Medical Statistics

Research on smoking and lung cancer

a landmark in the history of chronic disease epidemiology

cigarettes had been linked with fashion.

Lung cancer incidence was remarkably rising.

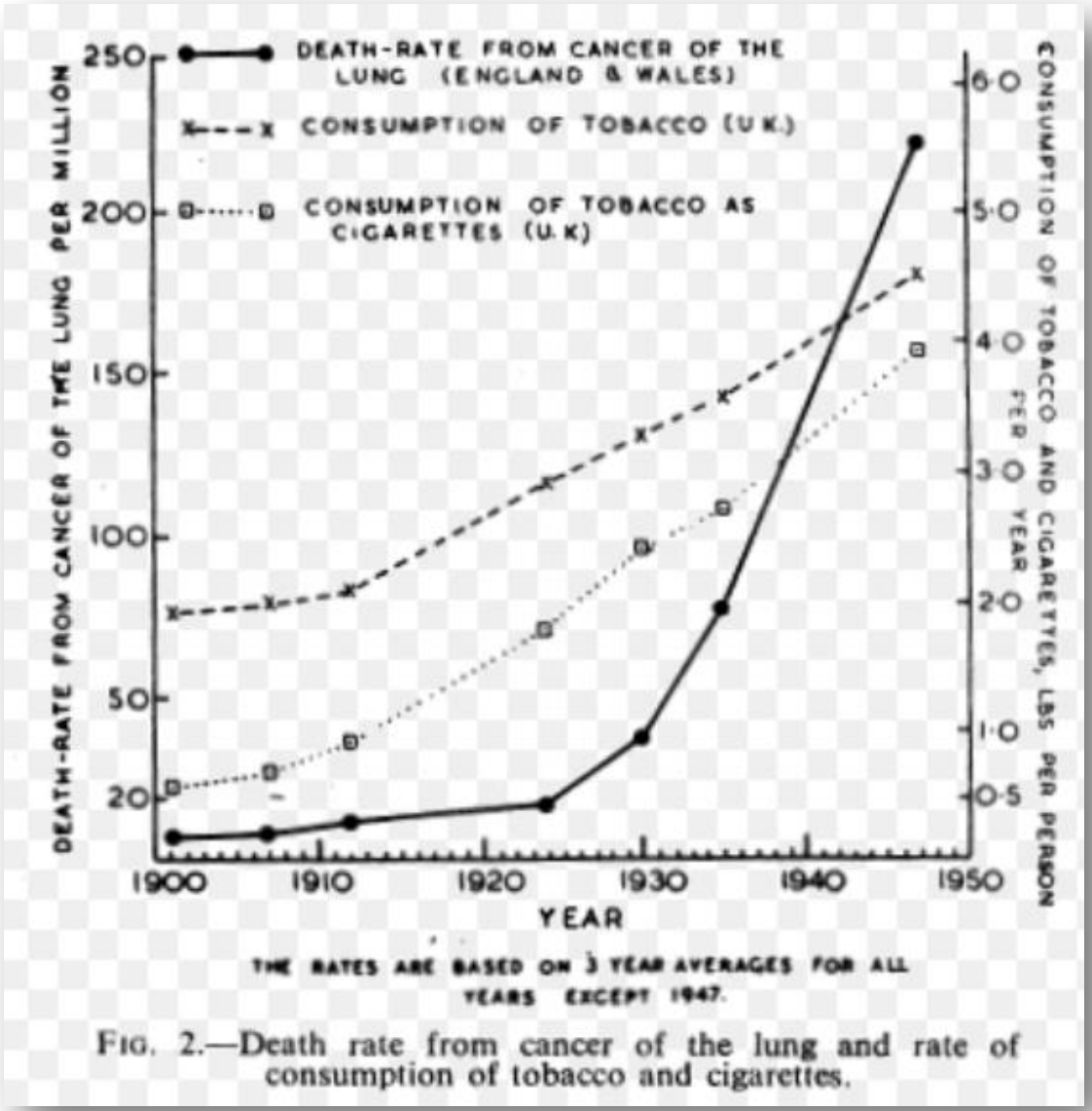
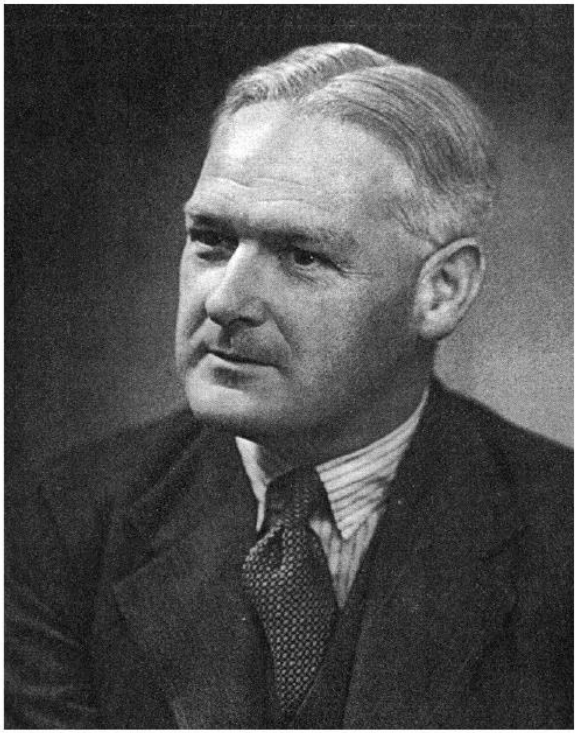


