

Segregation Analysis of Esophageal Cancer in a Moderately High-Incidence Area of Northern China

Weihua Zhang,¹ Joan E. Bailey-Wilson,² Weidong Li¹, Xiuqin Wang,¹ Chunlin Zhang,³ Xuezheng Mao,³ Zhihua Liu¹, Chuannong Zhou¹, and Min Wu¹

¹National Laboratory of Molecular Oncology, Cancer Institute, Chinese Academy of Medical Sciences, Beijing; ²National Human Genome Research Institute, National Institutes of Health, Baltimore; and ³Cancer Institute of Yangquan City, Yangquan City, Shanxi Province, China

In order to explore the mode of inheritance of esophageal cancer in a moderately high-incidence area of northern China, we conducted a pedigree survey on 225 patients affected by esophageal cancer in Yangquan, Shanxi Province. Segregation analysis was performed using the REGTL program of S.A.G.E. The results showed that Mendelian autosomal recessive inheritance of a major gene that influences susceptibility to esophageal cancer provided the best fit to the data. In the best-fitting recessive model, the frequency of the disease allele was .2039. There was a significant sex effect on susceptibility to the disease. The maximum cumulative probability of esophageal cancer among males with the AA genotype was 100%, but, among females, it was 63.5%. The mean age at onset for both men and women was 62 years. The age-dependent penetrances for males with the AA genotype by the ages of 60 and 80 years were 41.6% and 95.2%, respectively, whereas, for females, they were 26.4% and 60.5%, respectively. Incorporating environmental risk factors—such as cigarette smoking, pipe smoking, alcohol drinking, eating hot food, and eating pickled vegetables—into the models did not provide significant improvement of the fit of the models to these data. The results suggest a major locus underlying susceptibility to esophageal cancer with sex-specific penetrance.

Introduction

Unlike in most developed countries, esophageal cancer (EC) (MIM 133239) is a very common disease in many areas of China, especially the north. The disease ranks as the ninth most frequent cancer in the world. In China, it is the fourth most frequent cause of death from malignant tumors. It is said that the proportion of EC from China makes up 70% of all cases of this disease in the world (Parkin et al. 1993; Day and Varghese 1994). Remarkable variations in incidence and sex ratio are seen both between countries and by geographical and ethnic divisions within countries. In China, rates of EC vary more widely than those of any other common cancers, with some areas of remarkably high incidence (Liu and Li 1984). High-incidence areas in northern China are mainly in the areas where the borders of three provinces (Hebei, Henan, and Shanxi) meet on the south side of the Taihang Mountains. The mortality rate can be >100/100,000, such as in Linxian, Henan Province, and

Yangcheng, Shanxi Province. The highest county mortality rates are 254.77/100,000 for males and 161.11/100,000 for females in Linxian, which are 1,075-fold and 671-fold higher than the lowest sex-specific mortality in the country, respectively. The mortality rate in males is higher than that in females. The ratio of males to females can vary from .62 to 9.04 in the country and the nationwide ratio is 1.99 (National Cancer Control Office 1980; Li 1982). Most cases occur between the ages of 60 and 64 years. Since the prognosis of EC is very poor, prevention of this disease is very important.

Epidemiological studies have shown that certain environmental risk factors are associated with EC, though the odds ratios (ORs) vary considerably from study to study. The etiology of EC is related to a variety of smoking and alcohol-exposure history. Other factors include nutritional deficiencies, eating hot food, and intake of carcinogens like N-nitroso compounds present in poorly preserved foods such as pickled vegetables (Dayne and Munoz 1982; Li 1982; Day and Varghese 1994).

In the last two decades, increasing numbers of studies have shown that genetic susceptibility may influence the risk of developing EC. Familial aggregation of EC was found in high-incidence areas of northern China (Hu et al. 1992; Chang-Claude et al. 1997). Genetic susceptibility to EC was considered one of the important causes for the high prevalence and familial aggregation of this cancer in some areas of northern China (Wu et

Received March 13, 2000; accepted for publication May 11, 2000; electronically published June 6, 2000.

Address for correspondence and reprints: Dr. Min Wu, National Laboratory of Molecular Oncology, Cancer Institute, Chinese Academy of Medical Sciences, Panjiayuan, Chaoyang Qu, P.O. Box 2258, Beijing 100021, People's Republic of China. E-mail: wum@pubem.cicams.ac.cn

© 2000 by The American Society of Human Genetics. All rights reserved. 0002-9297/2000/6701-0015\$02.00

al. 1989). Thus, it is of great importance to understand the mechanisms of the genetic susceptibility and to establish effective ways of screening individuals who are highly susceptible to EC in high-incidence areas. The only published segregation analysis of EC (Carter et al. 1992) suggested an autosomal recessive Mendelian inheritance of EC in Linxian, the county with the highest mortality rate in China. However, in that study, 221 high-risk nuclear families were analyzed. Environmental factors and age at onset could not be analyzed, because of missing data. As the authors mentioned, the inference of Mendelian inheritance applies only to the subpopulation studied—that is, to nuclear families in Linxian in which at least one person is affected with EC and all offspring are aged ≥ 40 years. The patients included those with carcinoma of the gastric cardia.

In order to explore whether genetic factors play the same role in the etiology of EC in a moderate high-incidence area of northern China, we performed a genetic epidemiological survey of six towns in the suburbs of Yangquan City, Shanxi Province. The crude mortality from EC in this city is 40.12/100,000, which is in a middle level among the high-risk areas in northern China. EC accounts for $>40\%$ of cancer deaths occurring in this city. Previous analysis has shown positive familial aggregation of EC in this city and the results favor genetic etiology, at least as important as known environmental factors (Li et al. 1998). To further explore the mode of inheritance of EC in this area and to compare the result with that of the previous study in Linxian, the family data were subjected to complex segregation analysis in this study.

