Tobacco Addiction: Effect on Human Health

Chapter 1

Smoking and Primary Cancer among Chinese Urban Men: A Novel Study Design in Practice

Jingmei Jiang; Wei Han; Lei Hou; Boqi Liu*

Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences / School of Basic Medicine Peking Union Medical College, Beijing, China

***Correspondence to: Boqi Liu**, The Cancer Institute/Hospital, Chinese Academy of Medical Sciences, 17 Pan Jia Yuan Nan Li, Beijing (100021).*

Tel: 86-10-87788441; Fax: 86-10-85370653; E-mail: China. boqiliu@163.com

Abstract

In the 1980s, the hazard of smoking has been widely recognized by general public in western countries. While, during the same period, China had developed rapidly as the world's leading tobacco production and consumption country. However, due to the delay of smoking consumption peak comparing with developed countries, the smoking related research in China had lagged behind western developed countries for nearly 50 years and basic epidemiological information regarding to the health problems associated with smoking was lacking. Therefore, it was imperative to undertake nationwide research to depict the hazards related with smoking comprehensively and systematically, which was crucial for forming preventive public health policy as well as raising awareness of health risk caused by cigarette smoking among general population in China. However, it is almost impossible to achieve this goal in a short period by directly applying traditional epidemiological research designs, such as cohort design and case-control design, both of which are most frequently used in epidemiological etiological studies. In cohort studies, subjects need to be followed up for a period of time to observe the occurrence of the events of interest, which usually takes a long time for a cohort study to mature. While for case-control studies require control group to be an approximate random sample of base population, which is hardly to meet in large-scale population studies. Therefore, there was a call for methodological innovation, through which the impact of smoking on Chinese could be evaluated within relative short period with less resource consumption.

In 1989-1991, a nationwide retrospective mortality survey was conducted in China, which involved 24 major cities and 79 rural counties that covered 67 million of a population, and about 1.2 million adult (age 30 years or above) death during 1986 ~ 1988 were collected. As the first and largest mortality survey in China, the survey was characterized by simultaneously collecting the smoking information of the deceased individuals as well as their surviving spouses, by interviewing the living spouses of deceased people or other informants. This research had received great attentions and aroused considerable repercussions both home and abroad, and it played an important role in convincing opinion leaders, politicians and the general public of the importance both of the problem and of the urgent necessity to address it. The project's primary report, which was published on BMJ in 1998, implemented proportional mortality ratio analysis in evaluating the risks of smoking on different causes of death systematically and comprehensively and presented a whole picture of smoking hazards in Chinese adults. In proportional mortality ratio analysis, the case group consists of deceased people who died from a disease that assumed be related with smoking, and the control group includes deceased subjects with death causes assumed to be unrelated with smoking. Then, differences between the proportions of smokers in case and control groups can be used to calculate the risk ratios. The application of proportional mortality ratio analysis has the strength that any bias affecting assessment of the habits of those in the case group should similarly affect assessment of the habits in the control group. However, the control group has a poor representativeness for base population because the base population consist of living people, which might lead to underestimating the risk of smoking. In addition, we cannot investigate the association between smoking and those death causes involved by control group in proportional mortality ratio analysis.

To address these problems and make a progress from methodological perspective, sex-matched case-spouse control design had been proposed by Chinese epidemiologist Boqi Liu and his colleagues. By utilizing this novel control selection strategy in the data from 1989 nationwide survey, an approximate random sample can be drawn from the base population and the problems mentioned above could be successfully resolved. Using this design, series of studies had been conducted to investigate the hazards of smoking among Chinese adults from both etiology and public health awareness perspectives. In this chapter, we will mainly introduce the theoretical framework of this design, as well as the applications of this design based on the national mortality survey data by investigating the relationship between smoking and primary cancer death among Chinese urban men. In addition, the methodological evaluation of this novel design will also be discussed.

Key words: smoking; chinese; cancer death; case-control study; control selection

1. Introduction

In the 1980's, owing to findings from a large number of smoking-related researches and the tobacco control measures, the hazard of smoking had been gradually recognized by general population in western developed countries, the perception of cigarette smoking had undergone a complete change and the smoking had been viewed as a lethal addiction [1-7]. Meanwhile, China, as the largest developing country with 20% of the world's population, had experienced a dramatic soar in cigarettes consumption and accounted for 30% of the world's cigarette consumption at that time, owing in large part to the earlier and more intensive consumption of cigarette. It was estimated that nearly 67% of males and 4% of females become smokers before the age of 25 and the average cigarette also increased at the same period [8-10].

The epidemic of cigarette smoking in China in 1980s had posed a tremendous threat on health for Chinese population. Moreover, the dramatic growth of tobacco consumption, unless prevented, will result in not only human health but also an economic burden of medical and health costs. In contrast with this severe situation, China was at a different stage of smoking epidemic comparing with western developed country and the smoking-related research in China had lagged behind for nearly 50 years, with few reliable or country-wide research had been done. So the information on hazard caused by smoking is scanty and the awareness of the health risk of smoking was low among general population, which made it urgent to carry out nationwide study to investigate how large the effect of smoking on health and where in China the hazard is the greatest [11]. While, it is almost impossible to achieve this purpose within a relative short period by employing traditional epidemiological study designs, such as cohort and case-control study design, both of which are frequently used in epidemiological etiological research. Because in cohort studies, a number of subjects should be followed up for a period of time to observe the occurrence of the events of interest, which thereby usually takes a long time for a cohort study to mature [12]. While for case-control studies, despite being quick, inexpensive and easily to be carried out, has a requirement that the control group should be an approximate random sample of the base population, which is hardly to meet in large-scale population studies [13]. Therefore, there was an urgent need to found a time-saving and cost-effective method to assess the hazard associated with smoking in consideration of the severe smoking epidemic in China at that time.

In 1989-1991, a nationwide retrospective mortality survey was conducted in China, with 24 major cities being chosen non-randomly to represent a wide range of area, which included Beijing, Tianjin, Shanghai, Harbin, Hangzhou, Changsha, Xi 'an, Lanzhou, Chengdu, Chongqing, Zigong, Guiyang, Guangzhou, Nanjing, Fuzhou, Kunming, Huangshi, Yangquan, Changchun, Jilin, Shenyang, Dalian, Zibo, Luoyang. A total of 79 rural counties were also selected through stratified random sampling from the 2000 counties whose cancer rate in 1973–1975 were recorded in the Chinese cancer atlas (as shown in Figure 1). As the largest mortality

survey in China, the survey covered 67 million of a population and about 1.2 million adult (age 30 years or above) deaths during 1986~1988 were collected. Besides, by interviewing the surviving spouses or other informants, smoking information of deceased person as well as their surviving spouses (other family members or informants if there was no surviving spouses) were collected during the survey.

Figure 1. Location of study areas: 24 major cities (large circles with names) and 79 rural counties (unnamed circles).

As the first and largest nationwide smoking-related study, the 1989 nationwide mortality successfully combined descriptive and analytic epidemiology studies, which belong to different epidemiologic study types. Therefore, the hazard associated with smoking in China could be assessed within only three years, for which the western countries has spent also 50 years. And this research is considered to be a milestone in terms of describing the mortality pattern and smoking pattern, and estimating the harm caused by smoking in China, which had a profound effect on raising the concerns as well as consciousness of the smoking hazard both for the government and the general population. The project's primary report, which was published on BMJ in 1998, estimated the tobacco attributable mortality in middle or old age from neoplastic, respiratory, or vascular disease by employing proportional mortality ratio analysis (PMR) [14]. In PMR analysis, all deceased people are divided into two parts. Those deaths are assumed to be related with smoking were enrolled as cases, whereas for others whose death causes are assumed not related with smoking were enrolled as controls, including infective or parasitic, diabetes, parkinsonism, other nervous or mental disease, renal disease, hepatic disease (chiefly due to chronic hepatitis B infection), peptic ulcer, other digestive disorders, other medical disorders, road traffic accidents, suicide or homicide, and other non-medical reasons. The strength of proportional mortality analyses is that any bias affecting assessment

of the habits of those in the case group should similarly affect assessment of the habits in the control group. Hence, comparison of proportion of smokers can be made between case group and control group to calculate the risk ratios (smoker versus nonsmoker) for mortality from those causes [14]. Moreover, another prominent advantage for PMR is that the criteria for eligible controls can be established conveniently. However, a primary limitation of proportional mortality analysis is that the controls cannot represent the base population due to the reason that the base population consists of living subjects, Which is an obvious violation of study base principle for control selection in case-control study. In addition, it is inevitable that smoking is associated with some certain kinds of deaths in the reference group (for example, some of those from gastric ulcer), which means that proportional mortality analyses might underestimate the risk of smoking.