FIG. 2.—Death rate from cancer of the lung and rate of consumption of tobacco and cigarettes.

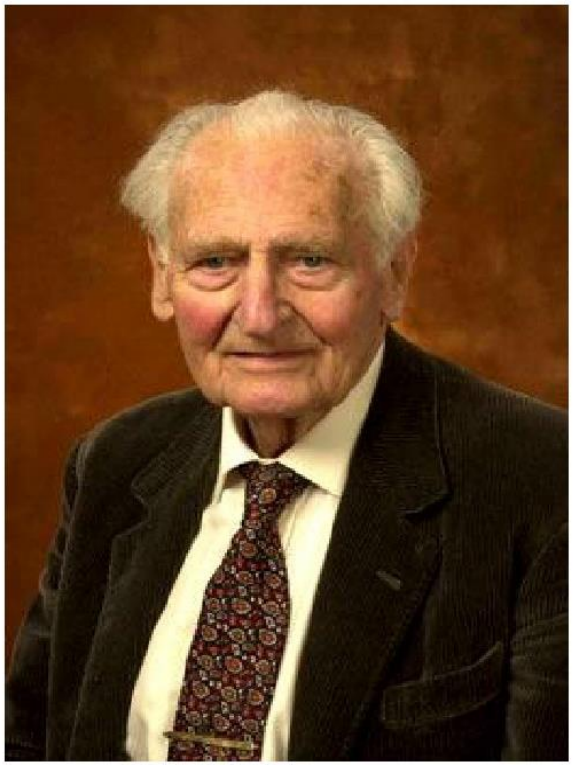
In 1948，Austin Bradford Hill and William Richard Doll，statisticians in the UK, started to explore the causes of lung cancer。



