

# Probability of Transition of Normal Esophageal Mucosa and Precancerous Lesion to the Esophageal Cancer

**Zhiyu Chen**

Sichuan University West China Second University Hospital

**Yong YU**

Chengdu shuangliu district center for disease control and prevention

**Xue YANG**

Sichuan University West China School of Public Health

**Jing-Xuan WANG**

Chengdu shuangliu District Center for Disease Control and Prevention

**Wen-Qiang Wei**

Beijing Union Hospital: Peking Union Medical College Hospital

**You-Lin Qiao**

Beijing Union Hospital: Peking Union Medical College Hospital

**Chun-Xia YANG**

Sichuan University West China School of Public Health

**Li DAI** (✉ [daili@scu.edu.cn](mailto:daili@scu.edu.cn))

Sichuan University West China Second University Hospital <https://orcid.org/0000-0001-6116-8024>

---

## Research

**Keywords:** Esophageal cancer, precancerous lesions, transition probability, Markov models

**Posted Date:** September 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-917455/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** To estimate the transition probabilities of esophageal cancer [EC] and its precancerous lesions by Markov model, which could provide important information for EC screening about choosing reasonable screening and follow-up intervals.

**Methods:** The transition probabilities among pathological stages were estimated by establishing Markov models for the natural history of EC and repeatedly adjusting and calibrating Markov models by comparing the modeled incidence and distributions of pathological stages (alone or combined) with observed data in real-world condition.

**Results:** In one year, the probabilities were 0.024, 0.05, 0.12 for people from health state progressing to mild dysplasia (mD), mild dysplasia (mD) to moderate dysplasia (MD), and moderate dysplasia (MD) to severe dysplasia/carcinoma in situ (SD/CIS), respectively. The age-specific transition probabilities were 0.08~0.18 for severe dysplasia/carcinoma in situ (SD/CIS) progressing to intramucosal carcinoma (IC), 0.4~0.87 for intramucosal carcinoma (IC) to submucosal carcinoma (T<sub>1</sub>N<sub>0</sub>M<sub>0</sub>) (SC), and 0.2~0.85 for submucosal carcinoma (T<sub>1</sub>N<sub>0</sub>M<sub>0</sub>) (SC) to invasive carcinoma (INC). The progression probabilities increased with age and the severity of the disease. Based on the estimated transition probabilities, we predicted the incidence of EC and distributions of its pathological stages. Comparisons between modeled results with observed data confirmed the validation of our transition probabilities.

**Conclusions:** An esophageal cancer transition model in high-risk areas of China has been established with validity. It could be a point of reference for further economic evaluation and policy formulation of esophageal cancer screening.

## Background

Esophageal cancer (EC) is one of the tumors with high morbidity and mortality. According to the International Cancer Registry, there were an estimated 572,000 new cases and 509,000 deaths worldwide in 2018, ranking seventh and sixth in terms of incidence and death of malignant tumors, respectively[1]. China is one of the countries with the highest incidence of EC. According to the latest data released by the National Cancer Center, the incidence of EC in China was 17.87/100,000 and the death rate was 13.68/100,000 in 2015, ranking the sixth and fourth in the incidence and death of malignant tumors respectively[2]. EC affects a wide range of areas in China, covering about 200 million people, mainly in Taihang mountain areas, called the high-risk areas, the mortality of EC reaches 68.3/100 000[3], far more than the national level of 13.68/100 000[2]. More than 95% of the patients treated for EC in the high-incidence areas were in the middle and advanced stages, and the 5-year survival rate of the total cases was less than 10%. However, if early detection, early diagnosis, and early treatment can be achieved, the 5-year survival rate of early EC can be greater than 95%[4]. Accordingly, early detection and treatment through screening programs can reduce the threaten of EC to populations in high-risk areas [5].

It is known that choosing reasonable screening intervals plays an important role in EC screening. And screening intervals are highly associated with the transition probabilities from one health status to another during esophageal carcinogenesis. There have been some follow-up studies in small cohorts of patients with precancerous lesions, and test-retest of endoscopy researches in a short period. They could provide some transition probabilities for parts of stages in the process of EC development. However, the variations of transition probabilities obtained from small populations were significant among studies. Two-fold differences were found in the transition probability of mild dysplasia (mD) progressing to moderate dysplasia (MD) in previous studies [6,7]. And even some probabilities were not available such as the progress probability from intramucosal carcinoma (IC) to submucosal carcinoma (SC)[8]. Thus, the applicability of previous transition probabilities is limited.

As far as we know, large-population-based prospective randomized studies are ethically difficult and expensive to conduct, and results would only be obtained in decades while the Markov model has been successfully used in studies of other cancers to address similar questions [9]. The Markov model is considered a powerful tool for simulating the natural history of chronic diseases. In Markov models, health states passed through by patients are defined separately; and then through modeling based on a system of transition probability among states within a cycle (usually one year), the development of diseases in a population (like disease-specific incidence and mortality) could be estimated [10]. On the other way around, we could estimate transition probability by repeatedly adjusting and calibrating Markov models by comparing the modeled incidence and mortality (alone or combined) with observed data in real-world conditions.

