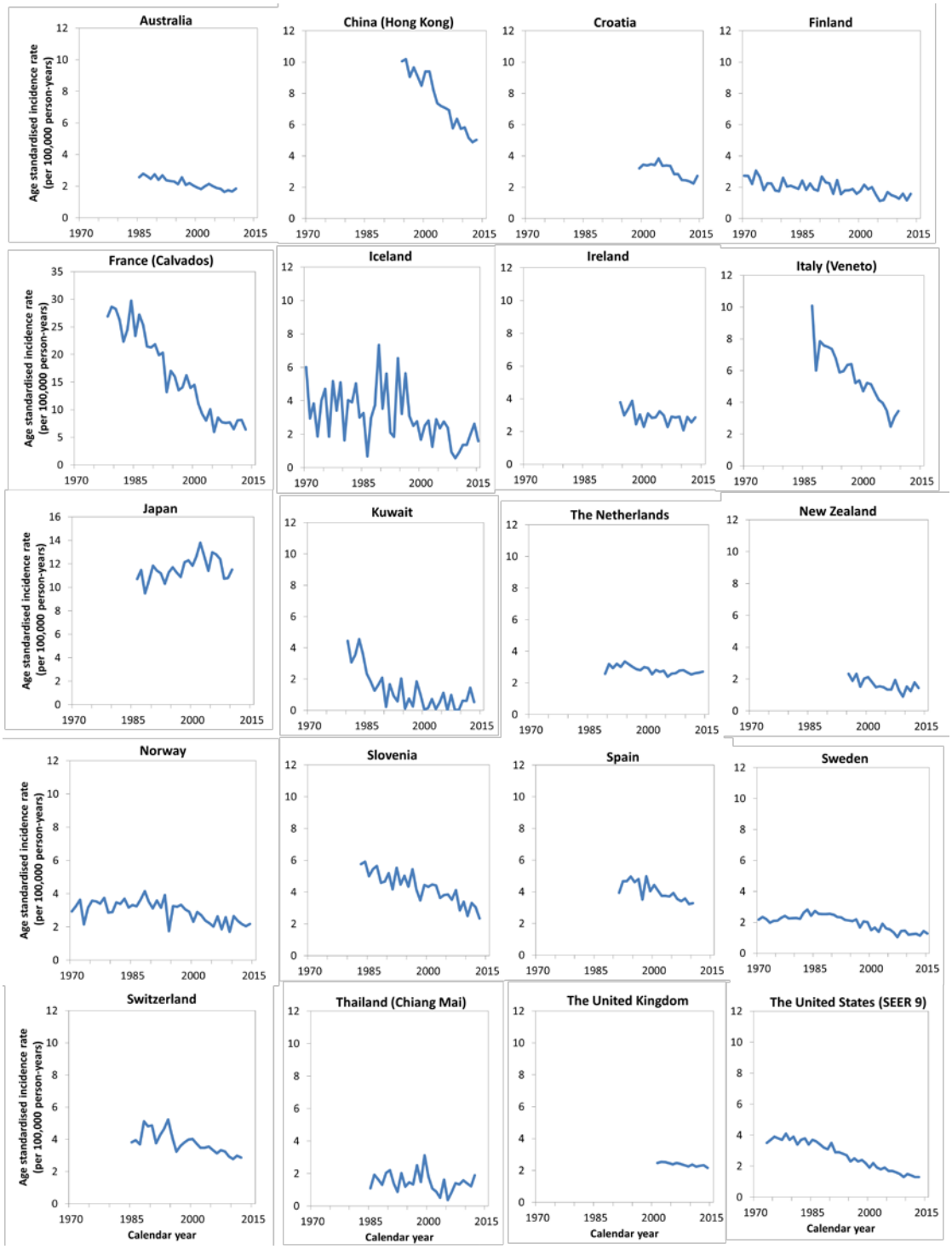


Figure 5. Global annual age-standardized incidence rates of ESCC in men

In men, the incidence of ESCC decreased from year 2000 in most countries with varying slopes and joinpoints, but the incidence rates remained high in Japan, France and Hong Kong, China (Figure 5). A sharp decline was found in Calvados, France and Hong Kong, China, whereas the incidence increased by 1% per year in Japan during 1986-2005. The age-period-cohort analysis revealed that calendar-period and birth-cohort effects might explain the decreasing trends, while the calendar-period effect contributed to the increasing trend in Japan (P=0.021).

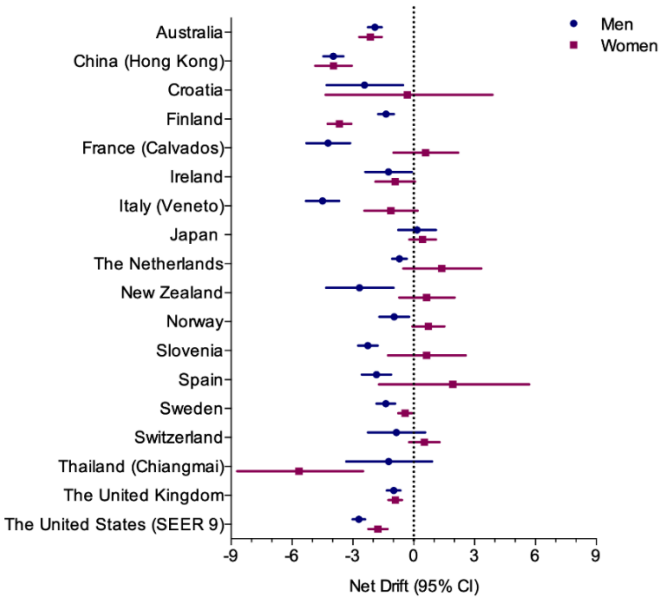


Figure 6. Average annual percentage change and 95% CI in the incidence of ESCC by sex

In women, the incidence was stable or decreased in most countries, but increasing trends were found in women in Japan, the Netherlands, New Zealand, Norway and Switzerland (Figure 6). The age-period-cohort analysis showed a significant calendar-period effect in Norwegian women (P=0.046), especially those aged 45-70 years, whereas a birth-cohort

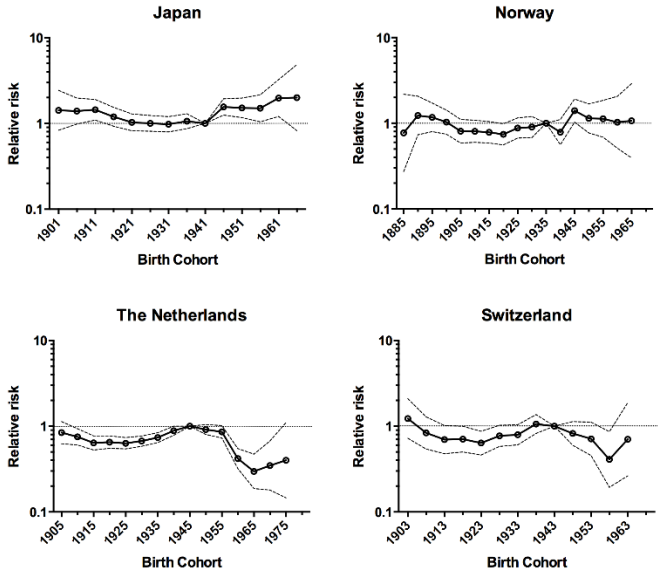


Figure 7. Relative risk for ESCC incidence by birth-cohort in women

effect was identified in women from Japan (P<0.001), the Netherlands (P<0.001), and Switzerland (P=0.014). Compared with the reference cohorts, the birth-cohort-specific relative risk increased in women born after 1945 in Japan and Norway, but not in the Netherlands or Switzerland (Figure 7).

5.2 STUDY II

In Study II, the initial search identified 15,009 publications, of which 41 were selected. Among these, 22 (54%) studies were found to be of high quality (score ≥ 7) and 18 (41%) reported results for years after smoking cessation.

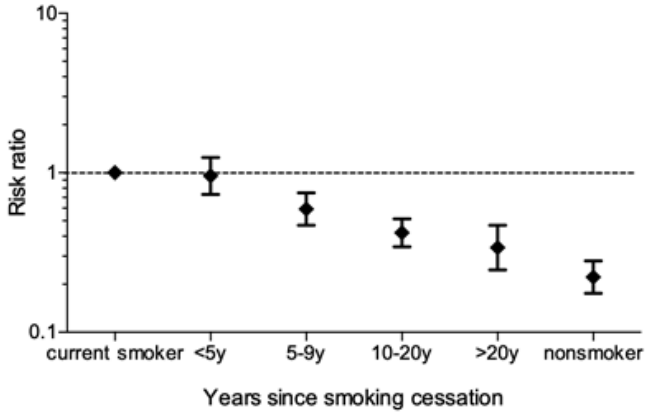
Using nonsmokers as a reference, the overall RR of ESCC was higher among current smokers (RR 4.18, 95% CI 3.42-5.12) than among former smokers (2.05, 95% CI 1.71-2.45) (Table 3). In North America, the risk difference between current smokers (RR of 5.75, 95% CI 3.56-9.26) and former smokers (RR of 2.45, 95% CI 1.83-3.27) was largest, while such big difference was not found in Asian populations. There were no major sex differences in the risk of ESCC among current or former smokers.

Table 3. Tobacco smoking status and risk of ESCC, using nonsmokers as the reference

Study characteristics	RR (95% CI) in former smokers	Studies (n)	RR (95% CI) in current smokers	Studies (n)
Overall	2.05 (1.71-2.45)	41	4.18 (3.42-5.12)	41
Study design				
Case-control	2.01 (1.67-2.43)	37	3.81 (3.06-4.74)	37
Cohort	2.50 (1.29-4.85)	4	6.95 (4.17-11.57)	4
Publication year				
≤1999	1.98 (1.54-2.54)	8	5.07 (3.35-7.68)	8
2000-2009	1.85 (1.47-2.34)	21	3.62 (2.83-4.64)	21
≥2010	2.57 (1.69-3.91)	12	4.29 (2.54-7.24)	12
Geographic origin				
North America	2.45 (1.83-3.27)	7	5.75 (3.56-9.26)	7
Europe	1.75 (1.15-2.65)	14	4.57 (3.19-6.54)	14
Oceania	2.18 (1.51-3.17)	1	4.58 (2.99-7.02)	1
Asia	2.47 (1.78-3.44)	12	2.82 (1.81-4.39)	12
South America	1.67 (1.37-2.04)	7	2.91 (2.41-3.50)	7
Sex*				
Men	2.00 (1.43-2.80)	10	3.77 (2.29-6.20)	10
Women	1.34 (0.71-2.53)	6	3.85 (2.20-6.74)	6
Unspecified	2.26 (1.86-2.76)	29	3.94 (3.12-4.99)	29
Response rate				
≥80%	1.92 (1.56-2.36)	15	4.21 (2.74-6.47)	15
<80%	2.42 (1.70-3.45)	10	4.80 (3.09-7.43)	10
Unknown	1.95 (1.36-2.78)	16	3.51 (2.69-4.60)	16
Tobacco types				
Cigarettes	2.38 (1.58-3.60)	15	4.02 (3.07-5.28)	15
Unspecified	1.94 (1.66-2.26)	26	3.96 (2.94-5.33)	26
Study quality				
Low (score<7)	2.16 (1.63-2.86)	19	3.60 (2.65-4.90)	19
High (score≥7)	1.97 (1.57-2.48)	22	4.45 (3.31-5.99)	22

** Two studies (Stefani et al.1990 and Victora et al.2007) reported RR for men and women separately and combined; one study (Kabat et al.1993) reported RR for men and women separately.*

A dose-response association was found between years after smoking cessation and risk of ESCC (Figure 8). Compared with current smokers, those who had quit smoking <5 years earlier had no decreased risk (RR 0.96, 95% CI 0.73-1.25), while the risk was decreased following smoking cessation for 5-9 years (RR 0.59, 95% CI 0.47-0.75), further decreased after 10-20 years (RR 0.42, 95% CI 0.34-0.51), and was lowest >20 years after smoking cessation (RR 0.34, 95% CI 0.25-0.47). The latter risk estimate was similar to that in never smokers (RR 0.22, 95% CI 0.18-0.28).