Material and Methods

Study Population

A total of 228 unrelated EC patients (probands) who were newly diagnosed with esophageal cancer between July 1, 1989–July 1, 1994 were identified in the survey of all 132,039 people in six towns of the suburbs of Yangquan City (about one-tenth of the population of the city in 1994), Shanxi Province. Families were ascertained through single probands. Where there was more than one patient who was newly affected with EC within that period in a family, the latest-affected patient was selected as the proband and one questionnaire was used for the family to eliminate duplication. Six such families were observed, containing 13 newly affected patients. All the patients/probands were diagnosed by X-ray and/or by cytology.

The study was approved by the Institutional Review Board of the Cancer Institute at the Chinese Academy of Medical Sciences, and informed consent was obtained from each participating individual.

Data Collection

Interviewers, who were doctors from each village, were trained before the formal investigation. Face-to-face interviews were conducted in the houses of each household, and a structured questionnaire was administered for each family. Since females often were married traditionally outside of the village and were not available for interviews, familial information was collected through the male head of household, aged 55–65 years, in each family, so that as much information could be obtained on four generations as possible. The head of household was usually the proband, if alive, or his or her father or son if he or she was aged <55 or >65 years, or her husband if the proband was female. When the proper male head of household could not be defined, a female head of household was interviewed (6.2% of families). Four consecutive generations were investigated: the head of household and spouse, parents, sibs, offspring, grandparents, aunts, uncles, and their spouses and children. Information about disease status, age of examination or death, cause of death, age of onset, clinical and pathological diagnosis, and lifestyle (smoking, alcohol consumption, and eating habits, etc.) was recorded.

To reduce false negatives, those older people (mostly grandparents) who were diagnosed by village doctors or were suspected of having died of EC because they suffered swallowing difficulty were considered to be EC patients. This is because the clinical symptoms of EC are very distinctive, and swallowing difficulty is the simplest sign to use in the diagnosis of the disease. Because of poor medical conditions in the past, it is difficult to ensure that all the patients were confirmed by laboratory examination. Of 380 patients (see Results), 51 (13.4%) were diagnosed in this way without X-ray or cytological diagnosis, 34 confirmed by village doctors and 17 simply by the recollection that they died of “swallowing difficulty.” Since the symptom of swallowing difficulty is also typical of the late-stage carcinoma of the gastric cardia in this area, some of these patients might be considered to have been affected with EC, whereas they were actually affected with carcinoma of the gastric cardia. However, the first symptom of carcinoma of the gastric cardia is anemia, as is not the case with EC, and the incidence of EC was over threefold higher than that of carcinoma of the gastric cardia (Guanrei and Sunliang 1987). Thus, the false-positive rate would not be high enough to cause a serious bias.

After the interviews, the finished questionnaires were checked by professionals. Those not completed clearly were picked out, and the heads of the households were reinterviewed. In the end, one-tenth of the completed questionnaires were randomly selected and the heads of households were reinterviewed by professionals for a

final check. The result was satisfactory to our strict criteria, with answers to >90% of all questions being consistent, and with much higher consistency in the important variables for this study (EC status, sex, and age at onset).

Three families were excluded from this study in the end, since the probands who died of EC had previously been diagnosed with primary carcinoma of the gastric cardia and the EC may have been metastatic. Thus, 225 families, with a total of 7,701 individuals, were studied. Among the probands, 125 were males and 100 were females, with a ratio of 1.25, which is lower than that of the city in 1990 (crude ratio 1.90, adjusted ratio 1.48) (Yangquan Cancer Control Office 1990). The reasons for the lower ratio in this study might be as follows: (1) The probands in the present study came from a considerably smaller population, and there may be variations in different parts of the area. (2) Carcinoma of the gastric cardia was combined with EC in the rate calculations for the city. This disease has a higher male-to-female ratio, which may partly conceal the real ratio of EC. (3) The ratio varies considerably in different areas. Generally, the ratio is lower in high-incidence areas—for example, 1.58 in Linxian, Henan Province, and 1.24 in Yangzhou County, Jiangsu Province.

There were 980 and 1,510 individuals who had the habit of eating pickled vegetables and eating hot food, respectively, but who were missing information on the number of years of consumption. We analyzed the distribution of age at interview, sex, EC status, proband status, age at onset, occupation, etc., in the individuals who were missing these data and in those who were not missing these data. All factors showed no difference between these two groups.

Statistical Methods

The REGTL program of S.A.G.E. (1997) (version 3.0) under a Linux 2.0 operating system was used to perform complex segregation analysis. This program uses maximum-likelihood methods to estimate parameters of mathematical models of disease occurrence in families. It assumes that, under a class D regressive model, a censored trait—such as age at onset or susceptibility to the disease—follows a logistic distribution (Bonney 1986; Elston and George 1989). If Mendelian transmission exists, it is assumed to be through a single autosomal locus with two alleles, A and B, A being associated with the affected state. Go et al. (1978) used “type” to describe the discrete factors that affect a person’s phenotype. The same concept was denoted “ousiotype” by Cannings et al. (1978), and genotype is the special case of type or “ousiotype” that is transmitted according to Mendelian mode. So we can use “type” to represent any kind of transmission. Two general models could be assumed. In

model 1, the genotype or type is assumed to influence age at onset of the affected state—through location and scale parameters (α and β)—but not susceptibility; the susceptibility only depends on sex and randomly distributed environmental factors. In model 2, which applies only to situations with a single affection class, type is presumed to influence susceptibility to the affected state, but not the parameters of the age-at-onset distribution, and groups of individuals of different types have the same mean age at onset. Analyses were performed under both models. EC was represented by a dichotomous variable y , in which $y = 1$ for affected and 0 for unaffected. Using the program, the following parameters were estimated: type frequencies Ψ_u ($u = AA, AB, BB$; if the type frequencies are in Hardy-Weinberg equilibrium proportions, then they are defined in terms of $q_A =$ frequency of allele A); transmission probabilities τ_u (the probability that a parent of type u transmits allele A to an offspring; under Mendelian transmission, $\tau_{AA} = 1$, $\tau_{AB} = 0.5$, $\tau_{BB} = 0$); baseline parameter β , which can be sex-dependent (β_s under both model 1 and model 2; $s = F$ for female and M for male) and/or type-dependent (β_{us} under model 1); covariate coefficients ξ_1, \dots, ξ_n ; age adjustment coefficient α ; and susceptibility parameter γ , which can be sex-dependent (γ_s under both model 1 and model 2) and/or type-dependent (γ_{us} under model 2). Susceptibility is defined as the maximum cumulative probability that an individual will be affected with the disease.