This paper aims to estimate the transition probabilities for EC and its precancerous lesions by establishing the Markov models based on the natural history of EC, which can provide important information for EC screening about choosing reasonable screening and follow-up intervals.

## Methods

### 1. Overview

The transition probabilities among health states of EC were estimated based on established Markov models for the natural history of EC. It is known that the natural history model can predict the incidence of EC based on the EC mortality, all-cause mortality, the prevalence of each pathological stage of EC, and transition probability. Similarly, as the above incidence/mortality/prevalence was available, we could estimate transition probability.

### 2. Study design

Based on the project of early diagnosis and treatment of EC which began in 2005 and is detailed described in our previous paper[11], this study through cluster sampling selected Hejian Town as the screening site in Linzhou County, Henan Province, and the population aged 40-69 in this town as the target screening population, among which those without contraindications for endoscopy were examined by endoscopic iodine staining, and those with positive results were examined by pathology, and their pathological results, age, gender, and other basic information were recorded in detail. For precancerous lesions and esophageal cancer diagnosed by screening, the treatment principals were as follows: (1) for severe dysplasia, carcinoma in situ and intramucosal carcinoma, endoscopic mucosal resection (EMR) or argon plasma coagulation (APC) was recommended; in the first year after treatment, they should be followed up by endoscopy; (2) for submucosal carcinoma (T1N0M0), esophagectomy was recommended; (3) for invasive carcinoma, common treatment modalities were chosen depending on disease severity and could include surgery, radiotherapy, or chemotherapy, or a combination. At the same time, the incidence of EC in each age group every 10 years from 40 to 69 years old in Linzhou County in 2005 and the statistical monitoring data of 321,737 deaths in 2004-2006 were obtained from Linzhou County Cancer Registry.

### 3. Markov models

#### 3.1 Natural history of EC

We developed an 8-state Markov model simulating the natural history of esophageal carcinogenesis: normal (EC-free), mild dysplasia (mD), moderate dysplasia (MD), severe dysplasia/carcinoma in situ (SD/CIS), intramucosal carcinoma (IC), submucosal carcinoma (T1N0M0) (SC), invasive carcinoma (INC), and death[12,13]. The state-transition Markov model based on the natural history of EC has been described in the literature and also showed in Figure 1 [14]. At the start of the model, a hypothetical cohort is distributed in these health states except the "death" state. During each Markov cycle (1 year), a person may remain in the same health state, progress to another state or regress to lesser stages, and die from other reasons or EC. The health state of EC in the next year relied only on the health state of this year and the corresponding transition probabilities [15]. Consider that in the real world, the eligible screening age is 40-69 years and the average life expectancy is 73 years[16], so in the Markov model, we assume that 100,000 participants enter the model at age 40 years and are followed up until age 70 years. TreeAge Pro 2009 Suite by TreeAge Software Inc, was used for all analysis.

#### 3.2 Parameters used in the Markov model

To establish the Markov model of natural history for EC, the following parameters were needed: initial probability, transition probability, and death probability. Initial probability refers to the prevalence of each health state for cohort members at the start of the modeling [17]. Transition probability denotes the likelihood of progression or regression from one health state to another in a Markov cycle. Death probability represents the probability that the population dies from EC or other causes in each model cycle.

#### 3.3 Parameter sources

##### 3.3.1 Initial probability

Initial probabilities were calculated in terms of Linzhou County’s screening results in the project called “Early Detection and Early Treatment of EC in Demonstration Centers in China” (EDETEC) during the years 2005-2008.

The project EDETEC was launched by the Chinese Central Government in 2005, aimed at increasing the early detection and treatment rate, the five-year survival rate of EC, and so forth [18]. Until 2008, 8267 participants aged 40-69 years have received endoscopic and nearly 3000 performed the pathological examination. Most of the esophageal cancer was Squamous carcinoma while only 5% of them were Barrett’s esophagus. The initial distribution probabilities of 7 Markov states except death (from “normal” to “INC”) were respectively 88.95%, 8.2%, 1.8%, 0.9%, 0.08%, 0.05%, and 0.02% in the 40-44 year age group [14].

### 3.3.2 Death probability

The previous study has demonstrated that persons with SD/CIS or lesser may not die from EC; that IC or SC cases may die from all causes including EC; INC patients were assumed to mainly die from EC [14]. Therefore, in the natural-history model, the corresponding death probabilities for the three kinds of the population above were converted from non-esophageal-cancer mortality, all-cause mortality, and case fatality rate of EC, respectively. And they all were obtained from the published data, which were counted according to Linzhou County Cancer Registry during 2004-2006, and the results of our prospective cohort study based on the EC chemoprevention trial of selenomethionine and celecoxib in “Early Detection of EC” (EDEC) program [14,19]. Table 1 demonstrates the age-specific death probabilities of different Markov states.