Smoking cessation	RR (95% CI)
Current smoker	1.00 (reference)
<5 years	0.96 (0.73-1.25)
5-9 years	0.59 (0.47-0.75)
10-20 years	0.42 (0.34-0.51)
>20 years	0.34 (0.25-0.47)
Nonsmoker	0.22 (0.18-0.28)

Figure 8. Risk ratio of ESCC by duration since smoking cessation

The meta-regression analysis revealed that sex and study quality caused heterogeneity in the analysis of former smokers, while study geographic origin and control selection method resulted in heterogeneity in the analysis of current smokers. No publication bias was found.

5.3 STUDY III

Study III included 411,603 metformin users with a mean follow-up time of 6.0 (±3.4 standard deviation [SD]) years and 4,116,030 nonusers of metformin who were followed up for a mean of 5.8 (±3.5 SD) years.

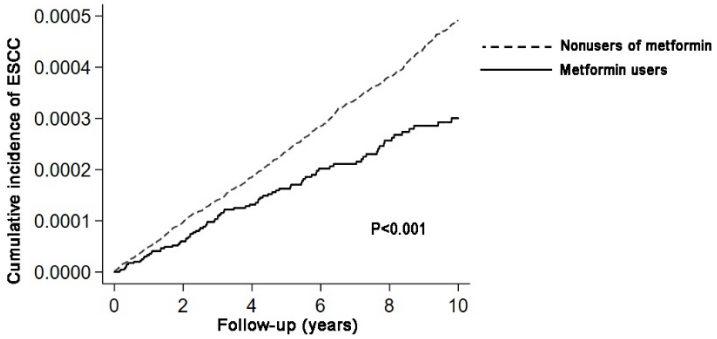


Figure 9. Cumulative incidence of participants developing ESCC during the follow-up among metformin users and nonusers of metformin

The incidence rate of ESCC in metformin users and nonusers were 3.5/100,000 person-years and 5.3/100,000 person-years, respectively (Figure 9). Its cumulative incidence was significantly lower in the metformin user group than in the nonuser group (P<0.001). Multivariable cause-specific proportional hazards model revealed an HR of 0.68 (95% CI 0.54-0.85) in ever metformin users and 0.44 (95% CI 0.28-0.64) in new metformin users (Table 4). Sensitivity analyses and the rule-out approach analysis supported the robustness of these main results.

Table 4. Risk of ESCC in metformin users compared with nonusers in Sweden in 2005-2015

Analysis	ESCC (n)	Model 1	Model 2	Model 3
Metformin use	1,348	0.66 (0.53-0.82)	0.67 (0.54-0.84)	0.68 (0.54-0.85)
New metformin use	627	0.46 (0.32-0.67)	0.46 (0.31-0.67)	0.44 (0.28-0.64)
Sensitivity analysis 1*	1,210	0.68 (0.54-0.85)	0.69 (0.55-0.87)	0.70 (0.56-0.88)
Sensitivity analysis 2†	1,142	0.66 (0.52-0.84)	0.67 (0.53-0.85)	0.68 (0.53-0.86)
Dose-response analyses				
≤175 DDD	22	Reference	Reference	Reference
175-300 DDD	24	0.94 (0.53-1.68)	0.90 (0.50-1.61)	0.91 (0.51-1.63)
> 300 DDD	27	0.83 (0.47-1.47)	0.88 (0.49-1.59)	0.89 (0.49-1.61)

*Model 1: crude analysis; Model 2: adjusted by age, sex, calendar year, residence, smoking, and alcohol overconsumption; Model 3: model 2 further adjusted by non-steroidal anti-inflammatory drugs/aspirin and statin use. * Censored those who developed any other cancers (except for non-melanoma skin cancer) before ESCC during follow-up; † Excluded those with less than one-year follow-up, regardless of exposure status and reasons for the end of follow-up.*

Compared with low-dose (<175 DDD) metformin use in the first year of enrollment, the intermediate-dose group (175-300 DDD) and high-dose group (>300 DDD) had slightly lower point estimates (HR 0.91, 95% CI 0.51-1.63 and HR 0.89, 95% CI 0.49-1.61, respectively) (Table 4).

5.4 STUDY IV

Study IV included 167 ESCC cases and 820 controls. After the predictor selection strategy, six variables remained in the full model: age, sex, tobacco smoking, alcohol overconsumption, a combined variable of education years and partnership years, and childhood residence. The simple model included age, sex, tobacco smoking, and alcohol overconsumption. The population attributable fraction was 0.95 for the full model and 0.85 for the simple model.

The full model had an AUC of 0.81 (95% CI 0.77-0.84) and the simple model had an AUC of 0.79 (95% CI 0.75-0.82) (Figure 10). The leave-one-out cross-validation analysis slightly reduced the AUC to 0.78 (95% CI 0.75-0.82) in the full model and 0.75 (95% CI 0.71-0.79) in the simple model.

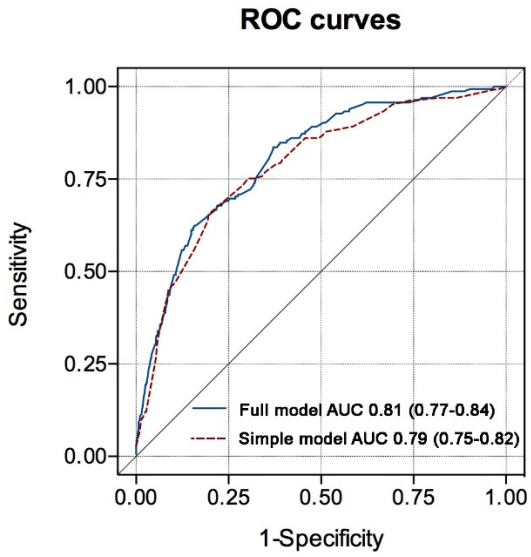


Figure 10. Receiver operating characteristic (ROC) curves based on the full model and the simple model

A heat chart based on the simple model is presented in Figure 11, containing an estimated absolute five-year risk in individuals with different risk exposures. In the highest absolute risk group (281.4/100,000 person-years), 355 subjects need to be surveyed to identify one case of ESCC within five years. Detailed absolute risk calculators for both models are available in the online appendixes with the link below: https://doi.org/10.1016/j.gie.2018.10.025

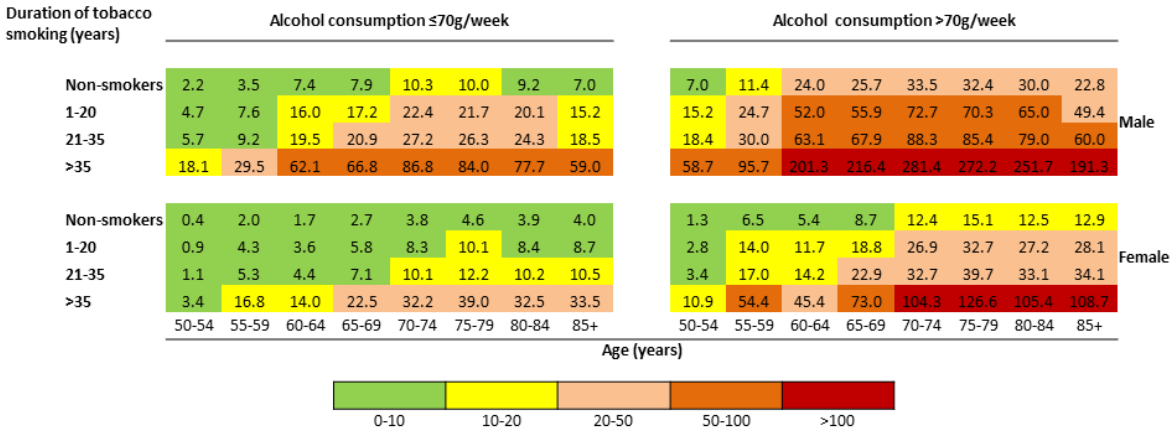


Figure 11. Absolute five-year risk of ESCC per 100,000 person-years, estimated from the simple model in individuals aged 50 years or above

6 DISCUSSION

6.1 METHODOLOGICAL CONSIDERATIONS

6.1.1 Study design

This thesis comprises a mix of study designs, i.e., a descriptive study (Study I), a systematic review and meta-analysis (Study II), a population-based cohort study (Study III), and a population-based case-control study (Study IV).

Epidemiological studies include both experimental and observational study designs, but this thesis only includes observational studies. The research questions under study are not feasible and unethical to examine in an experimental study (e.g., randomized clinical trial) because of exploration in etiology and limited prior evidence of preventive measures. The main observational study designs are cohort studies, case-control studies, cross-sectional studies, and ecologic studies. *Cohort studies* classify participants in a source population according to their exposure status and follow them over a certain period to assess disease incidence. *Case-control studies* identify cases and controls from the same source population and classify them according to their exposure history. Observational studies can also be categorized as either prospective or retrospective studies. It is not straightforward to use these terms, but one recommendation is to use these terms to elucidate whether the outcome could influence the exposure information.[116] Thus, prospective studies refer to studies where the outcome could not influence the exposure information and vice versa.

Descriptive studies often describe the characteristics or demographics of a population. A descriptive study collects quantitative information for statistical analysis, and typically uses a cross-sectional design. Study I is a descriptive study of global incidence trends.

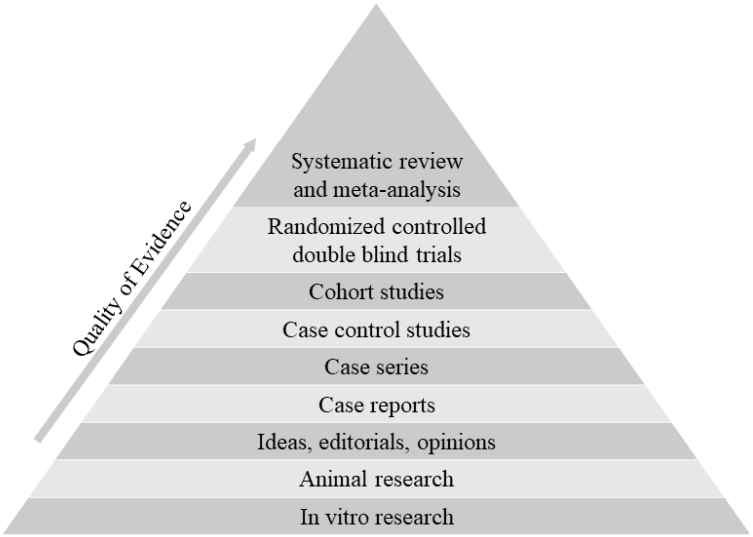


Figure 12. Evidence level for research

Systematic review and meta-analysis studies are regarded as the highest quality of evidence level in research by several researchers (Figure 12).[117] These studies aim to include all

available and relevant studies on a specific topic by using a systematic and comprehensive search strategy. They also evaluate the studies' quality and combine the results from different studies. In Study II, we identified published studies on smoking cessation and ESCC risk, assessed their validity, and analyzed the combined results.

In Study III, registry-based data were used and data on medication use (exposure) were collected before the onset of ESCC (outcome), suggesting a prospective design. In Study IV, cases of ESCC (outcome) were confirmed before collecting exposure information via interview, making it a retrospective study design, and recall bias can be an issue. However, our study included both ESCC and EAC patients and clearly different risk factor patterns were revealed for these two diseases, which indicate the limited influence of recall bias.

6.1.2 Measure of disease

6.1.2.1 Age-standardization

Incidence rate is a measure of the occurrence of a disease, computing number of new cases in the background population during a specific time frame. Incidence rate is frequently presented as a number of cases per 100,000 person-years in a rare disease setting.