In light of the these problems, Chinese epidemiologist Boqi Liu, who led the 1989 mortality survey, proposed sex-matched case-spouse control design along with his colleagues, which made it possible to draw an approximate random sample as a control group in large population. More importantly, the implementation of this design could provide an alternative method to give accurate estimate of early smoking-attributable mortality in a nationwide level, which investigate the relationships between smoking and broader range of disease, including those diseases being involved by control group in PMR methods. Since the development of this novel design, it has gained a wide spread attention and recognition both in home and abroad.

By using this innovative design method, we have carried out series of studies (Appendix 1), through which the tobacco hazards were investigated from the following aspects: (1) smoking and deaths from various cancers. Particularly, we proved the positive association between smoking and glioma, which is a rare cancer and its association with smoking remains controversial; (2) smoking and deaths from respiratory disease, including chronic obstructive pulmonary disease and tuberculosis; (3) smoking attributable deaths and life expectancy reduction, which bear significant implications in raising the awareness of hazards associated with smoking in the public for the sake of being easily comprehended and accepted by general individuals. In addition, the hazard of passive smoking, which has been attracted more and more attention, has also been investigated based on the nationwide population-based study. In the following sections, we will introduce the sex-matched case-spouse control design and illustrate the application of this novel design in practice by investigating the association between cigarette smoking and death from primary cancers among Chinese urban men based on data from the 1989-1991 nationwide survey.

2. Methods

2.1 Ascertainment of death cause

In 1989 mortality survey, causes of death were sought chiefly from official death cer-

tificates and supplemented, if necessary, by reviewing medical records or by discussing (a few years after the death) with local health workers, family members or both. The findings were recorded as part I and part II of a standard death certificate, Which were available from local government offices. The underlying cause of each death was coded according to the World Health Organization's International Classification of Diseases, 9th revision (ICD-9). Underlying causes were coded by 100 clerks in five teams, each under a trained oncologist from the Ministry of Health with experience of coding standard death certificates using ICD-9 (international classification of diseases, ninth revision). Some batches of data sheets were coded by two teams and the differences discussed to develop consistent coding conventions. At ages 35-69 only 0.4% of causes were ill defined (codes 780-99), but at older ages 4% were ill defined (3% urban, 6% rural).

2.2 Ascertainment of smoking exposure

All surviving spouses (28.1% of total interviewees in urban areas; 20.7% of total interviewees in rural areas), other family members (35.6% in urban areas; 27.1% in rural areas), or other informants including family relatives and local informants (36.3% in urban areas; 52.2% in rural areas) were interviewed. Interviews were conducted with either the living spouse or another member of the identified household when the spouse was deceased. A short structured questionnaire that had two sections were employed for asking the same questions, one for the deceased person and another for a living person (deceased person's spouse or the other informants). The information concerning smoking history was provided by living informants who described their own smoking habits as well as those of their deceased partners, including types of tobacco (cigarette, hand-made cigarettes or other forms of tobacco), duration of smoking, starting age and average number of cigarettes consumed per day. These data were used to determine whether people had ever smoked by 1980, a period of time prior to the onset of their disease. This approach was used to help guarantee that there was a reduced chance that the interviewee's smoking habits being changed by the disease eventually caused death during 1986–1988. A nonsmoker was defined as a person who had never smoked during his life or had only smoked infrequently during young age.

2.3 Sex-matched case-spouse control design

The scheme of sex-matched case-spouse control design is illustrated in Figure 2. In this design, people aged 30 years old or above who died from any cause–associated smoking during the years 1986-1988 may be defined as cases. Meanwhile, controls were the surviving spouses of those who died of any condition during the same period, whose age range was the same for the cases when his or her spouse had died. This control selection procedure is based on the assumption that in 1980s individuals in the control group had smoking habits similar to those from the base population, and there is no significant relationship between couples in

terms of tobacco use.

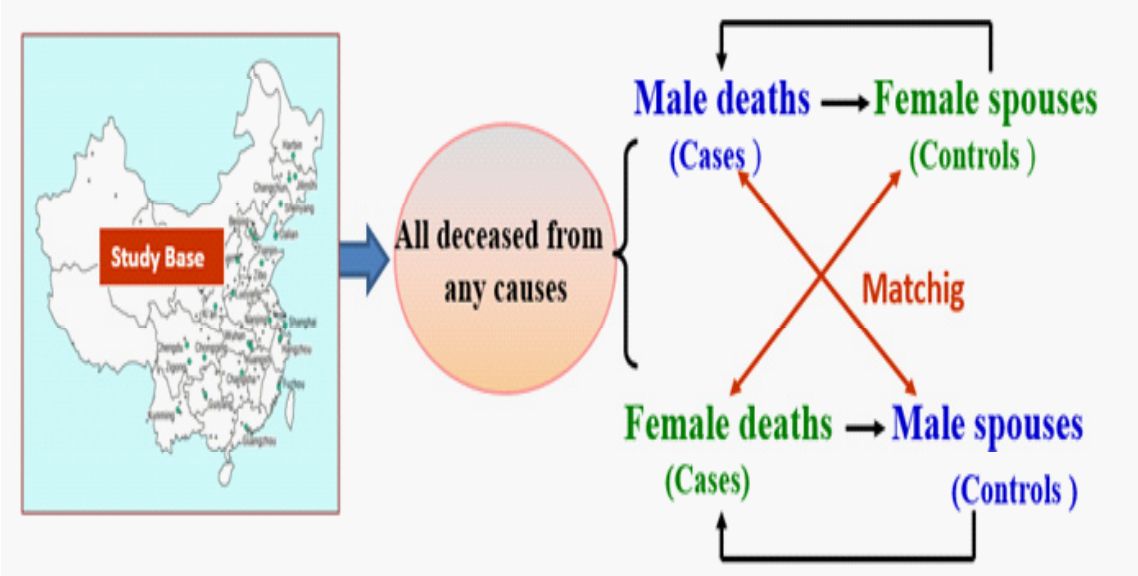

Figure 2: Sex-matched case-spouse control design

In our current study, by implementing living spouse control design, a total of 77,883 urban men aged 35 years or older who died from primary cancers during 1986-1988 were selected as cases, the primary cancer was defined as (n; %) lung cancer (24,794;31.8%), stomach cancer (19,044;24.5%), liver cancer (17,086;21.9%), esophagus cancer (9285;11.9%), pancreaticcancer (2638;3.4%), prostate cancer (1662;0.8%), bladder cancer (584;2.1%), and minor sites cancer which combining lip, oral cavity and laryngeal (2790; 3.6%). The control group was defined as living spouse of those all-caused deceased females during 1986-1988, with the age range was the same for the cases when her spouse had died, and 63,878 subjects were selected as controls.

2.4 Statistical methods

A non-conditional logistic regression model was used to estimate the risk of smoking and conduct a trend test with adjustment for confounders. Odds ratio (OR) with 95% confidence intervals (CIs) were used to estimate the effects of smoking history on the cancer deaths. Attributable fraction (AF), calculated as (OR–1)/OR, was employed to express risk attributing to cancer deaths for the smoking population.

Equivalence evaluation was carried out by testing the equivalence between sex-matched case-spouse control design (denoted by $O\hat{R}_1$) and PMR design (denoted by $O\hat{R}_2$), respectively. Then, the 95% confidence interval (95% CI) for $O\hat{R}_1 / O\hat{R}_2$ were calculated.