Bonney (1986) introduced logistic regressive models for dichotomous traits. The class A model assumes that sibs are dependent only because of common parentage. The class D model assumes that, given parental outcomes, the outcomes of offspring are equally predictive and depend on the numbers of older sibs who are affected and unaffected. The latter model is of particular interest when there is polygenic inheritance or a common sibling environment.

To compare with the results of the former study under a class D model, we generated the following covariates representing residual familial effects as described by Carter et al. (1992). Briefly, these are F_1 (affected father effect), F_2 (unaffected father effect), M_1 (affected mother effect), M_2 (unaffected mother effect), S_1 (affected spouse effect), S_2 (unaffected spouse effect), OS_1 (number of affected older sibs), and OS_2 (number of unaffected older sibs). F_1 , M_1 , and S_1 were coded 1 if father, mother or spouse were affected, respectively, and 0 if unaffected or missing. F_2 , M_2 , and S_2 were coded 1 if father, mother or spouse were unaffected, respectively, and were coded 0 if affected or missing. Thus, the regression coefficients ξ_{F1} is the increase (or decrease, if negative) in the logit (the risk of EC) if the father is affected, ξ_{F2} is the change in the logit if the father is unaffected, and a person’s logit is unchanged if the EC status of the father is un-

known. The other ξ 's for the familial covariates are similarly defined.

In each model, five hypotheses were tested against the likelihood of a general (unrestricted) model, in which all parameters were unrestricted and allowed to fit the empirical data; thus this general model gives the best fit to the data. The five hypotheses of transmission are as follows: (1) no major type (no major gene), (2) purely environmental major effect, (3) Mendelian dominant, (4) Mendelian recessive, and (5) Mendelian codominant (arbitrary major gene). The last hypothesis is a more general one, including the previous two as special cases in which the major gene is restricted to dominant or recessive inheritance; thus, it must fit the data better than the previous two. Under the purely environmental major effect hypothesis, an individual's phenotype depends on his or her own personal environmental exposures and is independent of the parental phenotype (i.e., there is no parent-offspring transmission of type); thus, the transmission probabilities are either restricted (1) to be equal ($\tau_{AA} = \tau_{AB} = \tau_{BB}$, allowing possible heterogeneity of exposure levels between generations), or (2) to be equal to the frequency of allele A ($\tau_{AA} = \tau_{AB} = \tau_{BB} = q_A$, assuming complete homogeneity of environmental exposures between generations). Under this model, the age at onset distribution or susceptibility may exhibit a mixture of two or three distributions, because of personal exposures to major random unmeasured environmental risk factors. The no-major-type hypothesis has no major gene or major environmental-type effects, but allows for random environmental factors that result in only one type distribution.

Twice the difference in the natural log likelihood (lnL) for the data under the hypothesis of interest and that under the unrestricted model was compared to the χ^2 distribution to assess departure from expectation. The degrees of freedom (df) for the χ^2 statistic are given by the differences in the number of estimated parameters between the hypothesis and the unrestricted model. If one or more parameters are fixed at a bound at the end of the estimation process, a range of the df and *P* values are given, when appropriate. A nonsignificant χ^2 indicates that the hypothesis can not be rejected. When more than one hypothesis cannot be rejected, Akaike's (1974) information criterion (AIC), which is defined as $AIC = -2\ln L + 2(\text{number of parameters estimated})$, was used to compare hypotheses. The hypothesis with the minimum AIC fits the data best.

The cumulative probability (CP) of being affected with EC predicted by the best fitting Mendelian model was calculated for various genotype, age, and sex combinations as follows: $CP = \gamma[e^f / (1 + e^f)]$, where $f = \beta + \alpha a + \xi_1(\chi_1) + \dots + \xi_n(\chi_n)$, where *a* is the age and α, β or γ can be type and/or sex dependent. The population average cumulative probability (*P*) weighted by the es-

timated population genotypic frequencies is calculated as $P = q_{AA}CP_{AA} + q_{AB}CP_{AB} + q_{BB}CP_{BB}$, where *q* is the genotype frequency of AA or AB or BB.

To correct for ascertainment bias, the likelihood of each pedigree was conditioned on the proband's EC status by age at examination or death and his or her age at onset. This assumes single ascertainment, which is a reasonable approximation since only 6 (2.7%) of 225 families had more than one patient eligible to be a proband.

Results

Of 7,701 individuals, 380 (4.9%), including probands, were affected with EC. Of 225 families, 93 (41.3%) had two or more EC patients. The mean age at onset was 60 years (range 34–84). The average age of the 36 probands alive at the time of the interview was 61 years (range 39–82 years). Table 1 gives a general description of the data.