To have a comparable disease burden in populations with different age structures, age-standardization methods are often used. The direct method or indirect method can be applied for age-standardization. Generally, direct age-standardization is favored over the indirect method. For age groups (I), the direct method calculates age-specific rates (r_i) in the target population with weights (w_i) from a standard population, using this formula:

$$ASR = \frac{\sum_{i=1}^{I}(r_i \times w_i)}{\sum_{i=1}^{I} w_i}$$

In Study I, we calculated ASR for each of the included countries during the study period, using the WHO World Standard Population (2000) as a reference, a most frequently used standard population. Therefore, the derived results for each population can be compared with other populations within this study and with other studies using the same reference.

6.1.2.2 Age-period-cohort analysis

The age-period-cohort model is a parametric statistical model that can summarize ESCC incidence rate trends over time. The complete age-period-cohort model can be written as in the formula bellow,[118] with both linear components (a, p, c) and non-linear components (x_{Ai}, x_{Pj}, x_{Ck}):

$$\ln\left(\frac{y}{n}\right) = \mu^L + \alpha^L a + \beta^L p + \gamma^L c + \sum_{i=1}^{n_a-1} \rho_{Ai} x_{Ai} + \sum_{j=1}^{n_p-1} \rho_{Pj} x_{Pj+} \sum_{k=1}^{n_c-1} \rho_{Ck} x_{Ck}$$

The model assumes that no period effect ($\beta^L = 0$) could yield a longitudinal model and no cohort trends ($\gamma^L = 0$) could lead to a cross-sectional model. As birth cohort (c) = time

period (*p*) - age (*a*), this model can be transferred into two two-factor models: age-cohort model and age-period model. The age-cohort model is generally preferred in cancer epidemiological settings, given that the exposures usually take a long time to influence the outcome, thus making the cohort better represent the pattern of exposure than period. *Net drift* represents the average annual percent change of incidence rate. Net drift equaling zero defines no changes over time, with proportional longitudinal and cross-sectional age curves. *Local drift* represents the annual percent change of incidence rate. Local drift equaling the net drift implies the same time trends in each age group. Using a reference period (or cohort), period rate ratios (or cohort rate ratios) can be estimated after adjusting for age and non-linear cohort effect (or non-linear period effect). *Period rate ratios* equaling one may imply constant time trends and that the cross-sectional age curve shows age incidence pattern in each period. *Cohort rate ratios* equaling one indicates that all local drifts equal zero and that the longitudinal age curve represents age incidence in each cohort. In Study I, all the above-mentioned terms were computed and used to interpret the time trends changes in different countries.

6.1.3 Systematic review and meta-analysis

6.1.3.1 Aggregation bias

Aggregation bias (or *ecologic bias*) occurs when we measure group outcomes based on means or rates of group exposure, rather than values of individual exposure. It is also common that meta-analyses compute the grouped results adjusting for the mean values of other covariates, which could distort the results. Aggregation bias, therefore, might exist in meta-analyses, and the aggregated results have to be interpreted with caution. In Study II, we performed several sensitivity analyses, all indicating the robustness of the results. The strong association between smoking cessation and decreased ESCC risk compared to continued smoking is also less likely to be affected by other covariates. Yet, careful interpretation of the results is still needed.

6.1.3.2 Exclusion bias

Exclusion bias (or *selection bias*) in systematic reviews might come from the inappropriate exclusion of studies, such as using an incomplete search strategy or exclusion of small sample size or low-quality studies, specific study type (e.g., case-control studies, or specific study population), and studies with less informative data. In Study II, we initially identified 15,009 publication records using a predefined and comprehensive search strategy and we enrolled all available studies, regardless of their sample size, study quality, or study type. To evaluate bias from differences in study quality, we stratified the analyses according to different characteristics of studies. Less informative studies might bias the overall results; however, this should be limited because we only identified four such studies.

6.1.3.3 Publication bias

Publication bias is a major source of bias in systematic reviews and it is usually related to other sources of bias, e.g., significance bias, study size bias, and suppression bias from sponsors. Such bias has to be carefully assessed before making conclusions. In Study II, we tested the publication bias using funnel plots, and Begg's and Egger's tests.[116] None of these methods identified publication bias in the study, lending validity to the findings of the study.

6.1.3.4 Heterogeneity

Population heterogeneity and methodological heterogeneity are common in systematic reviews. *Population heterogeneity* derives from differences in the study region, population age and sex, or risk factors in the diseases. *Methodological heterogeneity* comes from differences in study design, measurement of exposures and outcomes, adjustment for covariates, and statistical methods. Heterogeneity was assessed using Cochran's Q test and I^2 statistic in our study. Given the relatively large number of included studies, both subgroup analyses and meta-regression were applied to explore the source of heterogeneity, as well as the random-effects model which is usually more conservative than the fixed-effects model. Furthermore, quality scoring system and "move-one-out" sensitivity analysis were conducted, both of which support the robustness of the study results.

6.1.4 Internal validity

6.1.4.1 Selection bias

Selection bias might occur due to problems in study subjects' participation. A pitfall in selecting participants is neglecting those who are lost to follow up and those who are eligible for the study but do not participate, especially when their exposure and outcome patterns differ from those included. One type of selection bias is *self-selection bias*, e.g., patients exposed to the risk factor or with diseases are more likely to be involved in the study. In Study III, almost all metformin users in Sweden during the study period were included in the study. Non-metformin users were selected from the 8.4 million background population among about ten million national populations at that time. Disease status was obtained during the follow-up by linkage to national registries. Selection bias in that study should be less likely because of the complete follow-up of the cohort. In Study IV, the case-control study, a systematic sampling method (born on even dates) was applied to enroll half of all national ESCC cases, and controls were randomly selected from the national population registry. Both cases and controls had a rather high participation rate of 73%. In addition, a separate analysis showed no differences in baseline characteristics comparing non-participants and participants. Thus, selection bias should not strongly bias the results of Study IV.

6.1.4.2 Information bias

Information bias derives from measurement errors when collecting information for a study. *Non-differential misclassification* occurs when the misclassification of subjects' exposure is unrelated to the status of the participants' covariates, or disease. This error tends to

incorrectly turn the results to null values. *Differential misclassification* results from measurement errors that are not equal in the exposed (or diseased) and non-exposed (or non-diseased) group and it could bias the results in any direction. *Recall bias* might influence the results of case-control studies when researchers try to collect exposure or covariates information, but differences in the reporting of information occur because of the case or control status of the participants. The direction of recall bias is unpredictable and can either exaggerate or underestimate the estimates. Therefore it should be avoided or limited in the study. *Detection bias* happens during the disease information collection process when the possibility of being detected for the disease under study differs in exposed and non-exposed participants. In Study III, information bias was avoided by using registry-based data, indicating a prospective collection of exposure information and almost 100% completeness of disease information. Recall bias can exist in Study IV for the case-control setting, but it should be limited because of the distinct differences in etiological patterns found for ESCC and EAC patients in the same study.

6.1.4.3 Confounding

Confounding factors (confounders) can explain parts of or all differences between the measure of the association and the measure of the effect that could be achieved in an ideal counterfactual setting.[116] In other words, a confounder spuriously biases the exposure-outcome association by influencing both the exposure and the outcome, without being in the causal pathway (i.e., not a mediator). Confounding needs to be considered in observational studies either in the study design phase (e.g., matching by a confounder in a cohort study, restricting the participants regarding the status of confounders) or during the statistical analysis process (e.g., adjustment and stratification). To counteract confounding in Study III, we matched metformin users with nonusers by age and sex, adjusted for some other confounders in the statistical model, and stratified the analyses by potential confounders. Although residual or unmeasured confounding might still exist, any confounding should not alter the identified association to null, as indicated in the rule-out analysis.

6.1.4.4 Random error

Another main methodological issue is random error (or *chance variation*), which is inverse to *statistical precision*. It derives from unexplained variations in statistical measurements or the sampling process from the so-called “super population”. A risk of random error is unavoidable, but we can reduce it by increasing the sample size and avoiding multiple testing. The precision is often estimated as CI or P values in statistical analysis. *95% CI* is defined as, if repeatedly sampling from a “super-population” with different sample populations and 95% CIs are computed for each of the sample population, then at least 95% of these intervals include the true value of the “super-population”, providing no bias exists. Namely, there is 95% confidence that the true value from the “super-population” is included. The *P-value* (or *probability value*) is derived from significance testing and tests the probability of attaining the current observed results assuming that the null hypothesis is correct. It is the possibility of detecting a difference in sample populations (reject the null hypothesis) when no difference

exists in the "super population" (null hypothesis is true) (*Type I error*). To reduce Type I errors, one may either lower the significance testing level or perform a multiple testing correction if multiple testing has been conducted. But both of these methods result in an increased chance of *Type II error*, which is defined as the possibility of not rejecting the null hypothesis when it is false. The studies included in this thesis did not use any of these two methods. Instead, we limited the predefined study hypothesis testing and only included covariates according to subject-matter knowledge.

6.1.5 External validity

External validity or *generalizability* concerns whether the findings from a study are valid also in other populations or settings. Representativeness is often considered as a hindrance of good internal validity, and without internal validity, it is impossible to even discuss external validity. Therefore, internal validity is often given priority rather than generalizability. Study II merits high generalizability since we included many populations in the world. Despite high internal validity, only the Swedish population was considered in Study III and IV, which may limit the external validity of these two studies to other populations, especially non-Western populations.

6.1.6 Assessment of the performance of prediction models

6.1.6.1 Discrimination

Discrimination of a prediction model refers to the ability to discriminate those with the outcome from those without outcome. The discriminative ability can be assessed by the *concordance (c) statistic*, which is a rank-order statistic and identical to the *AUC* for binary-outcome studies. *Somers' D statistic* measures the direction and strength of predictions against observed outcomes, related to *c* statistic.[114] In addition, the discrimination slope assesses the absolute difference between the average predicted probability with and without the outcome and is usually visualized as a box plot or histogram. In Study IV, we evaluated the risk prediction model using both AUC and Somers' D statistic. Both the derivation model and the cross-validation model showed good discriminative abilities.

6.1.6.2 Calibration

Calibration is the agreement between predictions and observed outcomes within a certain period. A calibration plot is a common tool to assess the calibration of a risk prediction model, with predictions on the x-axis and observations on the y-axis. An ideal calibration plot should have a slope *b* of 1 and an intercept *a* of 0 (*calibration-in-the-large*, indicating if the model is systematically skewed). For binary outcome, albeit of being criticized as arbitrary grouping, the *Hosmer-Lemeshow goodness-of-fit test* is often applied to the plotted observed outcome by decile of predictions. However, calibration in Study IV could not be assessed due to the case-control study design and limited sample size of control participants, which makes it unfeasible to observe even one case during the next decades, given the low incidence rate of ESCC.

6.2 GENERAL DISCUSSION

6.2.1 Study I

Study I presented an overview of global incidence trends of ESCC by directly collecting data from 30 cancer registries in 20 countries. The incidence of ESCC in men had decreased in most of the 20 countries included in this study, although the rate was still high in Japan, France, and Hong Kong, China. The incidence of ESCC was lower in women than men and was stable or slightly decreased in women in most countries. However, an increasing trend was observed in women in Japan, the Netherlands, New Zealand, Norway, and Switzerland. Birth-cohort effects played a stronger determinant role in these countries, indicating potential etiological alterations. Age-period-cohort results suggested a further increasing incidence of ESCC in Japanese and Norwegian women in the coming decades.