We employed the proportion of similar response (PSR) to establish equivalence limit [15], which can be defined as follow

$$PSR = \int \min\{f(x), g(x)\}dx$$

f(x) and *g(x)* are probability density function for cumulative distribution function $f(x)$ and *g(x)*. The more closely the value of PSR approach to 1, the more similar of the two distributions. If the PSR is larger than a predefined value, the two distributions can be regarded as identical and equivalent. In practice, the default value is usually 0.9 or 0.8. In our study, we adopted 0.9 as equivalent standard and thereby the equivalent limit is set to be 0.2513. After conversion, we obtained the equivalent limit of (0.78 1.29), which indicates that equivalence could be concluded if $0.78 < \hat{OR}_1 / OR_2 < 1.29$.

Then, parameter interval method and bootstrap percentile method were employed in estimating the confidence interval. In parameter interval method, we firstly estimate the correlation coefficient *p* between $\hat{OR}_1$ and $\hat{OR}_2$ by using bootstrap technique. Then, confidence interval can be constructed using the following formula:

$$(\ln \hat{OR}_1 - \ln OR_2) \pm u_{0.05/2}\sqrt{se_1^2 + se_2^2 - 2\mathrm{r}\, se_1 se_2}$$

Where se_1 and se_2 denotes the standard error of $\hat{OR}_1$ and $\hat{OR}_2$

In bootstrap method, 100 samples were obtained using re-sampling methods, then 5th percentile and 95th percentile for $\hat{OR}_1 / \hat{OR}_2$ can be obtained and then be compared with the predefined equivalence limit to assess the equivalence.

By re-sampling from the original data set using stratified sampling strategy repeatedly, with the sample size range from 100-25000 (which could satisfy sample size requirement for most case-control study), we evaluate the consistence as well as the stability of the ORs for primary cancers obtained using PMR and the sex-matched living spouse control design respectively under various sample sizes.

All analyses were performed using SAS 9.3 statistical software package (SAS Institute). All *P* values were two-sided except for trend tests, in which one-sided *P* values were used; a *P* value less than 0.05 was considered statistically significant.

3. Results

3.1 Prevalence of smoking among cases and controls

In general, the smoking prevalence for case group and control group were 69.2% and 56.7% respectively. And as shown in Figure 3, the highest smoking prevalence was found in 35-40 age group for control group, while in the case group the peak was found at around 60 years old. In addition, The prevalence of smoking in case group was higher comparing with

control group in almost all age groups, and the gap gradually increased with age at the start and turned stable after 50 years old.

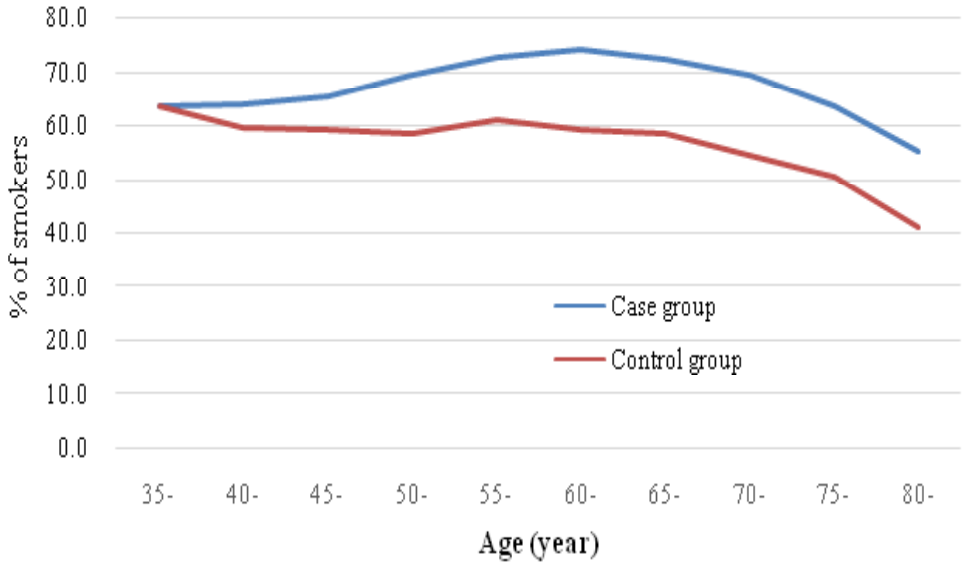

Figure 3: Smoking prevalence by age group in case group and control group, 1990

3.2 Risks for death from primary cancers

The crude and adjusted odds ratios (ORs) of deaths from various primary cancers between smokers and non-smokers are shown in Table 1. There is a relative large variation in risks across the primary cancers, among which lung cancer has the highest association with smoking, and the OR was 2.91 (95% CI: 2.81-3.01). While, the association between smoking and prostate cancer deaths is the lowest, and the OR was 1.04 (95% CI: 0.88, 1.22).

Table 2 to Table 5 present the adjusted ORs for association between smoking and deaths from four major cancers: lung cancer, stomach cancer, liver cancer and esophageal cancer. Although the ORs were influenced by age, the associations were significant for lung cancer across all age groups, and the peak of risk was found in 60-64 age group. However, for other major cancers, significant associations were not observed in lower age groups. For instance, the risk of smoking on stomach cancer death were not observed among individuals under 50 years old.

Table 1. Effect of smoking on various primary cancer deaths

	Crude OR (95% CI)	**Adjusted OR[a] (95% CI)**	**AFs (%)**
Primary cancers combined	1.72 (1.68, 1.76)	1.74 (1.70,1.78)	42.5
Lung cancer	2.91 (2.81,3.01)	2.98 (2.88,3.08)	66.4
Stomach cancer	1.27 (1.23,1.32)	1.32 (1.27,1.36)	24.2
Liver cancer	1.38 (1.33,1.43)	1.33 (1.28,1.38)	24.8
Esophagus cancer	1.72 (1.64,1.80)	1.84 (1.75,1.93)	45.7

Pancreatic cancer	1.33 (1.23,1.45)	1.34 (1.24,1.45)	25.4
Bladder cancer	1.25 (1.13,1.39)	1.45 (1.31,1.60)	31.0
Prostate cancer	1.04 (0.88,1.22)	1.25 (1.06,1.48)	20.0
Minor sitescancer[b]	1.67 (1.54,1.82)	1.66 (1.53,1.81)	39.8

Abbreviations: AFs, attributable fraction; [a]Adjusted for age and area of residence; [b]Combining the lip, oral and throat cancer

Table 2. Age-specific risk of lung cancer deaths associated with smoking

Age group (year)	No of smokers	Total deaths	Adjusted OR[a](95% CI)
35-	189	257	1.59 (1.19,2.13)
40-	278	389	1.70 (1.34,2.15)
45-	643	868	1.98 (1.67,2.34)
50-	1651	2119	2.52 (2.25,2.83)
55-	2934	3598	2.81 (2.55,3.09)
60-	3945	4726	3.47 (3.19,3.79)
65-	3995	4903	3.10 (2.85,3.36)
70-	3319	4172	3.23 (2.97,3.52)
75-	1810	2474	2.70 (2.44,2.99)
80-	864	1288	2.95 (2.59,3.37)
Total	19628	24794	2.98 (2.88,3.08)

[a]Adjusted for age and area of residence.

Table 3. Age-specific risk of stomach cancer deaths associated with smoking

Age group (year)	No of smokers	Total deaths	Adjusted OR[a](95% CI)
35-	165	261	0.99 (0.76,1.29)
40-	211	350	1.03 (0.82,1.30)
45-	382	624	1.09 (0.91,1.30)
50-	933	1464	1.26 (1.12,1.42)
55-	1552	2361	1.22 (1.11,1.34)
60-	2092	3168	1.34 (1.24,1.46)
65-	2437	3730	1.33 (1.23,1.44)
70-	2139	3479	1.33 (1.23,1.44)
75-	1285	2215	1.37 (1.24,1.51)
80-	694	1392	1.43 (1.27,1.62)
Total	11890	19044	1.32 (1.27,1.36)

[a]Adjusted for age and area of residence.