First, simple segregation analyses were performed that did not include regressive familial effects or environmental covariates in the models and in which age at onset and susceptibility were not sex-dependent. As a result, the Mendelian recessive major gene hypothesis was marginally not rejected under model 1 ($.02 < P < .1$) (table 2). Under model 2, the recessive and codominant Mendelian hypotheses were not rejected, and the

Table 1
General Characteristics of Data

Covariate	No. (%) of Individuals
Sex:	
Male	4,118 (53.5)
Female	3,583 (46.5)
Occupation:	
Farmer	5,744 (74.6)
Factory worker	942 (12.2)
Office worker	162 (2.1)
Student ^a	853 (11.1)
Cigarette smoking:	
Yes	1,827 (23.7)
No	5,870 (76.2)
Unknown	4 (0)
Pipe smoking:	
Yes	1,101 (14.3)
No	6,599 (85.7)
Unknown	1 (0)
Alcohol drinking:	
Yes	1,620 (21.0)
No	6,081 (79.0)
Eating pickled vegetables:	
Yes	5,877 (76.3)
No	1,824 (23.7)
Eating hot food:	
Yes	5,740 (74.5)
No	1,961 (25.5)

^a Including children too young to attend school.

Table 2
Segregation Analysis of EC under Model 1 (Type Influences Age At Onset)

PARAMETER ^a	HYPOTHESIS						
	Mendelian			No Major Type	Environmental		General
	Dominant	Recessive	Codominant		τ 's Equal	τ 's = q	
q_A	.1022	.3210	.32107551	.1481	.7150
τ_{AA}	1.0 ^b	1.0 ^b	1.0 ^b	...	0 ^c	.1481 ^d	.5403
τ_{AB}	.5 ^b	.5 ^b	.5 ^b	...	0 ^d	.1481 ^d	.9482
τ_{BB}	0 ^b	0 ^b	0 ^b	...	0 ^d	.1481 ^d	0 ^c
β_{AA}	-10.4260	-10.3268	-10.3068	-10.4257	-68.6000	-68.60	-57.0900
β_{AB}	-10.4260 ^d	-54.0750	-53.8200	-10.4257 ^d	-10.0804	-11.85	-60.8300
β_{BB}	-53.0640	-54.0750 ^d	-55.8000	-10.4257 ^d	-10.6945	-10.55	-10.4350
α	.1689	.1665	.1665	.1684	.1683	.1753	.1681
γ	.2852	.4981	.4981	.0945	.1457	.0975	.6294
χ^2	24.62	7.95	7.95	83.53	63.17	83.46	...
df	3-4	3-4	2-3	5-6	1-2	2-3	...
P value	<.001	.02< P <.1	.01< P <.05	<.001	<.001	<.001	...
$-2\ln L$	1841.19	1824.52	1824.52	1900.10	1879.74	1900.03	1816.57
AIC	1851.19	1834.52	1836.52	1906.10	1893.74	1912.03	1834.57

^a See Materials and Methods for definitions of the parameters.

^b Parameter is fixed at this value and is not estimated.

^c Parameter estimate went to bound.

^d Parameter is constrained to equal the preceding one and is not estimated.

dominant hypothesis was marginally not rejected ($P > .5$, $P > .3$, and $.01 < P < .3$, respectively). All other models were rejected at the .001 significance level. The recessive model had the minimum AIC (table 3). When tables 2 and 3 were compared, the Mendelian recessive major gene hypothesis under model 2 fitted the data best.

To explore the heterogeneity of age at onset, we separated the families into two groups: one with probands whose age at onset was <60 years (105 families, group 1) and the other with probands aged ≥ 60 years (120 families, group 2). Then the two data sets were analyzed under models 1 and 2, respectively, as above. Under model 1, the recessive and codominant hypotheses were not rejected ($P > .75$) in group 1, and the recessive model was not rejected in group 2. Under model 2, all the Mendelian hypotheses were not rejected in both of the groups. The recessive major gene model under model 2 fitted the data best. A χ^2 test for heterogeneity, $-2[\ln L(\text{group 1}) + \ln L(\text{group 2}) - \ln L(\text{complete data})]$ was performed, and was not significant (e.g., $\chi^2 = .85$, df 5, $P > .95$, for recessive hypothesis, model 2). Thus, there was no evidence of significant heterogeneity between these two subsets of the data, and all subsequent analyses used the complete data set.

Under either model 1 or model 2, the likelihoods showed no significant improvements when the age-at-onset distribution parameters α and/or β were sex dependent but did show significant improvement when the susceptibility was sex dependent ($\chi^2 = 7.1$, df 2, $P < .05$ for the recessive major gene hypothesis under model 2). This suggested that there was not a significant difference between the mean age of onset for males and

females but that males had a higher relative risk than did females. Table 4 presents the results under model 2, which shows that the best estimate of susceptibility (γ) was .6353 for females and 1 for males with AA genotype under the best-fitting recessive model with an average age at onset of 62 years. When sex was incorporated into the models as a covariate (coded 1 for male and 0 for female), there were no significant effects on the likelihoods.

When a class D regressive model was fitted to the data by incorporating the familial effect covariates into the models, there were no significant improvements in the likelihoods of any models except for the no major type and the environmental models, and these two models were still strongly rejected ($P < .001$). The results were similar to those in tables 2, 3, and 4. Model 2 always fit better than model 1. The Mendelian recessive major gene model under model 2 remained the best-fitting model (table 5).