There is a need for cautious interpretation of temporal incidence trends. Changes over time regarding, e.g., diagnostic techniques, screening programs, or completeness of registration systems, might contribute. However, such changes should not explain the decreasing incidence rates identified in the study. A more probable reason for the decrease is changes in risk factor exposures, which is supported by the significant cohort effect from the age-period-cohort analyses. Changes in smoking and alcohol consumption, the two main risk factors of ESCC, could contribute. The generally decreasing incidence of ESCC after the 1990s might be explained by the global smoking control activities since the 1980s and decreasing or stable prevalence of alcohol consumption.[119-121] Switzerland has an increasing prevalence of smoking in women (from 21% in 1980 to 26% in 1998), which might explain the increased incidence of ESCC.[119] In Japanese women, increased alcohol consumption has been reported, which could contribute to the increasing incidence.[120 122 123]

Compared with previous ESCC incidence trends studies, this study is more updated, includes more registries, covers a longer period, and is based on data directly retrieved from the registries. The age-period-cohort analyses also provide etiological clues for future studies. This study did not have national registries for some countries, e.g., Japan, Spain, and the United Kingdom. In these countries, we merged regional registries. Unfortunately, we were unable to collect data from mainland China, which contributes almost half of all ESCC cases in the world. This study indicates that although smoking control activities may have contributed to the decreasing trends of ESCC, more efforts towards alcohol control are warranted, especially in women.

6.2.2 Study II

Study II revealed a substantially lower risk of ESCC in former smokers than in current smokers and a clear time-dependent association between smoking cessation and ESCC risk. Compared with current smokers, a decreased risk started five years after quitting smoking, and long-term smoking quitters (i.e., 20 years' smoking cessation) had almost as low risk as never smokers.

Lower risk in former smokers than current smokers was also reported in a meta-analysis study of Japanese studies.[124] Although without specifying the histological subtype of esophageal cancer, about 90% is ESCC in Japan.[125] The geographic differences of ESCC risk could be partly justified by other risk factors (e.g., passive smoking, hot food and drinks, and pickled vegetables), which facilitate a high-risk baseline of ESCC in Asian countries. Quitting smoking alone might have a limited influence on decreasing the risk of ESCC in Asian populations.

This study is merited by its extensive search strategy, which identified a large number of studies and participants. Detailed subgroup analyses and dose-response analyses were, therefore, possible. Nevertheless, heterogeneity was discovered for the included studies, which was mainly caused by differences in sex distribution, study quality, geographic region, and source of controls. Therefore, stratified analyses by these factors were carefully performed, and similar results suggest the robustness. Additionally, restriction by English publications might bias results in the study, but this should be a limited concern because language restriction was only performed at a late stage of the inclusion process and records in other languages were few. The gradually decreased ESCC risk following smoking cessation is biologically plausible, and this study reinforces the value of recommending abstinence from smoking.

6.2.3 Study III

With a large population-based national cohort, Study III addressed the hypothesis that metformin use decreases the risk of ESCC. Metformin users were about 32% less likely to be diagnosed with ESCC than their comparators from the background population.

Previous studies have reported contradictory results regarding the association between metformin and ESCC risk. An increased risk of esophageal cancer associated with metformin use was reported in a case-control study from Taiwan (OR 2.84, 95% CI 0.99-8.18), whereas a decreased risk was suggested in a Taiwanese cohort study (HR 0.49, 95% CI 0.35-0.68) and a Dutch cohort study (HR 0.90, 95% CI 0.82-0.97).[88 89 126] A few other studies did not reveal any association.[86 87 127] A shared limitation of all these studies was that they did not separate ESCC from EAC, despite their distinct etiological patterns. The lack of adjustment for obesity in these studies probably introduced confounding because obesity is strongly associated with both diabetes (and thus metformin use) and EAC, whereas obesity is not a risk factor for ESCC and should therefore not confound the results of the present study.

One could argue that diabetes diagnosis could bias the results, but no association has been found between diabetes and ESCC.[128 129] Another concern might be bias from the use of insulin, which has been reported to increase the risk of esophageal cancer.[126] Patients who receive insulin usually have more severe diabetes and are therefore less likely to receive metformin. However, any confounding by insulin should be limited, given that this study only included 2% of insulin users. Compared with the group using the lowest dose of metformin, the higher-dose groups had no clearly further risk reduction. This may indicate a

possible benefit of a low-dose medication, which is supported by the finding of a lower risk of ESCC in new metformin users than in ever users. However, we cannot exclude that such lower risk may be due to different drug exposures between metformin new users and ever users, i.e., different exposure to other anti-diabetes drugs (e.g., sulfonylurea), which may increase the risk of ESCC.

Among the strengths of this study is the population-based cohort study design with long and complete follow-up using well-established registry data, which should reduce concerns about selection bias and information bias. The matched-cohort study design with competing risk analyses counteracted limit time-related bias. Yet, residual or unmeasured confounding (e.g., from dietary factors) possibly remains, although the rule-out analysis showed that such confounding should be limited. Finally, misclassification of metformin use may affect the association because we cannot know for sure that participants dispensed with metformin actually used the medication.

6.2.4 Study IV

In Western countries, routine endoscopic screening or surveillance of the general population for ESCC is not cost-effective or feasible. Instead, a prediction model could help select high-risk individuals. With readily identifiable predictors, this study constructed a full model and a simple model that had a good ability to predict an individual's absolute five-year risk of ESCC.

It is the first prediction model on ESCC risk in a Western population and is based on a national Swedish case-control study of high internal validity. However, several limitations of the model warrant attention. Recall bias cannot be completely excluded, although one could dispute this bias by the different risk factors profiles found for EAC in the same study. Some uncertainties in the precision of predicted risk estimates remained due to limited power in some subgroups. Despite good interval validation of the developed model, external validation using an independent population was not conducted. Thus, overfitting or overestimation is possible and the calibration of the model is difficult to evaluate.[130] Therefore, we have recently conducted a prediction model using two large independent cohorts (not included in this thesis), the Nord-Trøndelag Health Study (HUNT, as model derivation cohort), and the UK Biobank cohort (as external validation cohort). To accurately estimate the ESCC risk, we applied a competing risk model (the Fine and Gray model). The final model included similar predictors as those used in the present study, i.e., age, sex, smoking, alcohol, but also body mass index. The model had an AUC of 0.77 (95% CI 0.67-0.87) for a 15-year risk and 0.76 (95% CI 0.59-0.93) for a five-year risk in HUNT. External validation in UK Biobank found an AUC of 0.70 (95% CI 0.65-0.76). The Hosmer and Lemeshow goodness-of-fit test indicated a good fit (P >0.05 for all models), and the calibration plots are shown in Supplementary Figure 1.

This easy-to-use risk prediction model of ESCC could be useful for healthcare providers and public health decision-makers. The model can estimate an individual's risk of ESCC within

specific years. Stratifying populations into different risk groups may enable more cost-effective screening and surveillance programs, which could promote the detection of premalignant conditions and early-stage ESCC, thus may help reduce deaths from this aggressive tumor.

7 CONCLUSIONS

- The incidence of ESCC has decreased in most of the examined countries, but increasing trends were found in women in a few countries. Birth-cohort effects were important for the changes, indicating a need for preventive measures.
- Smoking cessation time-dependently reduced ESCC risk compared to continued smoking, and the risk reduction started after >5 years of smoking cessation and approached the level of never smokers after >20 years.
- Metformin use may decrease the risk of ESCC.
- A model based on four readily available predictors (age, sex, tobacco smoking, and alcohol use) showed good discriminative accuracy in predicting the absolute five-year risk of ESCC in a Western population.

8 FUTURE PERSPECTIVES

The decreasing incidence of ESCC is probably, at least partly, due to the reduced prevalence of tobacco smoking. Avoiding tobacco and alcohol overconsumption should also be emphasized in women. Etiological patterns of ESCC are more complex in Eastern countries, indicating a need for more research in these populations. Current knowledge of molecular characterization in ESCC is limited. Recent developments in molecular detecting techniques and methodological improvements could facilitate a deeper insight into the molecular mechanisms of ESCC, e.g., by omics studies such as genomics, microbolomics, and metabolomics.

Another potential primary prevention of ESCC might be chemoprevention, such as medication with metformin, NSAIDs/aspirin or statins, albeit more observational research in other populations and randomized clinical trials are required.

Early detection of ESCC could be possible by screening or surveillance of high-risk populations. More practical and feasible screening approaches are needed. Although ESCC risk prediction models have been developed to stratify populations into high- or low- risk groups, none of them have been externally validated. A well-established risk prediction model is still expected. A further investigation of biomarkers, such as alterations of serum microRNAs before ESCC diagnosis, may improve the risk prediction.

Additionally, more clinical studies regarding standardized treatments and follow-up patterns of the esophageal squamous dysplasia (ESCC precursor) are essential for an established and valid clinical guideline.

9 POPULAR SCIENCE SUMMARIES

9.1 Popular science summary

Esophageal cancer is a deadly cancer, with more than 80% of esophageal cancer patients died within five years after diagnosis. Esophageal squamous cell carcinoma (ESCC) is the main subtype of esophageal cancer. Because patients with early-stage tumors rarely present any typical symptoms, more than half of the ESCC patients are diagnosed at a late stage when the tumor is too advanced to allow curative treatment. Prevention and diagnosis at an earlier stage would substantially reduce the mortality in ESCC. Thus this thesis aims to provide a better understanding of possible measures for such prevention and early detection.

In Study I, we collected data on ESCC from 30 cancer registries in 20 countries between 1970 and 2015. We calculated the occurrence rate in different countries using the age-standardized method, which makes the rates comparable. For the year 2012, Japan had the highest occurrence of ESCC. During the past decades, most countries showed decreasing trends of ESCC rates, although increasing rates were found in women in Japan, the Netherlands, New Zealand, Norway, and Switzerland. We examined the reasons behind the change in the rates and found that the increased rates in women in these countries could be due to the increased use of smoking and alcohol drinking, or other factors related to an unhealthy lifestyle.

Study II is a pooled analysis of studies that have investigated the association between smoking and ESCC. By searching in databases, we identified 41 relevant studies among 15,009 publications. Compared with never smokers, current smokers had a four times higher risk of ESCC, while former smokers had a two times higher risk of ESCC. The benefits of stopping smoking were strong in Western populations but weak in the Asian populations. After smoking cessation, the risk of ESCC decreased by 41% after five-to-nine years, 58% after 10-20 years, and 66% after over 20 years. The latter category had a risk similar to never smokers.

In Study III, we assessed whether the use of metformin, a medicine commonly used to treat diabetes, prevents ESCC. We linked data from the Swedish Prescribed Drug Registry, Patient Registry, Cancer Registry, and Cause of Death Registry in 2005-2015. Among 8.4 million participants, 411,603 (5%) used metformin. Compared with ten times as many participants who did not use metformin, those who used metformin had a 32% decreased risk of developing ESCC.

In Study IV, we compared 167 ESCC patients with 820 control participants between 1995 and 1997 in Sweden. All participants were interviewed and their lifestyle information was collected. Then we developed a model that can be used to estimate an individual's absolute risk of developing ESCC during the next five years. According to information on age, sex, tobacco smoking, abuse of alcohol, years of education, duration of living with a partner, and place of residence during childhood, a model worked well in identifying ESCC patients.

Another simpler model was developed with only four factors (age, sex, tobacco smoking, and abuse of alcohol), which was almost as good as the first model. A calculator is available online for individuals who are interested in knowing their own risk of ESCC in the next five years.

In summary, although the occurrence of ESCC is decreasing, it remains common cancer globally. It is possible to prevent ESCC by quitting smoking. Using metformin might also reduce the ESCC risk, but this result has to be confirmed in future studies. A simple risk prediction model can help identify people at high risk of developing ESCC and may help them seek healthcare earlier and detect ESCC at an early and curable stage.