Table 4. Age-specific risk of liver cancer deaths associated with smoking

Age group (year)	No of smokers	Total deaths	Adjusted OR[a](95% CI)
35-	403	667	0.88 (0.73,1.05)
40-	538	866	1.11 (0.95,1.30)
45-	887	1467	1.06 (0.93,1.20)
50-	1604	2468	1.33 (1.21,1.47)
55-	2039	3026	1.32 (1.20,1.44)
60-	1923	2822	1.48 (1.35,1.61)
65-	1582	2383	1.39 (1.27,1.53)
70-	1126	1782	1.43 (1.29,1.59)
75-	604	1041	1.37 (1.20,1.56)
80-	291	564	1.54 (1.29,1.83)
Total	10997	17086	1.33 (1.28,1.38)

[a]Adjusted for age and area of residence.

Table 5. Age-specific risk of esophagus cancer deaths associated with smoking

Age group (year)	No of smokers	Total deaths	Adjusted OR[a](95% CI)
35-	39	57	1.25 (0.71,2.20)
40-	80	108	1.94 (1.25,3.00)
45-	194	253	2.27 (1.68,3.07)
50-	400	537	2.09 (1.71,2.56)
55-	862	1146	1.93 (1.68,2.22)
60-	1131	1534	1.93 (1.72,2.18)
65-	1264	1783	1.72 (1.54,1.92)
70-	1222	1764	1.88 (1.69,2.10)
75-	803	1283	1.66 (1.47,1.87)
80-	428	820	1.58 (1.36,1.84)
Total	6423	9285	1.84 (1.75,1.93)

[a]Adjusted for age and area of residence.

A dose-response relationship between smoking variables, such as years of smoking and number of cigarettes smoked daily with the deaths from four common cancer, were found (Table 6). For lung cancer death, the ORs were 1.21 (95% CI: 1.10–1.32), 2.06 (95% CI: 1.94–2.20), and 3.30 (95% CI: 3.18–3.42) for smokers with <20, 20–29, and ≥30 years of smoking(trend test P<0.001) and 1.70 (95% CI:1.60–1.81), 2.23 (95% CI: 2.13–2.33), and 4.57 (95% CI: 4.39–4.76) for smokers with 10, 10–19, and ≥20 cigarettes smoked daily (trend test P<0.001) after adjustment for age and area of residence compared with non-smokers. This trend was similar for other major cancers.

Table 6. Dose-response relationships between smoking and deaths from four major cancer (Odds ratio (95% CI))

	Lung cancer	Liver cancer	Stomach cancer	Esophagus cancer
Non-smoker				
Years of smoking				
<20	1.21 (1.10,1.32)	0.87 (0.80,0.94)	1.03 (0.94,1.12)	1.12 (0.98,1.29)
20–29	2.06 (1.94,2.20)	1.38 (1.30,1.46)	1.23 (1.15,1.32)	1.65 (1.50,1.81)
≥30	3.30 (3.18,3.42)	1.39 (1.34,1.45)	1.35 (1.31,1.40)	1.91 (1.82,2.00)
P for trend	<.0001	<.0001	<.0001	<.0001
Cigarettes smoked daily				
<10	1.70(1.60,1.81)	1.32(1.24,1.40)	1.23(1.16,1.30)	1.48(1.36,1.61)
10–19	2.23(2.13,2.33)	1.20(1.15,1.26)	1.20(1.15,1.26)	1.53(1.44,1.63)
≥20	4.57(4.39,4.76)	1.52(1.46,1.59)	1.49(1.42,1.55)	2.13(2.01,2.26)
P for trend	<.0001	<.0001	<.0001	<.0001

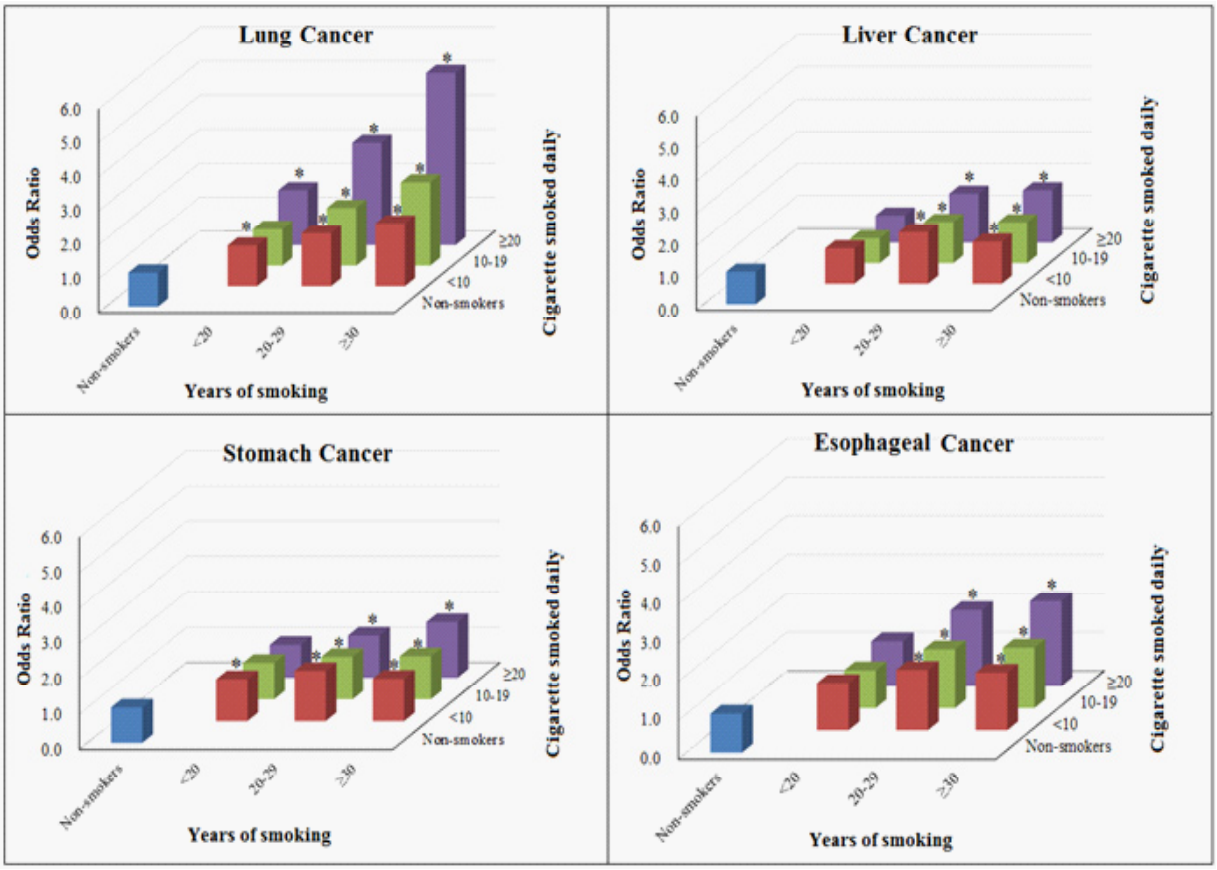

Figure 4. Effects of combining years of smoking with cigarettes smoked per day on death from lung, stomach, liver and esophagus cancer, respectively (adjusted for age and area of residence, **P*<0.05).

To express the synergistic effects of years of smoking and number of cigarettes smoked daily, by using non-smokers as the common reference group,and found that the risk dramatically increased with years of smoking and the number of cigarettes smoked daily (Figure 4).

In particular, in long term heavy smokers (defined as individuals with ≥30 smoking years) and ≥20 cigarettes (equivalent to one pack) smoked daily, the risk increased by 407% (OR=5.07; 95% CI: 4.87 –5.28) for cancer after adjustment for age and area of residence. While the corresponding increased risk were 62% (OR=1.62; 95% CI: 1.55 –1.70), 57% (OR=1.57; 95% CI: 1.51 –1.65) and 120% (OR=2.20; 95% CI: 2.07 –2.34) for liver cancer, stomach cancer and esophagus cancer, respectively.