**Table 1 Death probabilities for each health state of EC\***

| Age | Normal to SD/CIS | IC/SC    | INC      |
|-----|------------------|----------|----------|
| 40- | 0.002270         | 0.002438 | 0.581700 |
| 45- | 0.003073         | 0.003383 | 0.581700 |
| 50- | 0.007054         | 0.007967 | 0.581700 |
| 55- | 0.017061         | 0.019559 | 0.581700 |
| 60- | 0.019744         | 0.021985 | 0.581700 |
| 65- | 0.024105         | 0.027370 | 0.581700 |

\*: death probabilities were obtained from published literature[13]. SD/CIS: severe dysplasia/ carcinoma in situ; IC/SC: intramucosal carcinoma/submucosal carcinoma; INC: invasive carcinoma.

### 3.3.3 transition probability

The transition probabilities among health states were estimated using the approach taken by others [20-23]. Firstly, transition probability ranges were determined from published studies [6,7,24-30], cohort data in the chemoprevention trial of EDEC program mentioned previously [19], and experts’ opinions serving as an initial data set. Due to the limited sample size, the transition probabilities were significantly different among published literature, and even some parameters for progression and regression of the disease were unavailable. Thus, in the second step, the transition probabilities were hierarchically calibrated to make the modeled age-specific EC incidence curves fit the empirical ones observed in real-world settings. In the third step, we further adjusted the transition probabilities to obtain a distribution of each pathological stage for EC similar to that of surveillance data. Observed data from Linzhou County Cancer Registry during 2004-2006 and the screening results in the above EDETEC [18] project during 2005-2008 were used for the model calibration.

### 3.4 Model validation

Incidence and prevalence of EC were highly associated with the transition probabilities among its precancerous lesions. Internal validation of our results was performed by comparing the model predictions with observed epidemiological data from Linzhou County Cancer Registry during 2004-2006 and the screening results in the above EDETEC project during 2005-2008. Validation outcomes included the age-specific incidence of EC and the distribution of pathological stages of EC.

# Results

## 1. transition probabilities

The transition probabilities from a cancer-free state to an EC state and the age-specific transition probabilities from a preclinical cancer state progressing to the next preclinical cancer state are presented in Table 2. According to the Markov model, the sum of the transition probabilities from each initial state to the other states is equal to 1. At the same time, the transition probabilities of IC and SC vary greatly with age. Therefore, for IC and SC, we obtained the transition probabilities of five years old as an age group through consulting experts and repeated fitting. All transition probabilities were within or very close to the related ranges from the literature [6,7,24-30]. The progression probability increased with the severity of the disease. And the related regression probability decreased. The age-specific progression probability for SD/CIS, IC, and SC cases increased with age.

**Table 2 Transition probabilities among pathological stages of esophageal cancer in one year\***

| Transition among health status | Transition probabilities | Ranges from literature              | Transition among health status | Transition probabilities | Ranges from literature          |
|--------------------------------|--------------------------|-------------------------------------|--------------------------------|--------------------------|---------------------------------|
| Normal                         |                          |                                     | IC                             |                          |                                 |
| remaining Normal               | 0.976                    |                                     | remaining IC                   |                          | 0.4344 <sup>[26]</sup>          |
| progression to mD              | 0.024                    | 0.0454-0.0591 <sup>[6,24]</sup>     | 40-44 years old                | 0.6                      |                                 |
| mD                             |                          |                                     | 45-49 years old                | 0.5                      |                                 |
| regression to Normal           | 0.05                     |                                     | 50-54 years old                | 0.4                      |                                 |
| remaining mD                   | 0.9                      |                                     | 55-59 years old                | 0.2                      |                                 |
| progression to MD              | 0.05                     | 0.0868-0.1478 <sup>[6,7]</sup>      | 60-64 years old                | 0.15                     |                                 |
| MD                             |                          |                                     | 65-69 years old                | 0.13                     |                                 |
| regression to mD               | 0.08                     |                                     | progression to SC              |                          |                                 |
| remaining MD                   | 0.8                      |                                     | 40-44 years old                | 0.4                      |                                 |
| progression to SD/CIS          | 0.12                     | 0.0666-0.1478 <sup>[7,24]</sup>     | 45-49 years old                | 0.5                      |                                 |
| SD/CIS                         |                          |                                     | 50-54 years old                | 0.6                      |                                 |
| regression to MD               |                          | 0.1229-0.2212 <sup>[25,26]</sup>    | 55-59 years old                | 0.8                      |                                 |
| 40-44 years old                | 0.17                     |                                     | 60-64 years old                | 0.85                     |                                 |
| 45-49 years old                | 0.15                     |                                     | 65-69 years old                | 0.87                     |                                 |
| 50-54 years old                | 0.14                     |                                     | SC                             |                          |                                 |
| 55-59 years old                | 0.12                     |                                     | remaining SC                   |                          |                                 |
| 60-64 years old                | 0.11                     |                                     | 40-44 years old                | 0.8                      |                                 |
| 65-69 years old                | 0.09                     |                                     | 45-49 years old                | 0.7                      |                                 |
| remaining SD/CIS               |                          | 0.0476-0.5652 <sup>[7,25,27]</sup>  | 50-54 years old                | 0.55                     |                                 |
| 40-44 years old                | 0.75                     |                                     | 55-59 years old                | 0.2                      |                                 |
| 45-49 years old                | 0.75                     |                                     | 60-64 years old                | 0.17                     |                                 |
| 50-54 years old                | 0.74                     |                                     | 65-69 years old                | 0.15                     |                                 |
| 55-59 years old                | 0.74                     |                                     | progression to INC             |                          | 0.097-0.5276 <sup>[28,30]</sup> |
| 60-64 years old                | 0.73                     |                                     | 40-44 years old                | 0.2                      |                                 |
| 65-69 years old                | 0.73                     |                                     | 45-49 years old                | 0.3                      |                                 |
| progression to IC              |                          | 0.074-0.1730 <sup>[24, 28,29]</sup> | 50-54 years old                | 0.45                     |                                 |
| 40-44 years old                | 0.08                     |                                     | 55-59 years old                | 0.8                      |                                 |
| 45-49 years old                | 0.1                      |                                     | 60-64 years old                | 0.83                     |                                 |