We incorporated binary (yes or no) environmental covariates—such as smoking (cigarette or pipe, separately or combined in the analysis), alcohol consumption, wine consumption, eating pickled vegetables, and eating hot food—into the age at onset distribution of EC under both model 1 and model 2. No significant improvements in the likelihoods were found. We further analyzed quantitative variables: for example, the number of cigarettes smoked per day, years of cigarette smoking, the number of pipes smoked per day, years of pipe smoking, years of alcohol drinking, the frequency of alcohol drinking, the degree of alcohol consumption, years of eating pickled vegetables, and years of eating hot food. None of these factors significantly improved

Table 3
Segregation Analysis of EC under Model 2 (Type Influences Susceptibility)

PARAMETER ^a	HYPOTHESIS						
	Mendelian			No Major Type	Environmental		General
	Dominant	Recessive	Codominant		τ 's Equal	τ 's = q	
q_A	.0121	.1767	.17556476	.5560	.1913
τ_{AA}	1.0 ^b	1.0 ^b	1.0 ^b5067	.5560 ^c	1 ^d
τ_{AB}	.5 ^b	.5 ^b	.5 ^b5067 ^c	.5560 ^c	.4830
τ_{BB}	0 ^b	0 ^b	0 ^b5067 ^c	.5560 ^c	.0383
β	-10.3680	-10.2019	-10.2100	-10.4262	-10.5727	-10.4256	-10.3353
α	.1676	.1640	.1642	.1684	.1697	.1684	.1662
γ_{AA}	.7391	1 ^d	1 ^d	.0945	0 ^d	.3056	1 ^d
γ_{AB}	.7391 ^c	.0265	.0289	.0945 ^c	0 ^d	0 ^d	0 ^d
γ_{BB}	.0387	.0265 ^c	.0258	.0945 ^c	.5527	0 ^d	.0259
χ^2	5.76	1	1	83.13	67.83	86.13	...
df	1-4	2-4	1-3	3-6	1-4	2-3	...
P value	.01<P<.3	.5<P<.95	.3<P<.9	<.001	<.001	<.001	...
-2lnL	1819.73	1814.97	1814.97	1900.10	1881.80	1900.10	1813.97
AIC	1829.73	1824.97	1826.97	1906.10	1895.80	1912.10	1831.97

^a See Materials and Methods for definitions of the parameters.
^b Parameter is fixed at this value and is not estimated.
^c Parameter is constrained to equal the preceding one and is not estimated.
^d Parameter estimate went to bound.

the likelihoods of any of the models compared with the corresponding models when the coefficients of these factors were fixed at zero. These hypotheses about the effect of environmental covariates were tested by comparing more restricted models that incorporated the environmental covariates into the model but fixed their coefficients to zero against models where the coefficients of the covariates were estimated (results not shown). Finally, table 6 shows the age-, sex-, and genotype-specific penetrances of EC predicted by the best-fitting recessive model under model 2 (in table 4) in this population.

Discussion

In the 1970s, studies showed that more than half of EC patients had a family history of EC in some areas of northern China. A study in Yangcheng, Shanxi Province, showed that EC cases aggregated in <10% of the families (Li and He 1986). A 10-year followup study showed that more families with a prior EC history reported new EC deaths than those without a prior history (19% vs. 5%) (Hu et al. 1992). In this study, >40% of the families had multiple patients. Wu and colleagues (1989) conducted a series of population surveys and laboratory studies of EC beginning in the early 1980s. All the results show that genetic risk factors may be very important in the high-incidence areas of China (Wu et al. 1989).

In the world, time trends for EC rates range widely. The association between the trends in the rates of EC and those of the main, identified etiologic factors was weak. This is in contrast to cancers of the lung and larynx, for which trends are closely related to levels of

tobacco use and, for the larynx, alcohol consumption (Day and Varghese 1994). In China, especially in high-incidence areas, EC mortality rates have not significantly decreased. Genetic susceptibility may be a major etiologic factor in high-incidence areas in northern China, which could not be easily modified. This could partly explain why there has not been a significant decrease in the mortality of EC in these areas. A recent study (Hu et al. 1999) found that loss of heterozygosity (LOH) of a locus on 13q was more common in EC patients with a family history of upper gastrointestinal cancer than in those without such a history, suggesting that a gene in this area may be involved in genetic susceptibility to EC. A large number of studies have shown various molecular genetic changes related to EC, for example, allelic loss, genetic polymorphisms, and gene alterations, etc. (Montesano et al. 1996).

In the previous study (Carter et al. 1992), an autosomal recessive major gene was suggested, along with significant parental, spousal, and sibling correlation indicating unexplained environmental factors in the etiology of esophageal cancer in the subpopulation studied. The present study is different in several ways, as follows: (1) A moderate high-incidence area in northern China (Yangquan City, Shanxi Province) was studied, where the EC mortality rate (40.12/100,000) is lower than that in Linxian and Yangcheng (>100/100,000). (2) Environmental factors were incorporated into the analyses. (3) Four generations and all age ranges were studied. (4) Two different basic models (models 1 and 2) were tested, and age-at-onset information was completed that enabled us to conduct more complex analysis. (5) We incorporated covariates reflecting familial

Table 4
Segregation Analysis of EC under Model 2 (Type Influences Susceptibility) When Susceptibility Is Sex-Dependent

PARAMETER ^a	HYPOTHESIS						
	Mendelian			No Major Type	Environmental		General
	Dominant	Recessive	Codominant		τ 's Equal	τ 's = q	
q_A	.0129	.2039	.18826332	.5881	.1870
τ_{AA}	1.0 ^b	1.0 ^b	1.0 ^b4723	.5881 ^c	1 ^d
τ_{AB}	.5 ^b	.5 ^b	.5 ^b4723 ^c	.5881 ^c	.4971
τ_{BB}	0 ^b	0 ^b	0 ^b4723 ^c	.5881 ^c	.0215
β	-10.3950	-10.3008	-10.3075	-10.4216	-10.5587	-10.4212	-10.4116
α	.1680	.1660	.1662	.1682	.1695	.1682	.1680
γ_{AAF}	.5558	.6353	.7324	.0718	0 ^d	0 ^d	.7380
γ_{ABF}	.5558 ^c	.0203	0 ^d	.0718 ^c	0 ^d	0 ^d	0 ^d
γ_{BBF}	.0310	.0203 ^c	.0305	.0718 ^c	.3944	.4235	.0288
γ_{AAM}	.8577	1 ^d	1 ^d	.1155	.0465	.0848	1 ^d
γ_{ABM}	.8577 ^c	.0250	.0707	.1155 ^c	0 ^d	0 ^d	.0804
γ_{BBM}	.0456	.0250 ^c	.0102	.1155 ^c	.5049	.5083	0 ^d
χ^2	6.27	1.63	.04	84.92	69.34	84.92	...
df	1-5	2-5	1-3	4-8	1-2	2-3	...
P value	.01<P<.3	.3<P<.9	.8<P<1	<.001	<.001	<.001	...
-2lnL	1812.52	1807.88	1806.30	1891.18	1875.59	1891.18	1806.25
AIC	1826.52	1821.88	1824.30	1899.18	1895.59	1909.18	1830.25

^a See Materials and Methods for definitions of the parameters.