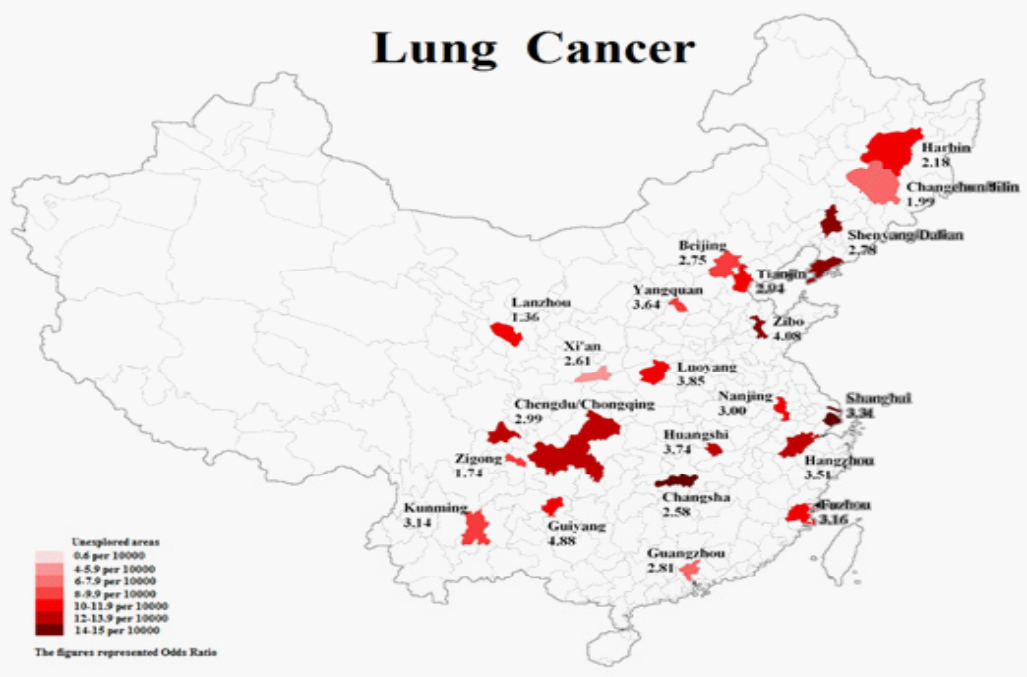

Figure 5. The distribution of mortality rate of lung cancer and its association with smoking among men in 24 cities in China

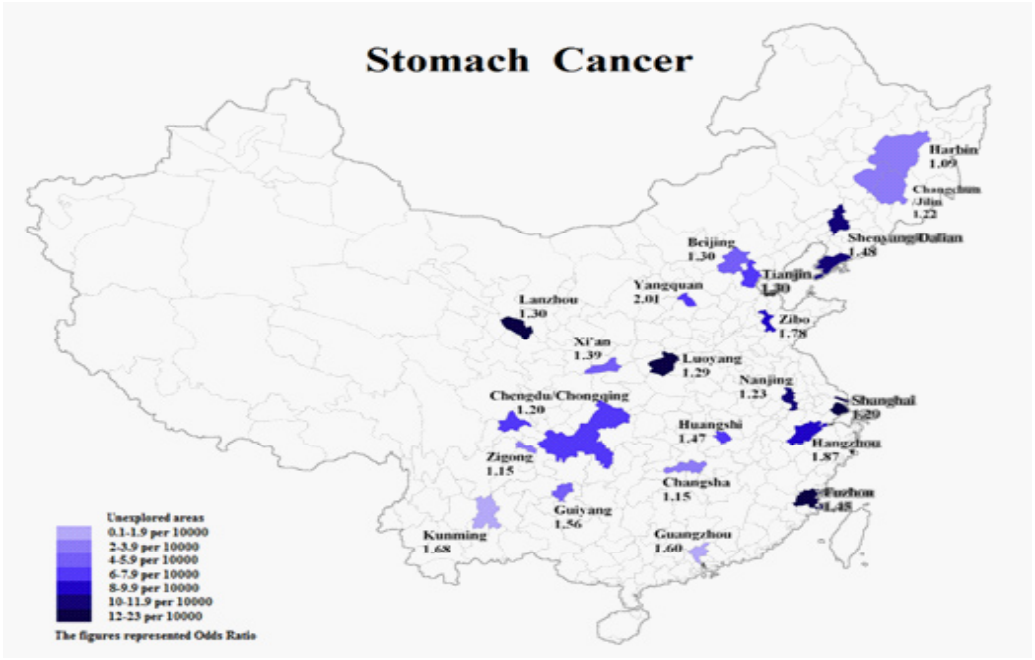

Figure 6. The distribution of mortality rate of stomach cancer and its association with smoking among men in 24 cities in China

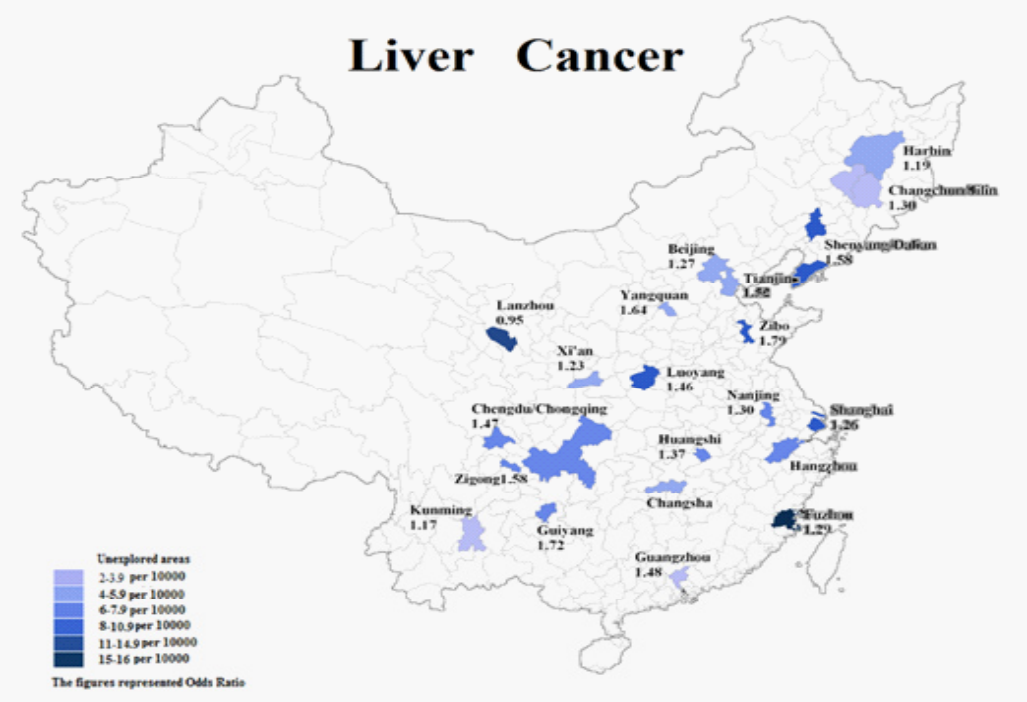

Figure 7. The distribution of mortality rate of liver cancer and its association with smoking among men in 24 cities in China