|                 |      |                     |                            |
|-----------------|------|---------------------|----------------------------|
| 50-54 years old | 0.12 | 65-69 years old     | 0.85                       |
| 55-59 years old | 0.14 | INC after treatment |                            |
| 60-64 years old | 0.16 | recovering          | 0.7696 <sup>[31]</sup>     |
| 65-69 years old | 0.18 | remaining INC       | 0.2304 <sup>[32, 33]</sup> |

\*: transition probabilities have been used in the health economic evaluation exploring preferable screening strategies for EC in high-risk areas of China [14]. mD: mild dysplasia; MD: moderate dysplasia; SD/CIS: severe dysplasia/ carcinoma in situ; IC: intramucosal carcinoma; SC: submucosal carcinoma; INC: invasive carcinoma.

## 2. Model internal validation

### 2.1 Comparison with Screening data on proportions

Based on the estimated transition probabilities above, we obtained the distribution of pathological stages of EC among the population aged 40-69 years. The validity analysis of EC Markov model was conducted by comparing the model estimates with the actual proportions of the EDETEC project during the period 2005-2008.

The estimated pathological stage distribution of EC is shown in Table 3, which closely approximated the results of the EDETEC project during 2005-2008, within the 99% confidence intervals of screening data. The proportions declined with the severity of the disease, mD stage ranked first (over 15%), followed by MD stage and SD/CIS stage, the proportion of EC stage (including IC, SC, and INC) was the lowest (about 0.8%).

And age-specific distributions of mD, MD, SD/CIS, and EC (including IC, SC, and INC) predicted by modeling were similar to related screening results. Figure 2 shows the proportions increased with age. 65-69 age group ranked first. This trend is consistent with the natural history of the EC and with reports from other high-risk areas in China[34,35].

**Table 3 Comparison of the distributions of EC between modeled results and screening data**

| Age   | mD            |                       | MD            |                       | SD/CIS        |                       | IC/SC/INC     |                       |
|-------|---------------|-----------------------|---------------|-----------------------|---------------|-----------------------|---------------|-----------------------|
|       | Estimated (%) | Screening (%; 99% CI) |
| 40-   | 10.7          | 8.2<br>(2.64-13.76)   | 2.2           | 1.8<br>(1.04-2.61)    | 0.9           | 0.7<br>(0.19-1.13)    | 0.26          | 0.1<br>(0.00-0.29)    |
| 45-   | 15.4          | 14.0<br>(7.85-20.15)  | 3.5           | 3.2<br>(2.04-4.43)    | 1.3           | 1.3<br>(0.51-2.00)    | 0.57          | 0.26<br>(0.00-0.61)   |
| 50-   | 18.1          | 16.8<br>(11.41-22.19) | 4.7           | 4.4<br>(3.16-5.64)    | 1.8           | 1.7<br>(0.95-2.50)    | 0.92          | 0.94<br>(0.37-1.52)   |
| 55-   | 19.6          | 17.8<br>(11.92-23.68) | 5.6           | 5.8<br>(4.25-7.38)    | 2.3           | 2.5<br>(1.45-3.49)    | 1.21          | 0.7<br>(0.15-1.24)    |
| 60-   | 20.6          | 20.6<br>(12.82-28.38) | 6.3           | 7.7<br>(5.26-10.09)   | 2.6           | 2.5<br>(1.47-4.49)    | 1.5           | 1.15<br>(0.00-2.08)   |
| 65-   | 21.3          | 17.4<br>(5.95-28.85)  | 6.6           | 5.5<br>(1.68-6.89)    | 2.9           | 4.7<br>(1.85-7.20)    | 1.9           | 2.86<br>(0.73-4.99)   |
| total | 15.4          | 14.9<br>(12.28-17.52) | 4.2           | 4.2<br>(3.65-4.82)    | 1.8           | 1.8<br>(1.42-2.18)    | 0.8           | 0.69<br>(0.44-0.91)   |

“Estimated”: refer to the proportions that were estimated through Markov models simulation. “Screening”: referred to the proportions were calculated based on the “Early Detection and Early Treatment of EC in Demonstration Centers in China” project screening results during 2005-2008. “99%CI”: 99% confidence interval. mD: mild dysplasia; MD: moderate dysplasia; SD/CIS: severe dysplasia/carcinoma in situ; IC/SC/INC: intramucosal carcinoma/submucosal carcinoma/invasive carcinoma.