^b Parameter is fixed at this value and is not estimated.

^c Parameter is constrained to equal the preceding one and is not estimated.

^d Parameter estimate went to bound.

effects (parent-child, sib-sib, spouse-spouse) into the models, which enabled us to analyze the data under the assumption of a class D regressive model while allowing for censoring of the time to disease onset and susceptibility.

In this study, the recessive major gene model under model 2 fitted the data best, even when the familial regressive effects (spouse, parent, and sibling) were incorporated into the models. The purely environmental model and the no-major-type model were rejected under both model 1 and model 2. This suggests that an autosomal recessive major gene that influences the susceptibility to EC may play a role in the inheritance of susceptibility to EC. As was not the case in the former study, all class D familial effect covariates did not significantly improve the fit of the models in this study. The familial effects found in the previous study may be partly due to the unusually high percentage of affected persons in the sample (40%). In this study, the percentage of affected persons was only 4.9%. We believe that the recessive major gene model suggested in this and the previous study indicate the presence of a gene that influences the susceptibility to EC. Although we did not find significant familial regressive effects in this study, we can not exclude the possibility that polygenic/multifactorial factors may play an important role in the etiology of EC. Development of advanced analytical methods and study of additional samples from more high-risk areas may be of help in better understanding EC etiology.

The susceptibility-allele frequency detected in this study (.2039) is very similar to that in Linxian (.1982) (Carter et al. 1992). This might suggest similar major genetic background in these two areas, whereas there might be some different environmental or other minor genetic factors modifying the difference of the mortality of EC in these areas. However, it is important to realize that parameter estimates in segregation analyses often have large standard errors, and, thus, should not be overinterpreted.

Many studies have shown that most of the common cancers—such as colon, breast, prostate, lung, and ovarian cancer—have one or more rare autosomal dominant gene(s) that increase the risk of these cancers, with susceptibility-allele frequencies in the range .002–.006, penetrance >60%, and the proportion of patients affected because of these genes decreasing with increasing age at onset (Claus 1995). In our study, no evidence was found for a major gene effect on the age at onset of EC nor for heterogeneity in age at onset of EC.

This study suggested a significant sex effect on susceptibility to EC. That the susceptibility for males with the AA genotype was estimated to be 100% and that for females to be 63.5% suggested that there should be different genetic and/or internal or external environmental risk factor exposures in the two sexes. This is consistent with previous analyses that showed that the increased risk of EC for males was greater than that for females in the first degree relatives of patients (10.49-fold vs. 7.69-fold) (Li et al. 1998). The population av-

Table 5
Segregation Analysis of EC, Incorporating Familial-Effect Covariates

PARAMETER ^a	HYPOTHESIS							
	Model 1				Model 2			
	Recessive	No Major Type	τ 's Equal	General	Recessive	No Major Type	τ 's Equal	General
q_A	.33250311	.2677	.17366043	.8580
τ_{AA}	1.0 ^b3326	1 ^c	1.0 ^b3207	1 ^c
τ_{AB}	.5 ^b3326 ^d	.0645	.5 ^b3207 ^d	.5136
τ_{BB}	0 ^b3326 ^d	.4786	0 ^b3207 ^d	.0745
β_{AA}	-10.8163	-11.2618	-6.3085	-10.5457	-10.5859	-11.2618	-10.3926	-10.6515
β_{AB}	-56.6500	-11.2618 ^d	-11.7224	-61.5000	-10.5859 ^d	-11.2618 ^d	-10.3926 ^d	-10.6515 ^d
β_{BB}	-56.6500 ^d	-11.2618 ^d	-71.7000	-54.7800	-10.5859 ^d	-11.2618 ^d	-10.3926 ^d	-10.6515 ^d
ξ_{S1}	-.0213	.1931	.1547	-.1776	-.2023	.1931	.1131	-.1270
ξ_{S2}	.4576	.7137	.8179	.4285	.4287	.7137	.7073	.5070
ξ_{F1}	.0492	.3844	-1.0750	-.6005	-.3428	.6704	.2905	-.4834
ξ_{F2}	-.1623	-.2955	-2.6694	-.6363	-.2983	-.0095	-.6864	-.3947
ξ_{M1}	.1128	1.3642	-2.4742	-.0452	.0535	1.0781	.3746	-.4034
ξ_{M2}	-.0499	.5229	-3.4406	.0757	.0808	.2369	-.4958	-1.020
ξ_{OS1}	.2266	.4925	2.1263	.3205	.1096	.4925	.6885	.2170
ξ_{OS2}	.0640	.0193	-.1268	.0606	.0903	.0193	-.0117	.0859
α	.1691	.1629	.1818	.1687	.1668	.1629	.1609	.1692
γ_{AA}	.4706	.1070	1 ^c	.6639	1 ^c	.1070	0 ^c	1 ^c
γ_{AB}	.4706 ^d	.1070 ^d	1 ^d	.6639 ^d	.0268	.1070 ^d	0 ^c	0 ^c
γ_{BB}	.4706 ^d	.1070 ^d	1 ^d	.6639 ^d	.0268 ^d	.1070 ^d	.4003	.0195
χ^2	10.39	73.39	41.30	...	2.85	75.18	49.41	...
df	3-4	5-6	2	...	2-4	3-6	1-2	...
P value	.01<P<.05	<.001	<.0012<P<.7	<.001	<.001	...
-2lnL	1820.77	1883.77	1851.68	1810.38	1811.44	1883.77	1858.00	1808.59
AIC	1846.77	1905.77	1881.68	1844.38	1837.44	1905.77	1888.00	1842.59