9.2 Populärvetenskaplig sammanfattning

Matstrupscancer är en dödlig cancer och mer än 80 % av patienter dör inom fem år efter diagnosen. Skivepitelcancer (ESCC) är den vanligaste typen av matstrupscancer. Eftersom patienter med tumör i tidigt stadium sällan uppvisar symtom, diagnostiseras de flesta ESCC-patienter i ett sent skede när tumören redan är för avancerad för att möjliggöra botande behandling. Diagnos i ett tidigare skede skulle minska dödligheten i ESCC betydligt. Denna avhandling syftar till att ge en bättre förståelse för möjliggöra förebyggande åtgärder och tidigare upptäckt.

I studie I använde vi 30 cancerregister från 20 länder och samlade in data om patienter med ESCC mellan 1970 och 2015. Sedan beräknade vi förändringar i förekomst i olika länder med hjälp av en åldersstandardiserad metod, vilket gör länderna dem jämförbara. År 2012 hade Japan den högsta förekomsten av ESCC. Under de senaste decennierna visade de flesta länderna minskade trender för ESCC även om ökande frekvenser hittades hos kvinnor i Japan, Nederländerna, Nya Zeeland, Norge och Schweiz. Vi undersökte orsakerna till förändringarna över tid och fann att den ökade förekomsten i vissa länder kan bero på ökad rökning, alkoholkonsumtion eller andra ohälsosamma livsstilsfaktorer.

Studie II är en analys av 41 publicerade artiklar om sambandet mellan rökning och ESCC. Dessa studier fann vi genom att söka i databaser bland 15,009 publikationer. Jämfört med personer som aldrig rökt hade nuvarande rökare fyra gånger högre risk för ESCC, medan tidigare rökare hade två gånger högre risk för ESCC. Fördelarna med att sluta röka var starka i västerländska befolkningar men svaga i asiatiska befolkningarna. När nuvarande rökare slutade röka minskade risken för ESCC med 41 % efter fem till nio år, 58 % efter tio till tjugo år och 66 % efter över 20 år. Den senare kategorin hade en nästan lika låg risk för ESCC som de som aldrig rökt.

I studie III undersökte vi om användning av metformin, ett läkemedel som ofta används för att behandla diabetes, minskar risken för ESCC. Vi använde data från Läkemedelsregistret, Patientregistret, Cancerregistret och Dödsorsaksregistret mellan åren 2005 och 2015. Av 8.4 miljoner personer använde 411,603 (5 %) metformin. Jämfört med tio gånger så många deltagare som inte använde metformin, hade de som använde metformin 32 % minskad risk för att utveckla ESCC.

I studie IV deltog 167 ESCC-patienter och 820 kontrollpersoner från befolkningen i en studie som via intervjuer samlade in uppgifter under åren 1995 och 1997 i Sverige. Med hjälp av informationen utvecklade vi en modell som kan användas för att beräkna individers risk för att utveckla ESCC under de kommande fem åren beroende på ålder, kön, tobaksrökning, missbruk av alkohol, utbildningsår, sammanboende med en partner och bostadområde under barndomen. En annan enklare modell utvecklades också med bara fyra faktorer (ålder, kön, tobaksrökning och missbruk av alkohol). Modellerna var bra på att utkristallisera högriskindivider, och den enkla modellen var nästan lika bra som den första modellen. En

kalkylator är tillgänglig online för personer som är intresserade av att få veta till sin egen risk för ESCC.

Sammanfattningsvis är ESCC en globalt sett ganska vanlig cancer även om dess förekomst lyckligtvis är på nedgående. Det är möjligt att minska risken för ESCC genom att sluta röka. Att ta läkemedlet metformin kan också minska ESCC-risken, men detta resultat måste bekräftas i framtida studier. Modellerna för riskprognos kan användas för att identifiera personer med hög risk för ESCC och kan hjälpa dessa att söka vård tidigt och därmed upptäcka ESCC i ett tidigare och botbart skede.

9.3 科普性总结

食管癌死亡率极高，80%以上的患者的生存时间短于五年。食管鳞癌是其最主要的亚型。鉴于肿瘤早期症状隐匿，绝大多数食管鳞癌患者在被诊断出来时已是中晚期，失去了根治的机会。因此，早期发现并预防食管鳞癌是降低其死亡率的重要策略。本论文旨在探寻其可行的预防和早发现措施。

第一部分研究中，我们收集了全球 20 个国家的 30 个癌症统计局在 1970 年至 2015 年间的食管鳞癌发病数据，用年龄标准化方法计算发病率用于不同国家地区间的比较。研究发现，2012 年，日本在所有纳入国家中发病率最高。在过去的几十年中，大多数国家的食管癌发病率呈下降趋势。然而，日本、荷兰、新西兰、挪威和瑞士的女性食管鳞癌发病率却呈现上升趋势。研究发现这些国家的女性发病率上升可能与增加的吸烟，饮酒或其他不健康生活方式有关。

第二部分研究是针对戒烟和食管鳞癌相关性开展的对已发表研究的汇总分析。通过搜索相关数据库，我们从 15009 篇文章中纳入了 41 个符合要求的研究。我们发现，目前吸烟者较从不吸烟者患食管鳞癌的风险高四倍，而过去吸烟者较从不吸烟者患食管鳞癌的风险高两倍。西方人戒烟后达到的降低食管鳞癌的风险收益比亚洲人更高。若目前吸烟者戒烟，那么其在五至九年内发生食管鳞癌的风险将下降 41%，十年至二十年内的风险下降 58%，二十年后更是下降 66%达到与从不吸烟者相似的风险。

第三部分研究中，我们旨在评估二甲双胍（一种治疗糖尿病的常见药物）在预防食管鳞癌发生中的作用。我们收集了在 2005 年至 2015 年期间瑞典统计局的相关数据，包其主要来源于处方药注册统计局，患者统计局，癌症统计局和死因统计局。在 840 万人群中，有 411603（5%）瑞典人使用过二甲双胍。研究发现，使用二甲双胍的人患食管鳞癌的风险比不使用该药物的人低 32%。

第四部分研究中，我们对比了 1995 至 1997 年期间的 167 例食管鳞癌患者和 820 例健康个体（非食管鳞癌患者），采访并收集其相关信息。我们利用这组数据建立了一个模型用于评估个体五年内发生食管鳞癌的可能性。根据年龄，性别，吸烟史，饮酒史，受教育年限，与伴侣共同生活的时间及童年期间的居住地等信息，该模型能够较好地预测食管鳞癌的发生。同时，我们仅用年龄，性别，吸烟史和饮酒史，建立了一个简单模型，该模型同样具有良好的预测效果。任何感兴趣了解自己相关风险的人可以使用我们的在线计算器进行估算。

综上所述，虽然食管鳞癌的发病率在下降，它仍然是目前最常见的肿瘤之一。戒烟可以降低食管鳞癌的发生风险。而服用二甲双胍类药物也可能会预防食管鳞癌的发生，但是该结论还需要被更多的研究加以证实。本文建立的风险预测模型可用于筛选出患食管鳞癌的高风险人群，有望帮助他们尽早发现食管鳞癌并接受治疗。

9.4 Résumé des sciences populaires

Le cancer de l'œsophage est un cancer mortel et plus de 80% des patients meurent dans les cinq ans suivant un diagnostic. Le carcinome épidermoïde de l'œsophage (ESCC) est le principal sous-type de cancer de l'œsophage. Comme les patients atteints de tumeurs à un stade précoce présentent rarement des symptômes typiques, plus de la moitié des patients ESCC sont diagnostiqués à un stade tardif, lorsque la tumeur est trop avancée pour permettre un traitement curatif. Vue que la prévention et le diagnostic à un stade précoce réduiraient la mortalité dans les ESCC, cette thèse vise à mieux comprendre les mesures possibles pour la prévention et la détection d'ESCC.

Dans l'étude I, nous avons collecté des données sur l'ESCC dans 30 registres du cancer dans 20 pays et collectées entre 1970 et 2015. Nous avons ensuite calculé le taux d'incidence dans différents pays en utilisant la méthode standardisée par l'âge, ce qui rend les taux comparables. Pour l'année 2012, le Japon avait l'incidence d'ESCC la plus élevée. Au cours des dernières décennies, le taux d'ESCC a eu une tendance à diminuer dans la plupart des pays, bien qu'on ait constaté des taux croissants chez les femmes au Japon, aux Pays-Bas, La Nouvelle-Zélande, la Norvège et la Suisse. Nous avons examiné les raisons de cette augmentation et avons constaté que l'augmentation du taux d'ESCC chez les femmes de ces pays pourrait être due à l'augmentation du taux de tabagisme, de consommation d'alcool ou d'autres modes de vie malsains.

L'étude II est une analyse groupée d'études qui ont examiné l'association entre le tabagisme et l'ESCC. En recherchant dans les bases de données, nous avons identifié 41 études pertinentes parmi 15,009 publications. Par rapport aux personnes n'ayant jamais fumé, les fumeurs actuels présentaient un risque quatre fois plus élevé d'ESCC, tandis que les anciens fumeurs présentaient un risque deux fois plus élevé d'ESCC. Les avantages de l'arrêt du tabagisme étaient importants dans les populations occidentales mais faibles dans les populations asiatiques. Lorsque les fumeurs actuels arrêtent de fumer, le risque d'ESCC diminue de 41% après cinq à neuf ans, de 58% après dix à vingt ans et de 66% après plus de 20 ans. Cette dernière catégorie présentait un risque semblable à celui des personnes n'ayant jamais fumé.

Dans l'étude III, nous avons cherché à évaluer si l'utilisation de la metformine, un médicament utilisé pour traiter le diabète, empêche l'ESCC. Nous avons couplé les données de l'année 2005 à l'année 2015 du Registre Suédois des Médicaments Prescrits, du Registre des Patients, du Registre du Cancer et du Registre des Causes de Décès. Parmi les 8.4 millions de participants, 411,603 (5%) ont utilisé la metformine. Ils présentaient une diminution de 32% du risque d'ESCC, comparativement à dix fois plus de participants qui n'utilisaient pas la metformine.

L'étude IV comparé 167 patients d'ESCC et 820 participants témoins entre 1995 et 1997 en Suède. Tous les participants ont été interviewés pour recueillir des informations. À partir de ces informations, nous avons développé un modèle de prédiction pour estimer le risque

absolu de développer une ESCC au cours des cinq prochaines années. Basé sur des données d'âge, de sexe, de tabagisme, d'abus d'alcool, d'années d'éducation, de durée de vie avec un partenaire et de lieu de résidence pendant l'enfance, le modèle a bien fonctionné pour identifier les patients présentant un ESCC. Un autre modèle plus simple comportant seulement quatre facteurs (âge, sexe, tabagisme et abus d'alcool) a également été développé, qui était presque aussi efficace que le premier modèle. Une calculatrice est disponible en ligne pour les personnes qui souhaitent connaître leur propre risque de développer une ESCC.

En résumé, bien que l'occurence de l'ESCC diminue, il reste un cancer commun à l'échelle mondiale. Il est possible de prévenir l'ESCC en cessant de fumer. La prise de metformine pourrait également réduire le risque d'ESCC, mais ce résultat doit être confirmé dans des études futures. Le modèle de prédiction du risque mis en point pourrait être utilisé pour identifier la population à risque élevé de développer une ESCC, qui seront encouragées à consulter un médecin plus tôt et à faire diagnostiquer leur ESCC à un stade précoce et curable.

10 ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards **Karolinska Institutet** and people who supported me during my doctoral education and helped me grow all the way up, both professionally and personally.

Jesper Lagergren, my main supervisor. Thank you for welcoming me to your distinguished research group and guiding me to the fascinating world of epidemiological and clinical research. Your great insights and endless passion for research, thoroughly critical thinking, excellent scientific writing skills, and outstanding leadership have inspired me a lot. I have gained a great education and courage from each conversation we had. I am incredibly grateful for your strong support of my doctoral education and allowing me to explore the world. And I am extremely proud of being your PhD student and you will always be my idol!