## 2.2 Comparison with observed data on incidence rates

Table 4 shows the comparison between the incidence of middle and advanced EC obtained by model fitting and the incidence of EC in Linzhou County from 2004 to 2006. The trend of age-specific incidence rates from 40 to 69 years estimated by the Markov model was similar to the report in Linzhou County, the incidence rates increased with age. The estimated value of the model is approximately 97-107% of the observed value. It can be seen that the incidence obtained by model fitting is highly consistent with the Cancer Registry Report in Linzhou County.

**Table 4 The incidence rates of EC in Linzhou County from 2004 to 2006**

| Age | Total population | Number of cases of EC | Incidences (per 100000 individuals) | Estimated per 100000 individuals |
|-----|------------------|-----------------------|-------------------------------------|----------------------------------|
| 40- | 440494           | 209                   | 47.44                               | 46.19                            |
| 50- | 277323           | 687                   | 247.77                              | 248.19                           |
| 60- | 156281           | 622                   | 398.00                              | 425.19                           |

“Estimated”: refer to the incidences that were estimated through Markov models simulation.

## Discussion

To our knowledge, our study is the first to comprehensively present transition probabilities for the natural history of EC based on the state-transition Markov model for further evaluation of EC screening projects. The transition probabilities among some health states estimated in our paper were partly different from published data. Most previous studies on the natural history of EC were conducted by repeated endoscopic screening or follow-up observation of precancerous lesions, but these methods require a long time, and the probability between some states might not be obtained. Moreover, the natural transition probabilities for some precancerous lesions were ethically difficult to get. Although previous studies have reported some transition probabilities for parts of pathological stages of EC, the data were significantly different among studies mainly owing to the small sample size [6, 7, 24–30]. In contrast, the Markov model has been successfully used to simulate the natural history of some malignant tumors [36, 37], so we tried to use it to simulate the natural history of EC and estimate the transition probabilities between its various states.

The Markov model is a useful method to describe the process of individuals passing through a series of states in continuous time. These types of patterns explain how patients transition between long-term disease states, which is popular in explaining the natural history of the disease [38, 39]. Moreover, the Markov model has another advantage that it can utilize data with limitations in quality and accuracy, which may be collected retrospectively [40]. In the present study, we constructed a Markov model to estimate the 1-year transition probabilities of various health states of EC (normal, mD, MD, SD/CIS, IC, SC, and INC). In particular, we fitted the age-specific proportion of EC at each pathological stage and predicted the incidence rates of EC in our model, which were finally compared with the screening data of Linzhou County.

In this study, we established the Markov model for the natural history of EC based on the transition probabilities. We set the disease (EC) to progress only one stage in a year, without considering the possibility that a few individuals could develop two stages in a year, but the proportion of such individuals in the population is very small. And we also tried to set up such a model in which the probabilities were very small, they had little effect on the outcome. Compared with the transition probabilities that can be found in the literature, the fitting transition probabilities in this study are relatively consistent except for the transition probability between healthy to mD and mD to MD (the sample size of the literature is small). But due to the small sample size of the early literature, the stability and generalization of the results may be limited. Besides, the age proportion of mD, MD, SD/CIS, IC, SC, and INC simulated according to the transition probabilities fitted by the model is close to the real proportion obtained by actual screening. And the transition probabilities of the model can still have such a good consistency after a cycle of 30 cycles. We believe that the model fitting transition probabilities have high credibility and good authenticity.

Most of the data used in our model came from the Linzhou County Cancer Registries, where extensive endoscopic screening has been conducted since the 1980s and established systematic cancer incidence and death registries. Therefore, the relevant data of Linzhou County are reliable. Although the model did not use near two years of data to estimate the transition probabilities of EC (because the model requires many parameters, and the data in the past two years is very limited, and many data cannot be obtained or searched), the change of EC in Linzhou County in recent 10 years was not significant, the transition probabilities in this model were obtained after 30 cycles of estimation, and the incidence of the disease in the middle and late stages of 40–69 years was close to the monitoring results of Linzhou County.

The model was internally validated by comparing the model predictions with epidemiological data from Linzhou County Cancer Registries. Model-predicted age-specific distributions of pathological stages of EC matched well with the screening results of Linzhou county, and the trends were similar to data reported in Ci county, another high-risk area of EC in China [41]. Owing to the inclusion of the younger age group, the results of the Anyang trial were not comparable to our predictions [35]. The age-specific incidence of EC estimated by modeling was quite close to Linzhou County Cancer Registry data. However, further comparisons with other areas were not conducted due to the unavailability of related reported data. In general, the estimated transition probabilities in the study were reliable for high-risk areas in China.