^a See Material and Methods for definitions of the parameters.
^b Parameter is fixed at this value and is not estimated.
^c Parameter estimate went to bound.
^d Parameter is constrained to equal the preceding one and is not estimated.

erage cumulative probabilities of EC predicted by the best-fitting model in this study (6.24% for males and 4.36% for females at age 80 years) are similar to or a little lower than the cumulative mortality rates of EC in the 1970s in the same city (15.26% for males and 6.44% for females at age 74 years) (National Cancer Control Office 1980). We should note that, in that time period, EC and carcinoma of the gastric cardia were combined in computing the rate, so the reported rate was higher than the true rate. In addition, other polygenic/multifactorial effects that were not significant in this study may explain part of the increased risk of developing the disease.

Environmental factors have been implicated by previous epidemiologic studies as being very important in EC development. However, in this study, all of the environmental factors—such as smoking, alcohol drinking, eating pickled vegetables, and eating hot food—did not significantly improve the likelihoods of the models, as either binary or quantitative variables. This is not unreasonable. The purpose of segregation analysis is to find a major gene or oligogene/polygene effect but not environmental effects. It is well known that case-control studies are more powerful for detection of environmen-

tal effects, whereas segregation analysis is more powerful for detection of genetic effects. Thus, when the ORs or relative risks (RRs) of environmental factors estimated from case-control studies are not high, they may not have a strong enough effect to influence the likelihoods in segregation analysis. According to the literature, the effects on EC risk of related environmental risk factors such as smoking, alcohol drinking, eating pickled vegetables and eating hot food seem to be moderate and vary from no or weak association with EC to OR or RR <5 (Gao et al. 1994; Kinjo et al. 1998), even in high-risk areas of northern China (Li et al. 1989; Wang et al. 1992; Guo et al. 1994). On the other hand, since many persons were missing information on some of the environmental risk factors, this data set may not have had adequate power to detect effects of these risk factors. In addition, this is a retrospective study, and all the information was obtained by recall from the head of household. It is very difficult to guarantee perfect accuracy of the information, especially the quantitative measurements for older people. Also, there may be other important environmental factors that have not been included in this study, such as nutritional factors. The results suggest that these environmental risk factors

could not significantly affect the risk of occurrence of EC in this sample—or, at least, that their effects were not large enough to be detected by this statistical method. Finally, these environmental risk factors may be of more importance in other, higher-risk, regions of China.

Although the EC mortality rates could be hundreds of times higher or lower in two different regions in China, the cause of the difference is not clear. There are some environmental factors (e.g., local economic status, geographical factors, nitrosamines, micronutrients, fungi, and even viruses) that might differ in different areas. The environmental factors analyzed in this study were considered as common risk factors for EC in high-incidence areas of China. However, there is a great deal of evidence that has shown that environmental factors are not the only important causes of EC: (1) EC mortality rates in high-incidence areas are very stable, even though some environmental risk factors for EC were identified many years ago. For example, the nitrosamine concentration was found higher in high-incidence areas such as Linxian. During the past 20–30 years, the change of water sources, and the reduction of pickled-vegetable eating have reduced nitrosamine exposure but have not changed the EC mortality in those areas significantly. From 1985 to 1991, two large nutrition intervention trials were conducted in Linxian, China. By the end of the intervention, no statistically significant reductions in the prevalence of esophageal dysplasia or EC were found in the group who supplemented their diets daily with multiple vitamins and minerals (Wang et al. 1994). (2) Incidences of EC among immigrants from high-incidence areas to lower incidence areas remained much higher than that of local residents (Wu et al. 1989). The differences of EC incidences between ethnic groups are at least as high as the differences between different areas. (3) With the same environmental exposures, EC occurrence differs a great deal from family to family, typified by the existence of highly aggregated “cancer families.” Familial aggregation may be more significant in high-incidence areas than in low-incidence areas (Wang et al. 1992). It is hard to say which factor (environmental or genetic) is more important. Both may play important roles in the etiology of EC. The mutation of EC gene(s) may act as a trigger for EC onset, or people with the gene(s) may tend to be more susceptible to EC under the effects of certain environmental risk factors.

In conclusion, our study showed that a Mendelian autosomal recessive major gene underlying susceptibility to EC may play an important role in the etiology of EC in a moderate high-incidence area of Northern China. This study provided some modeling parameters of EC (in this area) for further genetic linkage studies.

Table 6

Cumulative Probabilities of EC by Inferred Genotype, Age, and Sex, on the Basis of Model of Mendelian Autosomal Recessive Inheritance, Sex, and Genotype-Dependent Susceptibilities

SEX AND AGE (YEARS)	GENOTYPE		POPULATION AVERAGE
	AA	AB/BB	
Male:			
50	.1192	.0030	.0078
60	.4158	.0104	.0273
70	.7892	.0197	.0517
80	.9517	.0238	.0624
Female:			
50	.0757	.0024	.0054
60	.2642	.0084	.0190
70	.5014	.0160	.0362
80	.6046	.0193	.0436

The identification of the putative gene detected by the study is warranted.

Acknowledgments

The results reported in this paper were obtained by using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource Grant (1 P41 RR03655) from the National Center for Research Resources. We thank Dr. Adam W. Yao for his help on installing the Linux operating system, the software, and on tackling some technical problems. This work was supported by National Climbing Project 18 and 863 HighTech Project, and cosupported by China Key Program on Basic Research (G1998051021, G1998051205) and National Natural Science Foundation of China (39993420, 39990570).