Shaohua Xie, my co-supervisor. Thank you for being such a brilliant and supportive advisor. I appreciate greatly your unflagging patience and extraordinary guidance during my doctoral education. Whenever I was puzzled or felt insecure about the projects, you always gave me the most insightful and timely suggestions. Each of our discussions has been so fruitful and helpful for me. I admire your great knowledge of epidemiology, your enthusiasm, and highly-efficient working style. Your care and assistance bolster me to go further.

Edward L. Giovannucci, my advisor at Harvard. Thank you for offering me the opportunity to do an exchange study in Boston. Your glorious academic contributions, remarkable knowledge, the wisdom of life, kindness and modesty impress me greatly. I am thankful for your help with my career development. Your encouragement and support motivated me a lot. Many thanks for sharing your ideas with me for the “important” question, and I will let you know once I get mine.

Kayoko Hosaka, my mentor. Your positive attitude towards work and life has always been a huge motivation for me. I have gained a lot of strength each time I talked with you. I know you are always there whenever I need someone to talk to. Thank you for being my mentor, showing me the fun life in Sweden, and most importantly, your kindness and support.

Giola Santoni, biostatistician. Thank you for your pedagogical support in statistics and for clarifying many ambiguous concepts for me. I appreciate a lot of your kindness, patience and consideration, and thanks for being my office neighbor, making my work easier and funnier.

Poorna Anandavadivelan, I am so grateful to have met you in our office. You are such a nice and thoughtful friend. Whenever I feel upset, you always encourage me and give me a big hug. Thank you for lighting up my life, and I will forever treasure our time together.

Eva Doorakkers, thank you for being my good friend. You have introduced me to so many interesting things, which makes my office life much more exciting. You are so kind, strong, and warmhearted. Thanks for being there whenever I share happiness and sorrow with you.

Pernilla Lagergren, Karin Vikström, Kalle Mälberg, and **Cecilia Haddad Ringborg**, thank you for introducing the Swedish culture to me and helping me with the Swedish language. You make me feel at home in Sweden. Thanks.

Eivind Ness-Jensen, Joonas Kauppila, Wentao Li, Mingyang Song, Steven K. Clinton, Lorelei Mucci, NaNa Keum, all collaborators from cancer registries in Study I, and my coauthors and collaborators for this thesis or others not included, thank you for your valuable efforts and professional excellence.

Giulia Marras, Therese Kindåker, and **Ann-Britt Wikström**, our MMK department admins, thanks a lot for your great administrative support during my doctoral education.

Many thanks to dear former and current colleagues in the research groups Upper Gastrointestinal Surgery (UGIS) and Surgical Care Sciences (SCS), specifically **Jiaojiao Zheng, Yangjun Liu, Manar Yanes, Hanna Johans, Asif Johar, Fredrik Mattsson, Dag Holmberg, Zhao Cheng, Wenjing Tao, Eivind Gottlieb, John Maret-Ouda, Anna Schandl, Helen Rosenlund, Sheraz Markar, Sandra Nielsen, Wille Leijonmarck, Johannes Asplund, Sirus Rabbani, Fredrik Klevebro,** and **Karl Wahlin**. You are all so friendly, kind and helpful. Thank you for companying me all these years, creating a cozy and lovely working environment at the office.

Yunfeng Zhou and **Fuxiang Zhou**, my graduate advisors. Thanks a lot for guiding me to the clinical oncology and cancer research area. Your marvelous devotion and care for cancer patients have motivated me greatly. Your words are forever kept in my heart.

Jiang Yue, my undergraduate advisor. Thank you for your patient guidance and great time and effort in my first ever research experiment. Your excellent pedagogical skills have encouraged my curiosity about life research. Thank you.

My beloved Doctoral Student's Association (DSA) and union colleagues: **Eva Hesselmark, Susanne Neumann, Leif Karlsson, Mirco Martino, Sebastian Ols, Hannes Eichner, Nestor Vazquez Bernat, Benedek Bozoky, Chenhong Lin, Henna Salo, Yildiz Kelahmetoglu, Tage Mohammadat, Robin Palmberg,** and **Pil Maria Saugmann**. It is such a great pleasure to work with you, fighting for the same goal. You are all such brilliant PhD students with kind personalities. I will treasure all the laughter you brought and every fun moment we had together. I wish you guys all the best.

Thanks to **all of my dear friends**, for your love and great support. Special appreciation for:

The Mushroom Group: **Jiawei, Jingjing, Yujiao, Zhen, Yunhan, Hao, Zelong, Xin, Muyi, Xiaolei,** and **Mingming**. I am so lucky to have met you here. Thank you for your warm company and all the happiness you bring to me. You are always backing me up and glad to offer resourceful help in need. One day mushroom friends, forever friends in life!

Dear Harvard friends: **Derrick, Jinhee, Siyu, Xing, Zhangyan, Xiaoshuang, Yixing, Liang,** and **Zhangling**. It was such a fantastic experience in Boston together with you all. Your

intelligent brain, easygoing temper, and warm heart encourage me so much. Thank you for making my PhD exchange journey more enjoyable.

Xiangyi, thank you for helping me out of the trap and showing me the beauty of quantum and music. I am impressed by your extraordinary brilliance and artistic talent. Cheers for our friendship!

Donghao, Yafeng, Wei He, Yunlong, Wenyi, Jingru, and **Jiayao**, thank you for sharing your research or life experiences with me. Your support and encouragement inspire me greatly. Thanks.

Jie, Rong, Anqi, Xiangyu, Mingjun, Wenhao, and **Xiaohong,** thank you for being a long-lasting friend with me. Cheers!

Yang, thank you for being my special friend. You have delighted my life in the darkness, and I am so grateful for what you have done for me. Let us hope for the best.

To **my family**,

亲爱的爸爸妈妈和外公外婆，谢谢您们一直以来对我的支持，理解和关心。虽然隔着千山万水，我仍能感受到您们满满的爱。谁言寸草心，报得三春晖，惟愿您们身体安康，顺心如意。也谢谢哥哥嫂嫂和小晨晨，感谢你们同样在背后默默支持我帮助我。在医学的路上，我将继续锲而不舍，孜孜不倦前行。千言万语汇成一句话，我爱您们！

11 REFERENCES

1. Gray H. Gray's anatomy: the classic collector's edition: Avenel, New Jersey: Gramercy Books 1988.

2. Kuo. B, Urma. D. Esophagus - anatomy and development: GI Motility Online, 2006.

3. Eslick GD. Esophageal cancer: a historical perspective. *Gastroenterol Clin North Am* 2009;38:1-15.

4. Qian ZX. Investigation on esophageal cancer in the province of Xinjiang. Collected papers of the second symposium on esophageal carcinoma. *Chinese Acad Med Sci* 1961:74-8.

5. Karamanou M, Markatos K, Papaioannou TG, et al. Hallmarks in history of esophageal carcinoma. *J BUON* 2017;22:1088-91.

6. Lamy L. Étude de statistique clinique de 134 cas de cancer de l'oesophage et du cardia. *Archives des maladies de l'appareil digestif et de la malnutrition* 1910;4:451-6.

7. Gyser EG. De fame lethali ex callosa oesophagi angustia, cum detectis in quibusdam abdominis visceribus attentione dignis phaenomenis: Argentorati 1770.

8. Jain S, Dhingra S. Pathology of esophageal cancer and Barrett's esophagus. *Ann Cardiothorac Surg* 2017;6:99-109.

9. Ferlay J EM, Lam F, Colombet M, et al. Reprinted from "Global Cancer Observatory: Cancer Today." Lyon, France: International Agency for Research on Cancer. 2018. Available from https://gco.iarc.fr/today. Accessed January 15, 2020.

10. Arnold M, Soerjomataram I, Ferlay J, et al. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* 2015;64:381-7.

11. Longo DL, Fauci AS, Kasper DL, et al. Harrison's principles of internal medicine. In: RJ M, ed. Gastrointestinal Tract Cancer. 18 ed. New York: McGraw-Hill Medical Publishing Division.2008: 764–5.

12. Cheifetz AS, Brown A, Curry M, et al. Oxford American Handbook of Gastroenterology and Hepatology: Oxford University Press 2011:106.

13. Daly JM, Fry WA, Little AG, et al. Esophageal cancer: results of an American College of Surgeons Patient Care Evaluation Study. *J Am Coll Surg* 2000;190:562-72.

14. Pennathur A, Gibson MK, Jobe BA, et al. Oesophageal carcinoma. *Lancet* 2013;381:400-12.

15. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med* 2003;349:2241-52.

16. Wang GQ, Abnet CC, Shen Q, et al. Histological precursors of oesophageal squamous cell carcinoma: results from a 13 year prospective follow up study in a high risk population. *Gut* 2005;54:187-92.

17. Codipilly DC, Qin Y, Dawsey SM, et al. Screening for esophageal squamous cell carcinoma: recent advances. *Gastrointest Endosc* 2018;88:413-26.

18. Hamilton SR AL, World Health Organization, International Agency for Research on Cancer. Pathology and Genetics of Tumours of the Digestive System. 1st ed. Lyon: IARC 2000.

19. Shang L, Wang M. Molecular alterations and clinical relevance in esophageal squamous cell carcinoma. *Front Med* 2013;7:401-10.

20. Levine MS, Rubesin SE, Laufer I. Barium esophagography: a study for all seasons. *Clin Gastroenterol Hepatol* 2008;6:11-25.

21. Shimizu Y, Omori T, Yokoyama A, et al. Endoscopic diagnosis of early squamous neoplasia of the esophagus with iodine staining: high-grade intra-epithelial neoplasia turns pink within a few minutes. *J Gastroenterol Hepatol* 2008;23:546-50.

22. Muto M, Minashi K, Yano T, et al. Early detection of superficial squamous cell carcinoma in the head and neck region and esophagus by narrow band imaging: a multicenter randomized controlled trial. *J Clin Oncol* 2010;28:1566-72.

23. Lordick F, Mariette C, Haustermans K, et al. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2016;27:v50-v57.

24. Mehta K, Bianco V, Awais O, et al. Minimally invasive staging of esophageal cancer. *Ann Cardiothorac Surg* 2017;6:110-18.

25. Jiao R, Luo H, Xu W, et al. Immune checkpoint inhibitors in esophageal squamous cell carcinoma: progress and opportunities. *Onco Targets Ther* 2019;12:6023-32.

26. van Hagen P, Spaander MC, van der Gaast A, et al. Impact of a multidisciplinary tumour board meeting for upper-GI malignancies on clinical decision making: a prospective cohort study. *Int J Clin Oncol* 2013;18:214-9.

27. Hu Y, Puri V, Shami VM, et al. Comparative Effectiveness of Esophagectomy Versus Endoscopic Treatment for Esophageal High-grade Dysplasia. *Ann Surg* 2016;263:719-26.

28. Batra R, Malhotra GK, Singh S, et al. Managing Squamous Cell Esophageal Cancer. *Surg Clin North Am* 2019;99:529-41.

29. Sun F, Yuan P, Chen T, et al. Efficacy and complication of endoscopic submucosal dissection for superficial esophageal carcinoma: a systematic review and meta-analysis. *J Cardiothorac Surg* 2014;9:78.

30. Guo HM, Zhang XQ, Chen M, et al. Endoscopic submucosal dissection vs endoscopic mucosal resection for superficial esophageal cancer. *World J Gastroenterol* 2014;20:5540-7.

31. Muto M, Takahashi M, Ohtsu A, et al. Risk of multiple squamous cell carcinomas both in the esophagus and the head and neck region. *Carcinogenesis* 2005;26:1008-12.

32. Lagergren J, Smyth E, Cunningham D, et al. Oesophageal cancer. *Lancet* 2017;390:2383-96.

33. Allum WH, Blazeby JM, Griffin SM, et al. Guidelines for the management of oesophageal and gastric cancer. *Gut* 2011;60:1449-72.

34. Lagergren J, Mattsson F, Zylstra J, et al. Extent of Lymphadenectomy and Prognosis After Esophageal Cancer Surgery. *JAMA Surg* 2016;151:32-9.