Our study also has several limitations. Firstly, the transition probabilities were estimated based on limited published data and short-term cohort studies. Although model validation results confirmed a relatively high internal validity, the external validity of the model could not be assessed due to the lack of other available data sets. There is insufficient evidence to extend the results of this study to other regions. The results of the model depend on the quality of the registration and screening data to a large extent, which may have a potential risk of deviation. Therefore, long-term cohort studies with a large sample size were needed to test and improve our results. Nevertheless, this kind of epidemiological research is currently absent in China. Secondly, the transition probabilities were estimated based on the data from high-risk areas in China. Differences in social demographic structure and risk factors such as

lifestyle and diet habits across China made it prudent to transfer our results to low-risk areas of EC. Finally, in this study, the transition probabilities of all health states should vary with age. However, due to the small sample size of the literature and the unavailability of data, the transition probabilities of the normal state, mD state, and MD state were fixed, which would affect the results of the model to some extent [14]. Therefore, we should confirm our obtained transition probabilities by following up the natural history of a larger sample of people, and further improve our model of EC natural history and transition probabilities.

## Conclusion

An esophageal cancer transition model based on the natural history of esophageal cancer in high-risk areas in China was established, and the model is effective. This can provide a reference for economic evaluation and policy formulation of esophageal cancer screening.

## Abbreviations

EC: esophageal cancer

mD: mild dysplasia

MD: moderate dysplasia

SD/CIS: severe dysplasia/carcinoma in situ

IC: intramucosal carcinoma

SC: submucosal carcinoma

INC: invasive carcinoma

EMR: endoscopic mucosal resection

APC: argon plasma coagulation

EDETEC: Early Detection and Early Treatment of EC in Demonstration Centers in China

EDEC: Early Detection of EC

## Declarations

### Ethics approval and consent to participate

All procedures performed in studies involving human participants were following the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

### Consent for publication

All authors agreed on the final version of the manuscript.

### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author (e-mail: yangchunxia@scu.edu.cn) on reasonable request.

### Competing interests

All authors declare that there is no conflict of interest.

## Funding

This work was supported by The National Science and Technology Pillar Program of the 11th National Five-Year Plan of China, No.2006BAI02A15. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Authors' contributions

WqW, YIQ, and CxY designed the research and revised the manuscript; YX, and XjW participated in data collection and establishment of esophageal cancer metastasis model; ZyC, YY and LD reviewed the literature, analyzed the data and prepared the paper.

## Acknowledgments

The authors thank the health care workers of Linzhou tumor hospital for providing the field data and declare no competing interests.

## References

1. Bray F, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer Journal for Clinicians*. 2018;68(6):394–424.
2. SUN KX, et al., Report of Cancer Incidence and Mortality in Different Areas of China, 2015. *Cancer in China*, 2019. 28(01): p. 1–11.
3. Zhang SW, Chen WQ, Kong LZ, Li GL, Zhao P. An Annual Report: Cancer Incidence in 35 Cancer Registries in China, 2003. *Bulletin of Chinese Cancer* 2007. 16(7): 494–507.
4. Ciocirlan M, et al. Endoscopic mucosal resection for squamous premalignant and early malignant lesions of the esophagus. *Endoscopy*. 2007;39(1):24–9.
5. Qiao YL, Hou J, Yang L, He YT, Liu YY, Li LD, Li SS, Lian SY, Dong ZW. The Trends and Preventive Strategies of EC in High-risk Areas of Taihang Mountains, China. *ACTA Academiae Medicinae Scientiae*. 2001;23(1):10–4.
6. Wen DG, Zhang LW, Wang SJ, Li YS, Yu WF, Wang XL, Wang JH, Li SP, Li YW, Wang SP, Er LM, Cong QW, Ma CF. Sojourn time observation of esophageal and cardia precancerous lesions by periodic endoscopic screening of 301 subjects in Shexian. *Journal of Zhengzhou University (Medical Science)*. 2007;42(1):36–40.
7. Jacob P, Kahrilas PJ, Desai T, Hidvegi D, Walloch J, Yokoo H, Gurley AM, Ostrow JD. Natural history and significance of esophageal squamous cell dysplasia. *Cancer*. 1990;65(12):2731–9.
8. Wang GQ, Jiao GG, Chang FB, Fang WH, Song JX, Lu N, Lin DM, Xie YQ, Yang L. Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening. *Ann Thorac Surg*. 2004;77(5):1740–4.
9. Liao W, et al. Cost-effectiveness analysis of cabozantinib as second-line therapy in advanced hepatocellular carcinoma. *Liver international: official journal of the International Association for the Study of the Liver*; 2019.
10. Peirola R, Scalerandi M. Markovian model of growth and histologic progression in prostate cancer. *Physical Review E Statistical Nonlinear Soft Matter Physics*. 2004;70(1):011902.
11. Yang J, et al. Estimating the costs of esophageal cancer screening, early diagnosis and treatment in three high risk areas in China. *Asian Pac J Cancer Prev*. 2011;12(5):1245–50.
12. Wang GQ, Qiao YL. Study on precancerous lesions to control the development of esophageal cancer [J]. *Cancer research clinical*. 2003;15(01):3–4.
13. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making*. 1993;13(4):322–38.
14. Yang J, et al. Cost-benefit analysis of esophageal cancer endoscopic screening in high-risk areas of China. *World journal of gastroenterology*. 2012;18(20):2493–501.
15. Wang GQ. Development trend and countermeasures of precancerous lesions of esophagus [J]. *Chin J oncology*. 2002;24(02):44–6.
16. Chai YY, Zhang XY, Mai SMJ, Li Y. Analyzing Chinese residents' death status. *The Chinese health service management*. 2010;5:345–6.