Electronic-Database Information

The accession number and the URL for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim> (for EC [MIM 133239])

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Control* 19:716–723
- Bonney GE (1986) Regressive logistic models for familial disease and other binary traits. *Biometrics* 42:611–625
- Cannings C, Thompson EA, Skolnick MH (1978) Probability functions on complex pedigrees. *Adv Appl Prob* 10:26–61
- Carter CL, Hu N, Wu M, Lin P-Z, Murigande C, Bonney GE (1992) Segregation analysis of esophageal cancer in 221 high-risk Chinese families. *J Natl Cancer Inst* 84:771–776
- Chang-Claude J, Becher H, Blettner M, Qiu S, Yang G, Wahrendorf J (1997) Familial aggregation of oesophageal cancer

- in a high incidence area in China. *Int J Epidemiol* 26: 1159–1165
- Claus EB (1995) The genetic epidemiology of cancer. *Cancer Surv* 25:13–26
- Day NE, Varghese C (1994) Oesophageal cancer. *Cancer Surv* 19–20:43–54
- Dayne NE, Munoz N (1982) Esophagus. In: Schottenfeld D, Fraumeni J (eds) *Cancer epidemiology and prevention*. Saunders, Philadelphia, pp 596–623
- Elston RC, George VT (1989) Age of onset, age at examination, and other covariates in the analysis of family data. *Genet Epidemiol* 6:217–220
- Gao YT, McLaughlin JK, Gridley G, Blot WJ, Ji BT, Dai Q, Fraumeni JF Jr (1994) Risk factors for esophageal cancer in Shanghai, China. II. Role of diet and nutrients. *Int J Cancer* 58:197–202
- Go RCP, Elston RC, Kaplan EB (1978) Efficiency and robustness of pedigree segregation analysis. *Am J Hum Genet* 30:28–37
- Guanrei Y, Sunlian Q (1987) Incidence rate of adenocarcinoma of the gastric cardia, and endoscopic classification of early cardiac carcinoma in Henan Province, the People's Republic of China. *Endoscopy* 19:7–10
- Guo W, Blot WJ, Li JY, Taylor PR, Liu BQ, Wang W, Wu YP, et al (1994) A nested case-control study of oesophageal and stomach cancers in the Linxian nutrition intervention trial. *Int J Epidemiol* 23:444–450
- Hu N, Dawsey SM, Wu M, Bonney GE, He LJ, Han XY, Fu M, et al (1992) Familial aggregation of oesophageal cancer in Yangcheng County, Shanxi Province, China. *Int J Epidemiol* 21:877–882
- Hu N, Roth MJ, Emmert-Buck MR, Tang ZZ, Polymeropoulos M, Wang QH, Goldstein AM, et al (1999) Allelic loss in esophageal squamous cell carcinoma patients with and without family history of upper gastrointestinal tract cancer. *Clin Cancer Res* 5:3476–3482
- Kinjo Y, Cui Y, Akiba S, Watanabe S, Yamaguchi N, Sobue T, Mizuno S, et al (1998) Mortality risks of oesophageal cancer associated with hot tea, alcohol, tobacco and diet in Japan. *J Epidemiol* 8:235–243
- Li GH, He LJ (1986) A survey on the familial aggregation of esophageal cancer in Yangcheng county, Shanxi Province. In: Wu M, Nebert DW (eds) *Genes and diseases: proceedings of the first Sino-American human genetics workshop*. Science Press, Beijing, pp 43–47
- Li JY (1982) Epidemiology of esophageal cancer in China. *Natl Cancer Inst Monogr* 62:113–120
- Li JY, Ershow AG, Chen ZJ, Wacholder S, Li GY, Guo W, Li B, et al (1989) A case-control study of cancer of the esophagus and gastric cardia in Linxian. *Int J Cancer* 43:755–761
- Li W, Wang X, Zhang C, Han X, Chen D, Zhang T, Pan X, et al (1998) Genetic epidemiological survey on esophageal carcinoma in part of population of Yangquan City. *Natl Med J China* 78:203–206
- Liu BQ, Li B (1984) Epidemiology of carcinoma of the esophagus in China. In: Huang GJ, Wu YK (eds) *Carcinoma of the esophagus and gastric cardia*. Springer Verlag, Berlin, pp 1–24
- Montesano R, Hollstein M, Hainaut P (1996) Genetic alterations in esophageal cancer and their relevance to etiology and pathogenesis: a review. *Int J Cancer* 69:225–235
- National Cancer Control Office (1980) *Investigation of cancer mortality in China*. People's Health Publishing House, Beijing, pp 65–96
- Parkin DM, Pisani P, Ferley J (1993) Estimation of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 54:594–606
- S.A.G.E. (1997) *Statistical Analysis for Genetic Epidemiology, Release 3.0*. Computer program package available from the Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland
- Wang GQ, Dawsey SM, Li JY, Taylor PR, Li B, Blot WJ, Weinstein WM, et al (1994) Effects of vitamin/mineral supplementation on the prevalence of histological dysplasia and early cancer of the esophagus and stomach: results from the General Population Trial in Linxian, China. *Cancer Epidemiol Biomarkers Prev* 3:161–166
- Wang YP, Han XY, Su W, Wang YL, Zhu YW, Sasaba T, Nakachi K, et al (1992) Esophageal cancer in Shangxi Province, People's Republic of China: a case-control study in high and moderate risk areas. *Cancer Causes Control* 3: 107–113
- Wu M, Hu N, Wang X (1989) Genetic factors in the etiology of esophageal cancer and the strategy for its prevention in high-incidence areas in northern China. In: Lynch HT, Hirayama T (eds) *Genetic Epidemiology of Cancer*. Boca Raton, Florida, CRC Press, pp 187–202
- Yangquan Cancer Control Office (1990) Report of cancer mortality investigation of Yangquan City. *Cancer Prev* 1:1–22