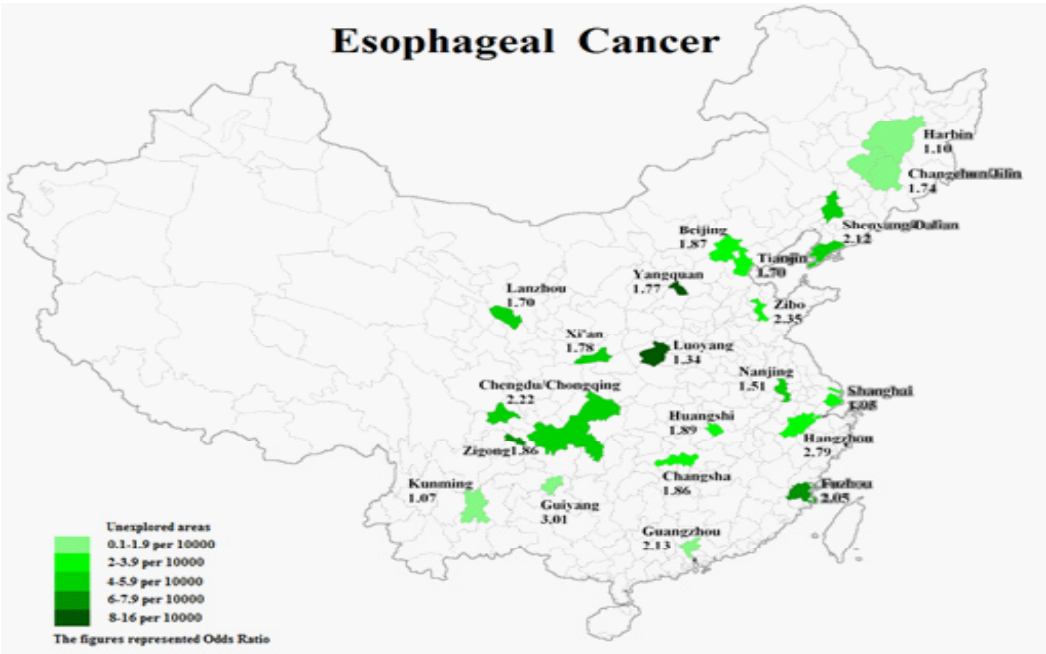

Figure 8. The distribution of mortality rate of esophagus cancer and its association with smoking among men in 24 cities in China

Figure 5 to Figure 8 present the regional features of mortality rates of four major cancers and their association with smoking. The background of the map represents the level of cancer mortality in the city and the darker color means a higher mortality rate. The figures next to the city represent the ORs. The results reveal that, although there is a considerable regional difference in cancer mortality, the associations between smoking and major cancer deaths are significant across all 24 cities. However, the distribution of the risk of smoking and major cancer death is not necessarily congruent with the distribution of cancer deaths.

3.3 Equivalence test and performance in the context of various sample size

Table 7 and 8 show the equivalence evaluation results using bootstrap percentile method and parametric interval method to calculate the 95% confidence interval of OR_1/OR_2, respectively. And the results suggest that whether for a primary cancer alone or in combination of them, there is no 95% confidence interval of OR_1/OR_2 exceeding the pre-defined equivalence limit (0.78-1.29), suggesting a high consistence between PMR method and sex-matched case-spouse control design method in estimating the risk of death from primary cancers.

Tables 7. Equivalence evaluation results using bootstrap percentile method

Death cause	Lower limit	Upper limit	Equivalence[a]
Combining primary cancer	0.973	1.015	√
Lung cancer	0.973	1.009	√
Stomach cancer	0.961	1.003	√
Liver cancer	1.001	1.047	√
Esophagus cancer	0.943	0.989	√
Pancreatic cancer	0.979	1.026	√
Bladder cancer	0.921	0.964	√
Prostate cancer	0.894	0.945	√
Minor sites cancer[b]	0.980	1.017	√

[a] Equivalence limit is 0.78-1.29; [b] Combining the lip, oral and throat cancer

Table 8. Equivalence evaluation results using parametric interval method

Death cause	Lower limit	Upper limit	Equivalence[a]
Combining primary cancer	0.974	1.020	√
Lung cancer	0.971	1.012	√
Stomach cancer	0.957	1.010	√
Liver cancer	0.995	1.051	√
Esophagus cancer	0.936	0.993	√
Pancreatic cancer	0.974	1.027	√
Bladder cancer	0.910	0.969	√
Prostate cancer	0.889	0.946	√
Minor sites cancer[b]	0.975	1.023	√

[a]Equivalence limit is (0.78-1.29); [b]Combining the lip, oral and throat cancer

By employing re-sampling techniques, we obtained ORs for lung cancer using PMR and the sex-matched case-spouse control design under various sample size settings, and the results are shown in Figure 9.

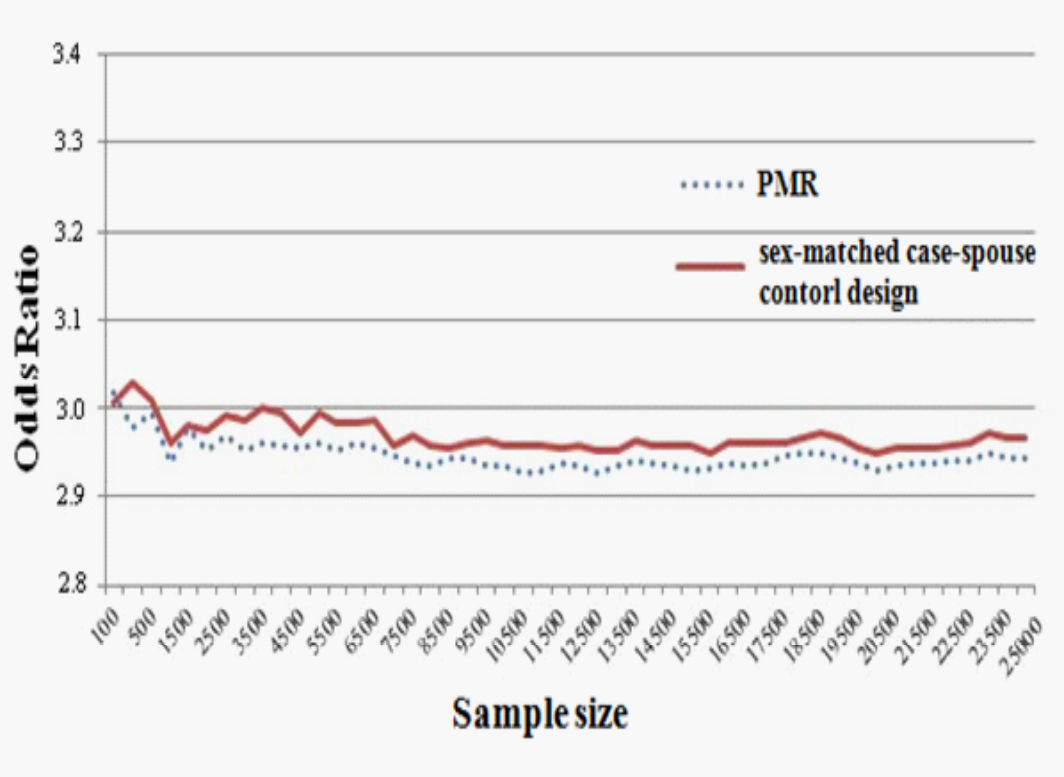

Figure 9. ORs for lung cancers obtained using PMR and the sex-matched case-spouse control design under various sample sizes scenarios

Highly consistent results were observed in the comparison between ORs obtained using two control selection strategies in the context of various sample size, except for the scenario with small to medium sample size (100 to 1500) for which small fluctuations (two contrasting trends) were observed, and with the increase of sample size, the ORs tend to be stable. Moreover, we found that the estimates of ORs for living spouse design methods were always higher than for PMR method.

4. Discussion

4.1. Theoretical exploration for sex-matched case-spouse control design

As a study design with considerable long history, case–control study has been firmly ensconced and being in wide spread use in etiological epidemiology studies, and it is a very efficient way of identifying an association between an exposure and an outcome [16]. Selection of an appropriate control group is critical in case-control study because study conclusions are based on a comparison of the exposure histories provided by cases and controls, therefore, the validity of the results in case-control study heavily relies on the appropriate selection of controls [17]. Overall, the key issue in control selection is that control group should be an unbiased sample of the base population that generated the cases, and theoretically speaking, the "simplest" way to do this is to randomly sample controls from the base population. However, in practice, it might be difficult to define or characterize the base population, especially for large-scale study with nearly a million subjects being investigated, it might be too time-

consuming or otherwise infeasible to obtaining a probability sample of controls from the base population, which makes the selection of controls from the base population the primary challenges.