17. Pinsky PF. An early- and late-stage convolution model for disease natural history. *Biometrics*. 2004;60(1):191–8.
18. Wei WQ, Wang JB, Yang J. EC. In: Asian Pacific Organization for Cancer Prevention. *Cancer report 2010*. Ankara: MN Medical & Nobel Publishing Company; 2010. pp. 209–14.
19. Limburg PJ, Wei WQ, Ahnen DJ, Qiao Y, Hawk ET, Wang G, Giffen CA, Wang G, Roth MJ, Lu N, Korn EL, Ma Y, Caldwell KL, Dong Z, Taylor PR, Dawsey SM. Randomized, placebo-controlled, esophageal squamous cell cancer chemoprevention trial of selenomethionine and celecoxib. *Gastroenterology*. 2005;129:863–73.
20. Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. *Am J Epidemiol*. 2000;151(12):1158–71.
21. Jin X, Carlin BP. Multivariate parametric spatiotemporal models for county level breast cancer survival data. *Lifetime Data Anal*. 2005;11(1):5–27.
22. Hsieh HJ, Chen TH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Stat Med*. 2002;21(22):3369–82.
23. Cook AR, Otten W, Marion G, et al. Estimation of multiple transmission rates for epidemics in heterogeneous populations. *Proc Natl Acad Sci USA*. 2007;104(51):20392–7.
24. Wang SJ, Zhang LW, Wen DG, Li YS, Yu WF, Wang XL, Wang JH, Li SP, Ma CF, Li YW, Wang SP, Er LM. Analysis of Endoscopic Screening for the Natural History of Precancerous Lesions in Esophagus and Cardia. *Chinese Journal of Clinical Oncology*. 2007;34(7):370–3.
25. Bai JX, Yin DL, Ding MJ, et al. A study on pathological evolution and gamma 2GT response in patients with severe esophageal cancer after 1 year. *J gastroenterology hepatology*. 1999;8(2):103–4.
26. Benedetti G, Sablich R, Vitalba A, Lacchin T, Guido E, Del Bianco T. Intraepithelial carcinoma of the oesophagus: report of 25 cases from north-east Italy. *Eur J Surg*. 2000;166(8):622–7.
27. Liu YY, Wang GQ, Zhou JL, et al. Observation and comprehensive prevention of precancerous esophagus. *Chinese journal of public health*. 1998;17(4):193–5.
28. Wang GQ, et al. Histological precursors of oesophageal squamous cell carcinoma: results from a 13 year prospective follow up study in a high risk population. *Gut*. 2005;54(2):187–92.
29. Dawsey SM, Lewin KJ, Wang GQ, Liu FS, Nieberg RK, Yu Y, Li JY, Blot WJ, Li B, Taylor PR. Squamous esophageal histology and subsequent risk of squamous cell carcinoma of the esophagus. A prospective follow-up study from Linxian. *China Cancer*. 1994;74(6):1686–92.
30. Wang LD, Yang HH, Fan ZM, Lu XD, Wang JK, Liu XL, Sun Z, Jiang YN, He X, Zhou Q. Cytological screening and 15 years' follow-up (1986–2001) for early esophageal squamous cell carcinoma and precancerous lesions in a high-risk population in Anyang County, Henan Province, Northern China. *Cancer Detection Prevention*. 2005;29(4):317–22.
31. Zhang HL, Ping YM, Du XQ, et al. Analyzed the prognostic factors after resection of esophageal cancer by cox model. *Chinese journal of oncology*. 1999;21(1):2–34.
32. Wang GQ, Jiao GG, Chang FB, Fang WH, Song JX, Lu N, Lin DM, Xie YQ, Yang L. Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening. *Ann Thorac Surg*. 2004;77(5):1740–4.
33. Zhang YW, Xiang JQ, Li FM, Hu H, Miao LS, Ma LF, Li HC, Zhang J. Cause of death and pattern of recurrence after esophagectomy for patients with thoracic esophageal carcinoma. *China Oncology*. 2006;16(10):831–4.
34. Wang G, et al. Significance of screening by iodine staining of endoscopic examination in the area of high incidence of esophageal carcinoma. *Chinese journal of cancer*. 2003;22(2):175–7.
35. He Z, Zhao Y, Guo C, et al. Prevalence and risk factors for esophageal squamous cell cancer and precursor lesions in Anyang, China: a population-based endoscopic survey. *Br J Cancer*. 2010;103(7):1085–8.
36. Gardiner JC, Luo Z, Bradley CJ, Sirbu CM, Given CW. A dynamic model for estimating changes in health status and costs. *Stat Med*. 2006;25(21):3648–67.
37. Siebert U, Sroczynski G, Hillemanns P, Engel J, Stabenow R, Stegmaier C, Voigt K, Gibis B, Hölzel D, Goldie SJ. The German cervical cancer screening model: development and validation of a decision-analytic model for cervical cancer screening in Germany. *Eur J Pub Health*. 2006;16(2):185–92.
38. Jackson C. *Multi-state modelling with R: the msm package*. Cambridge: MRC Biostatistics Unit; 2016.