35. van der Schaaf M, Johar A, Wijnhoven B, et al. Extent of lymph node removal during esophageal cancer surgery and survival. *J Natl Cancer Inst* 2015;107(5).

36. Ohashi S, Miyamoto S, Kikuchi O, et al. Recent Advances From Basic and Clinical Studies of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 2015;149:1700-15.

37. Ando N, Kato H, Igaki H, et al. A randomized trial comparing postoperative adjuvant chemotherapy with cisplatin and 5-fluorouracil versus preoperative chemotherapy for

localized advanced squamous cell carcinoma of the thoracic esophagus (JCOG9907). *Ann Surg Oncol* 2012;19:68-74.

38. Stahl M, Stuschke M, Lehmann N, et al. Chemoradiation with and without surgery in patients with locally advanced squamous cell carcinoma of the esophagus. *J Clin Oncol* 2005;23:2310-7.

39. Bedenne L, Michel P, Bouche O, et al. Chemoradiation followed by surgery compared with chemoradiation alone in squamous cancer of the esophagus: FFCD 9102. *J Clin Oncol* 2007;25:1160-8.

40. Vellayappan BA, Soon YY, Ku GY, et al. Chemoradiotherapy versus chemoradiotherapy plus surgery for esophageal cancer. *Cochrane Database Syst Rev* 2017;8:CD010511.

41. al-Sarraf M, Martz K, Herskovic A, et al. Progress report of combined chemoradiotherapy versus radiotherapy alone in patients with esophageal cancer: an intergroup study. *J Clin Oncol* 1997;15:277-84.

42. Conroy T, Galais MP, Raoul JL, et al. Definitive chemoradiotherapy with FOLFOX versus fluorouracil and cisplatin in patients with oesophageal cancer (PRODIGE5/ACCORD17): final results of a randomised, phase 2/3 trial. *Lancet Oncol* 2014;15:305-14.

43. Herskovic A, Martz K, al-Sarraf M, et al. Combined chemotherapy and radiotherapy compared with radiotherapy alone in patients with cancer of the esophagus. *N Engl J Med* 1992;326:1593-8.

44. Rustgi AK, El-Serag HB. Esophageal carcinoma. *N Engl J Med* 2014;371:2499-509.

45. Lagergren J, Mattsson F. Diverging trends in recent population-based survival rates in oesophageal and gastric cancer. *PLoS One* 2012;7:e41352.

46. Chau I, Norman AR, Cunningham D, et al. Multivariate prognostic factor analysis in locally advanced and metastatic esophago-gastric cancer--pooled analysis from three multicenter, randomized, controlled trials using individual patient data. *J Clin Oncol* 2004;22:2395-403.

47. Gavin AT, Francisci S, Foschi R, et al. Oesophageal cancer survival in Europe: a EUROCARE-4 study. *Cancer Epidemiol* 2012;36:505-12.

48. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.

49. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87-108.

50. Lin Y, Totsuka Y, He Y, et al. Epidemiology of esophageal cancer in Japan and China. *J Epidemiol* 2013;23:233-42.

51. Wang Z, Tang L, Sun G, et al. Etiological study of esophageal squamous cell carcinoma in an endemic region: a population-based case control study in Huaian, China. *BMC Cancer* 2006;6:287.

52. Gallus S, Bosetti C, Franceschi S, et al. Oesophageal cancer in women: tobacco, alcohol, nutritional and hormonal factors. *Br J Cancer* 2001;85:341-5.

53. Zambon P, Talamini R, La Vecchia C, et al. Smoking, type of alcoholic beverage and squamous-cell oesophageal cancer in northern Italy. *Int J Cancer* 2000;86:144-9.

54. Arnold M, Laversanne M, Brown LM, et al. Predicting the Future Burden of Esophageal Cancer by Histological Subtype: International Trends in Incidence up to 2030. *Am J Gastroenterol* 2017;112:1247-55.

55. Cook MB, Dawsey SM, Freedman ND, et al. Sex disparities in cancer incidence by period and age. *Cancer Epidemiol Biomarkers Prev* 2009;18:1174-82.

56. Xie SH, Rabbani S, Petrick JL, et al. Racial and Ethnic Disparities in the Incidence of Esophageal Cancer in the United States, 1992-2013. *Am J Epidemiol* 2017;186:1341-51.

57. Gomez SL, Le GM, Clarke CA, et al. Cancer incidence patterns in Koreans in the US and in Kangwha, South Korea. *Cancer Causes Control* 2003;14:167-74.

58. Edgren G, Adami HO, Weiderpass E, et al. A global assessment of the oesophageal adenocarcinoma epidemic. *Gut* 2013;62:1406-14.

59. Castro C, Bosetti C, Malvezzi M, et al. Patterns and trends in esophageal cancer mortality and incidence in Europe (1980-2011) and predictions to 2015. *Ann Oncol* 2014;25:283-90.

60. Pandeya N, Olsen CM, Whiteman DC. Sex differences in the proportion of esophageal squamous cell carcinoma cases attributable to tobacco smoking and alcohol consumption. *Cancer Epidemiol* 2013;37:579-84.

61. Abnet CC, Arnold M, Wei WQ. Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 2018;154(2):360-73.

62. Xie SH, Lagergren J. Risk factors for oesophageal cancer. *Best Pract Res Clin Gastroenterol* 2018;36-37:3-8.

63. Brown LM, Hoover R, Silverman D, et al. Excess incidence of squamous cell esophageal cancer among US black men: Role of social class and other risk factors. *American Journal of Epidemiology* 2001;153:114-22.

64. Wang JB, Jiang Y, Wei WQ, et al. Estimation of cancer incidence and mortality attributable to smoking in China. *Cancer Cause Control* 2010;21:959-65.

65. Lubin JH, Cook MB, Pandeya N, et al. The importance of exposure rate on odds ratios by cigarette smoking and alcohol consumption for esophageal adenocarcinoma and squamous cell carcinoma in the Barrett's Esophagus and Esophageal Adenocarcinoma Consortium. *Cancer Epidemiol* 2012;36:306-16.

66. Pandeya N, Williams GM, Sadhegi S, et al. Associations of duration, intensity, and quantity of smoking with adenocarcinoma and squamous cell carcinoma of the esophagus. *Am J Epidemiol* 2008;168:105-14.

67. Steevens J, Schouten LJ, Goldbohm RA, et al. Alcohol consumption, cigarette smoking and risk of subtypes of oesophageal and gastric cancer: a prospective cohort study. *Gut* 2010;59:39-48.

68. Ishiguro S, Sasazuki S, Inoue M, et al. Effect of alcohol consumption, cigarette smoking and flushing response on esophageal cancer risk: a population-based cohort study (JPHC study). *Cancer Lett* 2009;275:240-6.

69. Rafiq R, Shah IA, Bhat GA, et al. Secondhand Smoking and the Risk of Esophageal Squamous Cell Carcinoma in a High Incidence Region, Kashmir, India: A Case-control-observational Study. *Medicine (Baltimore)* 2016;95:e2340.

70. Yuan JM, Knezevich AD, Wang R, et al. Urinary levels of the tobacco-specific carcinogen N'-nitrosonornicotine and its glucuronide are strongly associated with esophageal cancer risk in smokers. *Carcinogenesis* 2011;32:1366-71.

71. Engel LS, Chow WH, Vaughan TL, et al. Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst* 2003;95:1404-13.

72. Wang JB, Fan JH, Liang H, et al. Attributable causes of esophageal cancer incidence and mortality in China. *PLoS One* 2012;7:e42281.

73. Castellsague X, Munoz N, De Stefani E, et al. Independent and joint effects of tobacco smoking and alcohol drinking on the risk of esophageal cancer in men and women. *Int J Cancer* 1999;82:657-64.

74. Allen NE, Beral V, Casabonne D, et al. Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst* 2009;101:296-305.

75. Joo MK, Park JJ, Chun HJ. Additional Benefits of Routine Drugs on Gastrointestinal Cancer: Statins, Metformin, and Proton Pump Inhibitors. *Dig Dis* 2018;36:1-14.

76. Li P, Cheng R, Zhang S. Aspirin and esophageal squamous cell carcinoma: bedside to bench. *Chin Med J (Engl)* 2014;127:1365-9.

77. Jankowski JAZ, de Caestecker J, Love SB, et al. Esomeprazole and aspirin in Barrett's oesophagus (AspECT): a randomised factorial trial. *Lancet* 2018;392:400-08.

78. Corley DA, Kerlikowske K, Verma R, et al. Protective association of aspirin/NSAIDs and esophageal cancer: a systematic review and meta-analysis. *Gastroenterology* 2003;124:47-56.

79. Bosetti C, Rosato V, Gallus S, et al. Aspirin and cancer risk: a quantitative review to 2011. *Ann Oncol* 2012;23:1403-15.

80. Zhi H, Wang L, Zhang J, et al. Significance of COX-2 expression in human esophageal squamous cell carcinoma. *Carcinogenesis* 2006;27:1214-21.

81. Choi JH, Lee SH, Huh G, et al. The association between use of statin or aspirin and pancreatic ductal adenocarcinoma: A nested case-control study in a Korean nationwide cohort. *Cancer Med* 2019;8:7419-30.

82. Wang K, Gerke TA, Chen X, et al. Association of statin use with risk of Gleason score-specific prostate cancer: A hospital-based cohort study. *Cancer Med* 2019;8:7399-407.

83. Alexandre L, Clark AB, Bhutta HY, et al. Statin use is associated with reduced risk of histologic subtypes of esophageal cancer: a nested case-control analysis. *Gastroenterology* 2014;146:661-8.

84. Pawalowska M, Markowska A. The influence of metformin in the etiology of selected cancers. *Wspolczesna Onkol* 2012;16:223-29.

85. Pawalowska M, Markowska A. The influence of metformin in the etiology of selected cancers. *Contemp Oncol (Pozn)* 2012;16:223-9.

86. Becker C, Meier CR, Jick SS, et al. Case-control analysis on metformin and cancer of the esophagus. *Cancer Causes Control* 2013;24:1763-70.

87. Lee MS, Hsu CC, Wahlqvist ML, et al. Type 2 diabetes increases and metformin reduces total, colorectal, liver and pancreatic cancer incidences in Taiwanese: a representative population prospective cohort study of 800,000 individuals. *BMC Cancer* 2011;11:20.

88. Tseng CH. Metformin and esophageal cancer risk in Taiwanese patients with type 2 diabetes mellitus. *Oncotarget* 2017;8:18802-10.

89. Ruiter R, Visser LE, van Herk-Sukel MP, et al. Lower risk of cancer in patients on metformin in comparison with those on sulfonylurea derivatives: results from a large population-based follow-up study. *Diabetes Care* 2012;35:119-24.

90. Yang J, Wei WQ, Niu J, et al. Cost-benefit analysis of esophageal cancer endoscopic screening in high-risk areas of China. *World J Gastroenterol* 2012;18:2493-501.

91. Wang KK, Wongkeesong M, Buttar NS. American Gastroenterological Association technical review on the role of the gastroenterologist in the management of esophageal carcinoma. *Gastroenterology* 2005;128:1471-505.

92. Liu MF, 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.

93. Chang J, Huang Y, Wei L, et al. Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. *Carcinogenesis* 2013;34:1782-6.

94. Wei WQ, Abnet CC, Lu N, et al. Risk factors for oesophageal squamous dysplasia in adult inhabitants of a high risk region of China. *Gut* 2005;54:759-63.

95. Etemadi A, Abnet CC, Golozar A, et al. Modeling the risk of esophageal squamous cell carcinoma and squamous dysplasia in a high risk area in Iran. *Arch Iran Med* 2012;15:18-21.

96. International Agency for Research on Cancer. Cancer Incidence in Five Continents Volumes I to X, Indices of Data Quality, 2017.

97. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Smokeless Tobacco and Some Tobacco-specific N- Nitrosamines. *IARC Monogr Eval Carcinog Risks Hum* 2007;89:1-641.

98. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum* 2004;83:1-1438.

99. Xie SH, Santoni G, Mattsson F, et al. Cohort profile: the Swedish Prescribed Drugs and Health Cohort (SPREDH). *BMJ Open* 2019;9:e023155.

100. Wettermark B, Hammar N, Fored CM, et al. The new Swedish Prescribed Drug Register-opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007;16:726-35.

101. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.

102. Barlow L, Westergren K, Holmberg L, et al. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncol* 2009;48:27-33.

103. Lindblad M, Ye W, Lindgren A, et al. Disparities in the classification of esophageal and cardia adenocarcinomas and their influence on reported incidence rates. *Ann Surg* 2006;243:479-85.

104. Brooke HL, Talback M, Hornblad J, et al. The Swedish cause of death register. *Eur J Epidemiol* 2017;32:765-73.

105. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016;31:125-36.

106. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997;16:791-801.

107. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.

108. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.

109. GA Wells BS, D O'Connell, J Peterson, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed October 19, 2016.

110. Pan A, Wang Y, Talaei M, et al. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol* 2015;3:958-67.

111. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. [updated March 2011] Available from www.cochrane-handbook.org: The Cochrane Collaboration. Accessed December 22, 2015.

112. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.

113. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006;15:291-303.

114. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38.

115. Ludvigsson JF, Haberg SE, Knudsen GP, et al. Ethical aspects of registry-based research in the Nordic countries. *Clin Epidemiol* 2015;7:491-508.

116. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3 ed. Philadephia, PA, USA: Lippincott Williams & Wilkins 2008.

117. Ahn E, Kang H. Introduction to systematic review and meta-analysis. *Korean J Anesthesiol* 2018;71:103-12.

118. Zheng T, Boffetta P, Boyle P. Epidemiology and Biostatistics. Lyon, France: International Prevention Research Institute 2011.

119. Institute for Health Metrics and Evaluation. Tobacco Visualization: Seattle, WA: IHME, University of Washington, 2017.

120. World Health Organization. Global Health Observatory data repository: Recorded alcohol per capita consumption, 1980-1999 by country. Available from http://apps.who.int/gho/data/node.main.A1024?lang=en&showonly=GISAH. Accessed July 7, 2017.

121. La Vecchia C, Bosetti C, Bertuccio P, et al. Trends in alcohol consumption in Europe and their impact on major alcohol-related cancers. *Eur J Cancer Prev* 2014;23:319-22.

122. Center for Cancer Control and Information Services. Cancer Statistics in Japan-2016: Trends in Consumption of Tobacco, Alcohol and Food. Available from http://ganjoho.jp/en/professional/statistics/brochure/2016_en.html. Accessed July 7, 2017.

123. Oze I, Matsuo K, Wakai K, et al. Alcohol drinking and esophageal cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Jpn J Clin Oncol* 2011;41:677-92.

124. Oze I, Matsuo K, Ito H, et al. Cigarette Smoking and Esophageal Cancer Risk: An Evaluation Based on a Systematic Review of Epidemiologic Evidence Among the Japanese Population. *Jpn J Clin Oncol* 2012;42:63-73.

125. Kitagawa Y, Uno T, Oyama T, et al. Esophageal cancer practice guidelines 2017 edited by the Japan Esophageal Society: part 1. *Esophagus* 2019;16:1-24.

126. Cheng KC, Chen YL, Lai SW, et al. Risk of esophagus cancer in diabetes mellitus: a population-based case-control study in Taiwan. *BMC Gastroenterol* 2012;12:177.

127. Murff HJ, Roumie CL, Greevy RA, et al. Metformin use and incidence cancer risk: evidence for a selective protective effect against liver cancer. *Cancer Causes Control* 2018;29:823-32.

128. Giovannucci E, Harlan DM, Archer MC, et al. Diabetes and cancer: a consensus report. *CA Cancer J Clin* 2010;60:207-21.

129. Inoue M, Iwasaki M, Otani T, et al. Diabetes mellitus and the risk of cancer: results from a large-scale population-based cohort study in Japan. *Arch Intern Med* 2006;166:1871-7.

130. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98:683-90.

12 SUPPLEMENTARY DOCUMENTS

Supplementary Table 1a. Cancer staging categories for ESCC

Category	Criteria
T category	
TX	Tumor cannot be assessed
T0	No evidence of primary tumor
Tis	High-grade dysplasia, defined as malignant cells confined by the basement membrane
T1	Tumor invades the lamina propria, muscularis mucosae, or submucosa
T1a	Tumor invades the lamina propria or muscularis mucosae
T1b	Tumor invades the submucosa
T2	Tumor invades the muscularis propria
T3	Tumor invades adventitia
T4	Tumor invades adjacent structures
T4a	Tumor invades the pleura, pericardium, azygos vein, diaphragm, or peritoneum
T4b	Tumor invades other adjacent structures, such as aorta, vertebral body, or trachea
N category	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1–2 regional lymph nodes
N2	Metastasis in 3–6 regional lymph nodes
N3	Metastasis in 7 or more regional lymph nodes
M category	
M0	No distant metastasis
M1	Distant metastasis
G category	
GX	Differentiation cannot be assessed
G1	Well-differentiated. Prominent keratinization with pearl formation and a minor component of nonkeratinizing basal-like cells. Tumor cells are arranged in sheets, and mitotic counts are low
G2	Moderately differentiated. Variable histologic features, ranging from parakeratotic to poorly keratinizing lesions. Generally, pearl formation is absent
G3‡	Poorly differentiated. Consists predominantly of basal-like cells forming large and small nests with frequent central necrosis. The nests consist of sheets or pavement-like arrangements of tumor cells, and occasionally are punctuated by small numbers of parakeratotic or keratinizing cells
L category*	
LX	Location unknown
Upper	Cervical esophagus to lower border of azygos vein
Middle	Lower border of azygos vein to lower border of inferior pulmonary vein
Lower	Lower border of inferior pulmonary vein to stomach, including esophagogastric junction

‡, if further testing of “undifferentiated” cancers reveals a squamous cell component, or if after further testing they remain undifferentiated, categorize as squamous cell carcinoma G3; *, location is defined by epicenter of esophageal tumor.

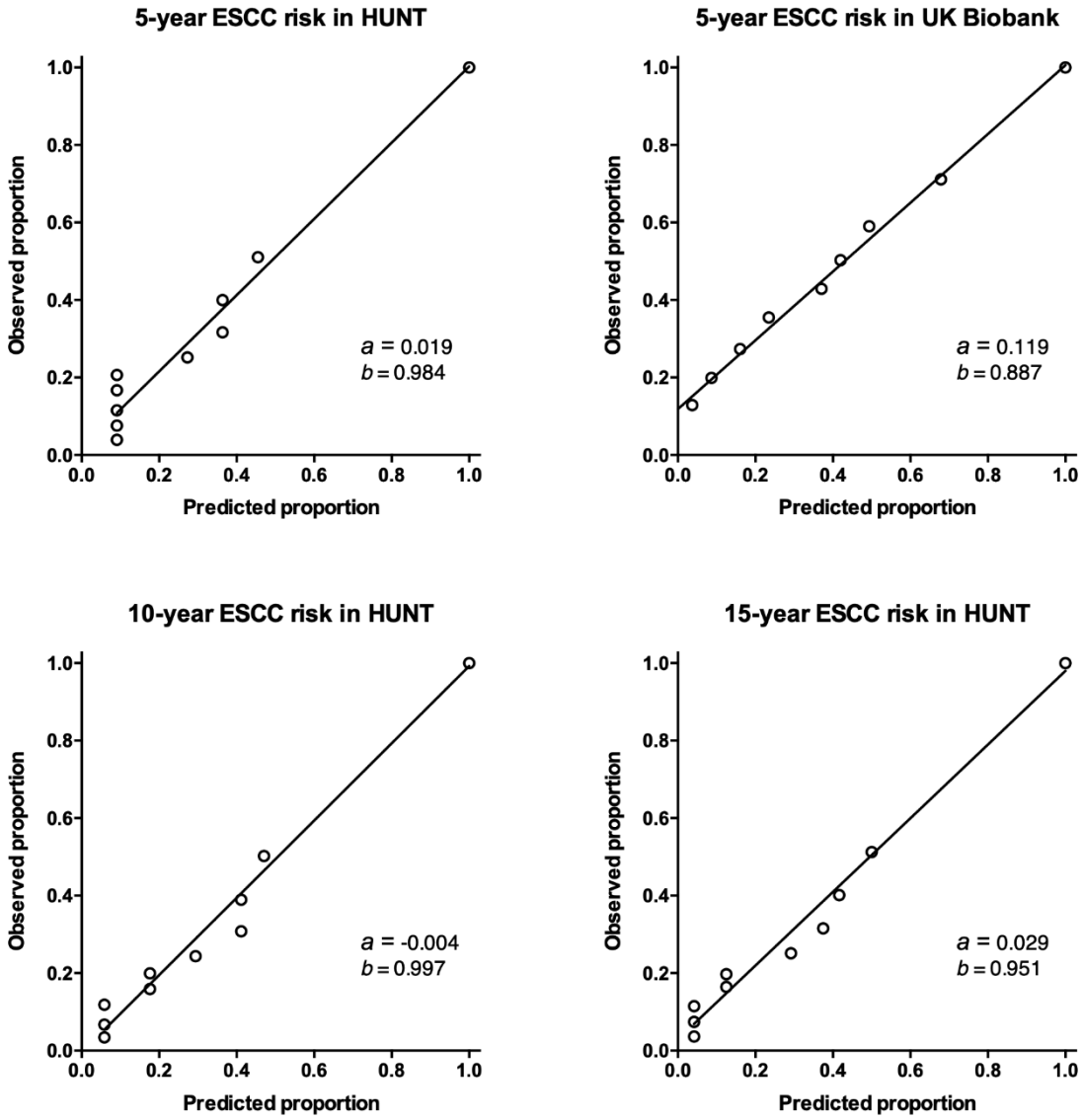
Supplementary Table 1b. Clinical (cTNM) stage groups for ESCC

cStage group	cT	cN	cM
0	Tis	N0	M0
I	T1	N0–1	M0
II	T2	N0–1	M0
	T3	N0	M0
III	T3	N1	M0
	T1–3	N2	M0
IVA	T4	N0–2	M0
	T1–4	N3	M0
IVB	T1–4	N0–3	M1

Supplementary Table 1c. Pathologic (pTNM) stage groups for ESCC

pStage group	pT	pN	pM	pGrade	pLocation
0	Tis	N0	M0	N/A	Any
IA	T1a	N0	M0	G1, X	Any
IB	T1b	N0	M0	G1, X	Any
	T1	N0	M0	G2–3	Any
	T2	N0	M0	G1	Any
IIA	T2	N0	M0	G2–3, X	Any
	T3	N0	M0	Any	Lower
	T3	N0	M0	G1	Upper/middle
IIB	T3	N0	M0	G2–3	Upper/middle
	T3	N0	M0	X	Any
	T3	N0	M0	Any	X
	T1	N1	M0	Any	Any
IIIA	T1	N2	M0	Any	Any
	T2	N1	M0	Any	Any
IIIB	T4a	N0–1	M0	Any	Any
	T3	N1	M0	Any	Any
	T2–3	N2	M0	Any	Any
IVA	T4a	N2	M0	Any	Any
	T4b	N0–2	M0	Any	Any
	T1–4	N3	M0	Any	Any
IVB	T1–4	N0–3	M1	Any	Any

Supplementary Figure 1. Calibration of observed cumulative proportion of ESCC cases and predicted cumulative risk of ESCC in HUNT and UK Biobank cohort



Note *a* denotes the intercept; *b* denote the slope