By enrolling the living spouses of deceased as control group in a nationwide mortality survey, sex-matched case-spouse control design successfully address the problem of selecting controls in a population-based case-control study in large retrospective mortality survey. The development of this design was totally complying with the basic principles for control selecting in case-control study, and there are several crucial assumptions in applying this method, among which the most important one is that the distribution of all causes of deaths in the base population is approximately random, so is their living spouses. Therefore, an approximate random sample of the base population can be obtained by selecting the living spouses as controls. What is noteworthy is that all causes deaths in the base population are assumed to be approximately random, rather than the death from a particular cause, because, as a matter of fact, the distribution of some kinds of deaths have regional characteristics, such as esophagus cancer. Therefore, although such death might happen non-randomly within base population, the all-cause death can still be regarded to be random because any systematical variation can be offset within such vast base population covering 67 million population.

The randomness mentioned above is vital in ensuring the representativeness of the living spouse to the base population. Therefore, it is reasonable to conclude that the living spouses had smoking habits similar to those of the base population, which is reinforced by a study finding that the spouse based prevalence were highly consistent with those in the 1984 and 1996 nationwide surveys of smoking prevalence [18,19].The representativeness of the base population or even general population is crucial in estimating the prevalence of disease and the attributable risk.

Another important assumption for this novel design is that there is no significant relationship in tobacco use between couples, in other words, the smoking status of either party of a couple does not influence the smoking status of his or her spouse. Therefore, the smoking status remains mutually independent between each other for any couples. This assumption is crucial for effect estimation in case-control studies because a correlated smoking habits within couples could undermine the representativeness of living spouse in terms of smoking. Besides, a strong association of smoking habits between couples could also attenuate the estimated risk. To validate this assumption, the Kappa coefficient of agreement test on smoking habits of couples were calculated, which are 0.076 in urban areas, and 0.163 in rural areas, indicating a very weak association between couple's smoking habit. Nevertheless, couples share some common environment (such as living habit, dietary habit, et.al), which seems to be reasonable in China. These shared habits could essentially be removed by cross-matching, through which the distribution of some unknown or unmeasured confounders tend to be same [20].

In the face of a severe and still growing epidemic of smoking in Chinese population during 1980s, the development of sex-matched case-spouse control design method provided a time-saving and resource-economic way in evaluating the hazards of smoking on Chinese population health. In living spouse control design, the exposure information of cases and controls can be collected simultaneously. The convenience of the information collection makes it relatively practical and feasible to be conducted in large-scale study (e.g. a nationwide population based case-control study). Besides, this new design has the advantage in estimating smoking-related deaths for all causes, That is, one exposure to risk of all causes of deaths can be investigated by conducting one study, and risk of smoking on a variety causes of death can be evaluated using one control group. Moreover, from methodological perspective, there is no "gold standard" in epidemiological surveys, therefore,the inclusion of multiple control groups selected by different criteria is preferable to only one control group [21-25]. The development of sex-matched case-spouse control design provide an alternative control selection strategy, affording opportunity in preventing disastrous consequence caused by potential bias and demonstrating consistence in the finding, which thereby enhances the validity of the study.

4.2 Methodological evaluation of the sex-matched case-spouse control design

Although the underlying rationale of the design has gradually gained recognition both home and abroad, there was still a need of methodological evaluation to further validate the methodological rationality and reinforce the validity of this novel design. In view of this point, we used the PMR method to create an "active control" group, which has been treated as a routine control selection design. Then, the comparison could be made between the ORs obtained from the sex-matched case-spouse control design and PMR, and the equivalence of the results could add weight to the rationality and validity of the new design.

Equivalence study has been widely employed in clinical medicine development and clinical trials, with most clinical study activities are aimed at showing that equivalence can also be claimed for generic versions of innovator drugs and for such diverse entities as medical protocols, surgical techniques and medical devices [26-30]. However, there is no such standard criteria for how to evaluating and supporting such equivalence claim in epidemiological survey data [31-33]. Therefore, in our current study, for the first time we applied the equivalence test method in evaluating the consistence of estimated risks obtained through different control selection strategies in case-control study.

The results of equivalence evaluation suggest that the 95% confidence intervals were entirely covered by equivalent limit, indicating a high consistence between the two control groups when estimating the association between smoking and primary cancer death.Our study bears the advantage that a large adequate sample size in each compared group can insure consistency between the initial design and final analysis based on confidence intervals. Moreover,

a large adequate sample size in each compared group will make a high probability (1- β, β is type II error) to insure that the upper/low limit of CIs will not excess the selected criterion. However, it should be noted that an absolute equivalence between the two methods could never be confirmed, and it can only be concluded that the actual difference is unlikely to exceed a particular range, which is the equivalent limit determined by the size of the study and the specified probability of error.

In addition, it is worth noting that the sex-matched case-spouse control design method was initially developed and applied in the context of nationwide mortality survey with extreme large sample size. To investigate the stability of this novel design under various sample size, which is crucial for its further extensive application, we took lung cancer as an example and employed re-sampling techniques to evaluate the stableness of the estimation of risk. The result shows that, although both PMR and living spouses control method showed a certain degree of fluctuation for sample sizes less than 2000, the absolute volatility is negligible (with minimum of 2.97 and maximum of 3.02), which provides a guarantee for its application in case-control studies with relative small sample size. Moreover, the estimate of OR for living spouses design is higher than PMR under various sample size, which indicates that the novel design is more sensitive and may provide more accurate estimate of the hazard of tobacco smoking in general population.

4.3 The hazard of smoking on primary cancer among Chinese urban adult men

To illustrate the application of sex-matched case-spouse control design, we investigated the relationship between smoking and death from various primary cancers among urban adult men in China, these cancer deaths accounted for a considerable fraction of the overall cancer deaths.The results indicated that the effect of the smoking is highest for lung cancer, which is the fourth leading cause of cancer death for Chinese men [34], and is the primary cause of cancer mortality in some cities. Our findings clearly illustrate that lung cancer is about three times as common among urban male smokers as among otherwise similar non-smokers. Moreover, despite the effect is relative trivial in magnitude, we demonstrated a significant association between cigarette smoking and prostate cancer, which remains a matter of debate and previous published data suggested no clear evidence of a causal relationship [35].

After stratified by age groups, we observed increased risk of death from the four major cancers (lung cancer, stomach cancer, liver cancer, esophagus cancer). However, the risk peaked at around 60-70 years old and then declined, which might be explained by potential competitive risk from other diseases, such as the respiratory disease and cardiovascular disease, which also cause considerable burden in China, especially for older people.

Estimating the dose-response relationships between smoking and cancer deaths can serve as important tools in assessing the long-term harm of smoking on health. Although most indi-

viduals realize that smoking is dangerous to their health, many smokers do not comprehend the actual magnitude of long-term exposure. Our findings clearly illustrate that the risk in various kinds of cancer deaths increased consistently with years of smoking and the number of cigarettes smoked daily regardless of age and area. Especially for lung cancer, the risk observed in long term heavy smokers (≥30 years and ≥20 cigarettes/day) was higher than other duration and dose groups. Those findings illustrate the accumulative effects of smoking is potential and harmful, and for current smokers, immediate cessation should be highly encouraged because it offers the only realistic way to avoid a substantial increase in lung cancer mortality caused by further continuation of smoking.

It is a remarkable fact that the risks of smoking in this study for some cancer death are not as strong as those observed in western developed countries (e.g. esophagus cancer and lung cancer) [36], which might be due to:(1) the absolute level of death is higher in China; (2) although China has a long history of smoking, the peak of consumption of cigarette lagged behind the developed countries for nearly 50 years, and the harm caused by smoking was still in an early stage.