39. Tada T, et al. Natural history of liver-related disease in patients with chronic hepatitis C virus infection: An analysis using a Markov chain model. *J Med Virol.* 2019;91(10):1837–44.
40. Tanaka J, Kumada H, Ikeda K, Chayama K, Mizui M, Hino K, Katayama K, Kumagai J, Komiya Y, Miyakawa Y, Yoshizawa H. Natural histories of hepatitis C virus infection in men and women simulated by the Markov model. *J Med Virol.* 2003;70(3):378–86.
41. Lu XJ, Chen ZF, Guo CL, et al. Endoscopic survey of EC in a high-risk area of China. *World J Gastroenterol.* 2004;10(20):2931–5.

## Figures

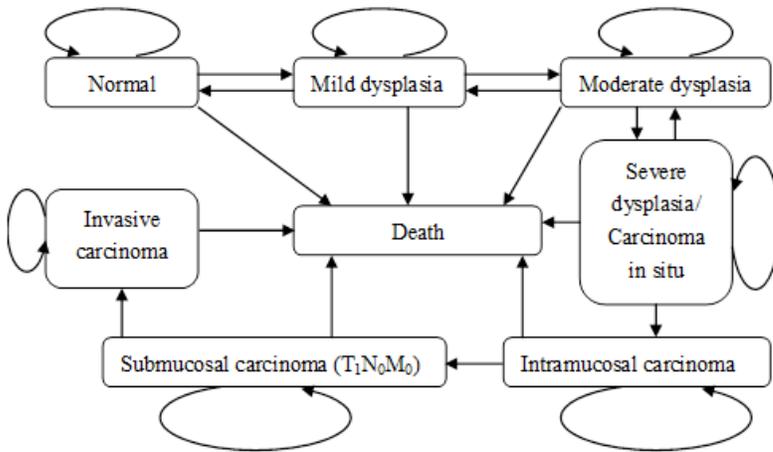
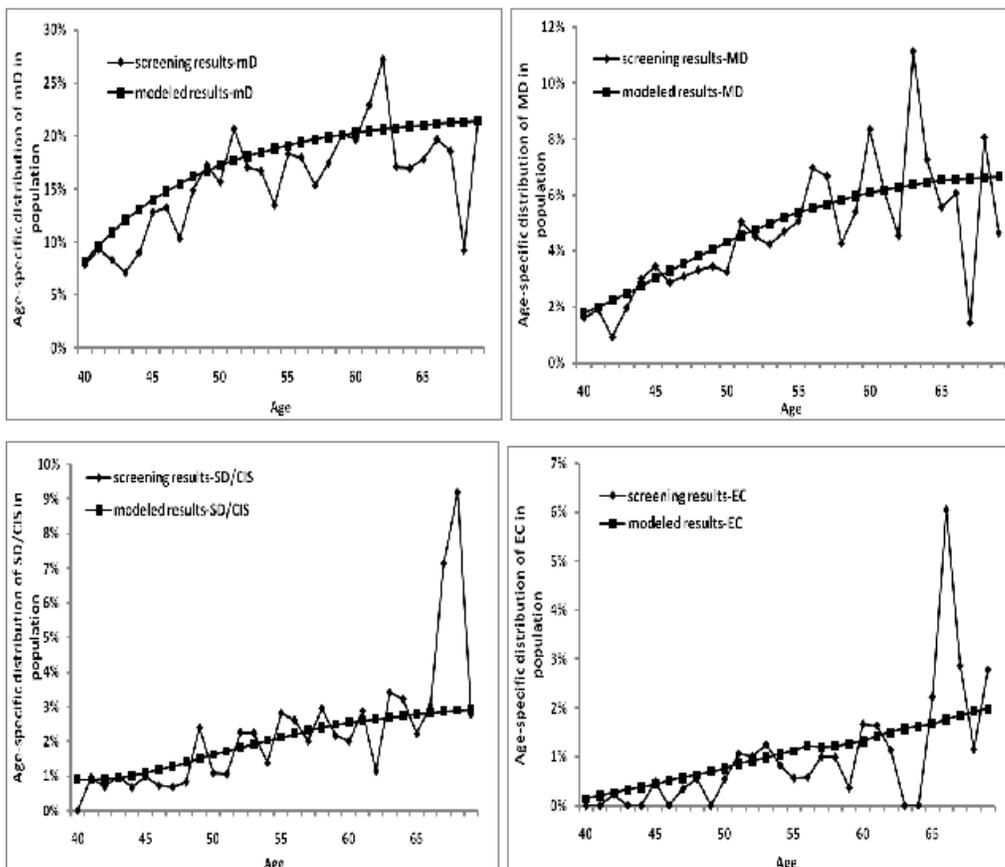


Figure 1

Natural history of EC. Each rectangle: health state; Arrows between different states: from current health state to another; Half-circle arrows on the rectangles: remain in the same state.



## Figure 2

Comparison of the age-specific proportion of EC between modeled results and screening data. mD: mild dysplasia; MD: moderate dysplasia; SD/CIS: severe dysplasia/carcinoma in situ; EC: EC, including intramucosal carcinoma, submucosal carcinoma, and invasive carcinoma.