In addition, although the smoking prevalence and cancer mortality varied widely among different regions, the association between smoking and primary cancer deaths were significant statistically in all the cities. However, the regional distributions of the risk of smoking on death from main cancers (lung cancer, stomach cancer, liver cancer and esophagus cancer) were not congruent with the regional distribution of the mortalities for primary cancer, that is, the risk is not necessarily high in cities with a relative high cancer mortality, while for some cities with lower mortality rates, higher relevance between smoking and cancer deaths be observed. For example, the strength of the association between smoking and lung cancer deaths in Luoyang is the third highest among the 24 cities, with OR up to 3.85, While lung cancer mortality was 8.50/ million, which was below the average level. This could be explained by the relative high absolute mortality rate for those cities with high cancer death rates, which might attenuate the relative risk for smokers.

Some limitations of this study must be considered when interpreting the results. First, the mortality survey included 90% of all deaths, and 10% of the identified death certificates failed to list an informant (spouse and others). These non-ascertained cases and controls might differ with regard to their smoking status when compared to the ascertained cases and controls, which may lead to selection bias. However, given the overall high ascertainment rate and the large sample size, the results are not likely to be materially affected. Second, not all of the deaths were pathologic diagnosed. For instance, 14.4% of urban esophagus deaths were diagnosed only by clinical symptoms or by other methods, which might lead to some misclassification. Selection bias could occur if these misclassified causes of death were associated with a lower prevalence of smoking and the relative risk of smoking would tend to be underestimated

in this study. Third, recall bias, a measurement error from surviving spouses that might attenuate the association between smoking and deaths from primary cancers. Finally, we made no direct adjustment of those variables that also influence the occurrence of cancer, such as alcohol consumption, environmental pollution and socioeconomic status.

5. Conclusion

As an alternative control selection method in case-control study design, sex-matched case-spouse control design can be considered to be a valid and efficient control selection method for population based large-scale case-control study (e.g. nationwide case-control study). The most prominent advantage of this design is that an approximate random sample can be drawn through this method in a time-saving and resource-economical manner.

6. Acknowledgments

Our work was supported by Cancer Research UK, the UK Medical Research Council, the US National Institutes of Health, the Chinese Ministry of Health, and the Chinese Academy of Medical Sciences. We thank Professor Richard Peto, who gave us great support for the project. The cooperation of the local government, the thousands of doctors, nurses, and other field workers who conducted the surveys, and the million interviewees are gratefully acknowledged.

7. References

1. Johnston L. Smoking and Carcinoma of the Lung. Bmj Clinical Research. 1950;2(4686):1004.

2. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors: Springer-Verlag; 1982. : 901–11. p.

3. Dayan A. Tobacco. A Major International Health Hazard. Journal of Clinical Pathology. 1988;41(2):238.

4. Schultz H. Tobacco or health: A global status report1997. 1-57 p.

5. Doll R. Tobacco: a medical history. Journal of Urban Health. 1999;76(3):289-313.

6. Hamann S. Tobacco or health? Physiological and social damages caused by tobacco smoking. Jama the Journal of the American Medical Association. 2010;304(21):2419.

7. Peto, Richard. Talking points imperial cancer research fund press release: "mortality from smoking in developed countries 1950-2000".

8. Peto R. Tobacco: UK and China. Lancet. 1986;2(8514):1038.

9. Peto R. Tobacco-related deaths in China. Lancet. 1987;2(8552):211.

10. Yu JJ, Mattson ME, Boyd GM, Mueller MD, Shopland DR, Pechacek TF, et al. A Comparison of Smoking Patterns in the People's Republic of China With the United States: An Impending Health Catastrophe in the Middle Kingdom. Jama the Journal of the American Medical Association. 1990;264(12):1575.

11. Mackay J, Crofton J. Tobacco and the developing world. British Medical Bulletin. 1996;52(1):206-21.

12. Thadhani R, Tonelli M. Cohort studies: marching forward. Clinical Journal of the American Society of Nephrology. 2006;1(1):1117-23.

13. Barker DJP, Hall AJ. Practical epidemiology. 4th edition 1991.

14. Liu BQ, Peto R, Chen ZM, Boreham J, Wu YP, Li JY, et al. Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. Bmj Clinical Research. 1998;317(7170):1411.

15. Inman HF, Bradley EL, Jr. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. Communication in Statistics- Theory and Methods. 1989;18(10):3851-74.

16. Paneth N, Susser E, Susser M. Origins and early development of the case-control study: Part 1, Early evolution. International Journal of Public Health. 2002;47(5):282-8.

17. Wacholder S, McLaughlin Jk Fau - Silverman DT, Silverman Dt Fau - Mandel JS, Mandel JS. Selection of controls in case-control studies. I. Principles. Am J Epidemiol.135(9):1019-28.

18. Yang G, Fan L, Tan J, Qi G, Zhang Y, Samet JM, et al. Smoking in China: findings of the 1996 National Prevalence Survey. Jama the Journal of the American Medical Association. 1999;282(13):1247-53.

19. Weng XZ, Hong ZG, Chen DY. Smoking prevalence in Chinese aged 15 and above. Chinese Medical Journal. 1987;100(11):886-92.

20. Jiang J, Liu B, Nasca PC, Han W, Zou X, Zeng X, et al. Comparative study of control selection in a national population-based case-control study: Estimating risk of smoking on cancer deaths in Chinese men. International Journal of Medical Sciences. 2009;6(6):329.

21. Ibrahim MA, Spitzer WO. The case control study: The problem and the prospect *. Journal of Chronic Diseases. 1979;32(1-2):139-44.

22. Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. New England Journal of Medicine. 1978;299(20):1089-94.

23. Hulka BS, Grimson RC, Greenberg BG, Kaufman DG, Jr FW, Hogue CJ, et al. “Alternative” controls in a case-control study of endometrial cancer and exogenous estrogen. 1969;112(3):376-87.

24. Boissel, JeanPierre, Collet, JeanPaul, Laurence, Ducruet, et al. A randomized comparison of the effect of four antihypertensive monotherapies on the subjective quality of life in previously untreated asymptomatic patients: field trial in general practice. Journal of Hypertension. 1995;13(9):1059-67.

25. Ma X, Buffler PA, Layefsky M, Does MB, Reynolds P. Control Selection Strategies in Case-Control Studies of Childhood Diseases. American Journal of Epidemiology. 2004;159(10):915-21.

26. Wellek S, Blettner M. Establishing Equivalence or Non-Inferiority in Clinical Trials: Part 20 of a Series on Evaluation of Scientific Publications. Deutsches ?rzteblatt international. 2012;109(41):674-9.

27. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. Bmj Clinical Research. 1996;313(7048):36-9.

28. Greene WL, Concato J, Feinstein AR. Claims of Equivalence in Medical Research: Are They Supported by the Evidence? Annals of Internal Medicine. 2000;132(9):715-22.

29. Fuller RW, Hallett C, Dahl R. Assessing equivalence of inhaled drugs. Respiratory Medicine. 1995;89(8):525-7.

30. Dong BJ, Hauck WW, Gambertoglio JG, Gee L, White JR, Bubp JL, et al. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. Jama the Journal of the American Medical Association. 1997;277(15):1205.

31. Mikolajczyk R. Methods and Concepts of Epidemiology: Springer New York; 2009. 193-208 p.

32. Lopez AD. Counting the dead in China. Measuring tobacco's impact in the developing world. Bmj Clinical Research. 1998;317(7170):1399-400.

33. Schuirmann DJ. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Pharmacodynamics. 1987;15(6):657-80.

34. Li B, Li JY. [National survey of cancer mortality in China]. Zhonghua zhong liu za zhi [Chinese journal of oncology]. 1980;2(1):1.

35. Huncharek M, Haddock KS, Reid R, Kupelnick B. Smoking as a Risk Factor for Prostate Cancer: A Meta-Analysis of 24 Prospective Cohort Studies. American Journal of Public Health. 2010;100(4):693-701.

36. Vioque J, Barber X, Bolumar F, Porta M, Santibáñez M, Mg DLH, et al. Esophageal cancer risk by type of alcohol drinking and smoking: a case-control study in Spain. BMC Cancer. 2008;8(1):221