Sir Austin B. Hill: The Father of Modern Epidemiology

现代流行病学的开山鼻祖 Austin B. Hill 爵士

1897-1991



Sir Richard Doll: The Godfather of Epidemiology

流行病学泰斗、三朝元老 Richard Doll 爵士

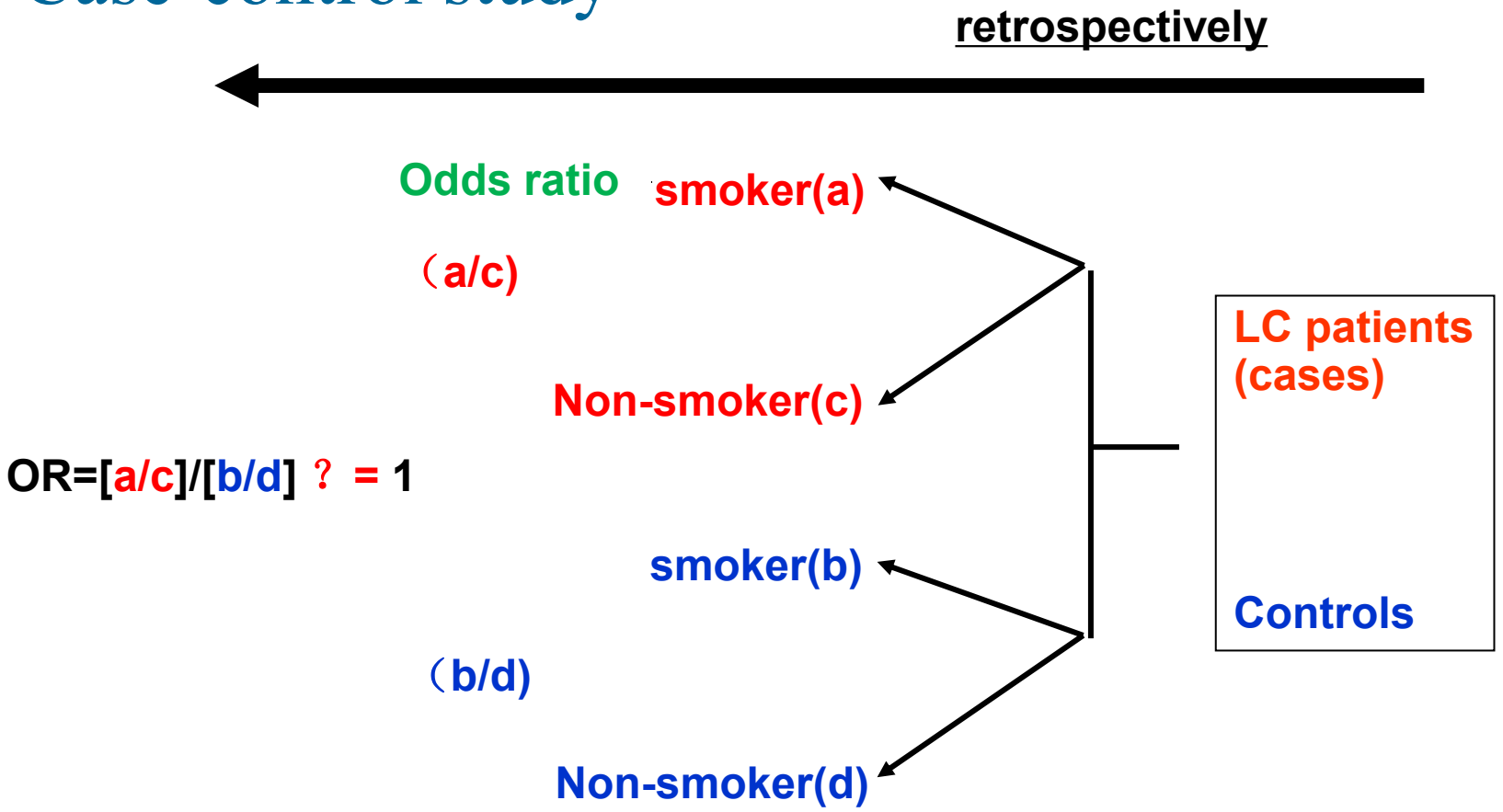
1912-2005

British Doctors Study/Doll and Hill study on Lung cancer

- **Case-control study—in 1948, Doll and Hill enrolled patients with and without lung cancer, who had been admitted to 20 large London hospitals. The results, documented in the paper "Smoking and Carcinoma of the Lung: Preliminary Report," indicated that a relationship existed between tobacco smoking and lung disease.**

Observational study

Case-control study



Note: as you begin with people who already have the disease, you cannot calculate incidence or prevalence

$$\text{Odds} = \frac{\text{Probability of event}}{\text{Probability of non-event}}$$

Data managing of case control study

Exposure/ smoking	Disease		total
	case	control	
Yes	a	b	$a+b=n_1$
no	c	d	$c+d= n_2$
total	$a+c=m_1$	$b+d=m_2$	**a+b+c+d**=T

- OR（odds ratio），OR= **[a/c]/[b/d]**.
- **OR how many times more likely were the cases to have been exposed than the controls.**
- **OR＞1 indicates increased occurrence of an event, is an positive association/risk factor.**
- **OR＜1, decreased, negative, protective factor.**

Characters, advantage and disadvantage

- Characters:
  - Enroll individuals with/without disease
  - Exposure was obtained via recall
  - Outcome-cause。
  - Association, instead of cause

Characters, advantage and disadvantage

Advantage

- Easy to conduct as no follow up is required
- gives faster results
- Inexpensive
- Suitable for rare and newly identified diseases
- More than one risk factors can be studied simultaneously
- Ethical problem lesser as disease has already occurred

Characters, advantage and disadvantage

Disadvantage

- The study depends upon the history given by the subject. Hence recall bias can occur
- Sometimes finding and appropriate control is difficult
- Selection bias
- can only establish an association (NOT causal) between the disease and the risk factor as we don't know what occurred first
- can't know the incidence hence can't calculate the relative risk (RR) directly from a case-control study

Doll and Hill' case control study：

709 lung cancer patients (20 hospitals in London)

709 individuals without LC from the same hospitals

smoke	case	control	total
yes	688 (a)	650 (b)	1338
no	21(c)	59(d)	80
total	709	709	1418

OR：(a/c)/(b/d)=(688/21)/(650/59)=2.97

The probability of LC among smokers is 2.97 times to nonsmokers

Subgroup analysis: male

smoke	case	control	total
yes	647(a)	622(b)	1269
no	2(c)	27(d)	29
Total	649	649	1298

OR：(a/c)/(b/d)=(647/2)/(622/27)=14.04

Doll quitted smoke during the research

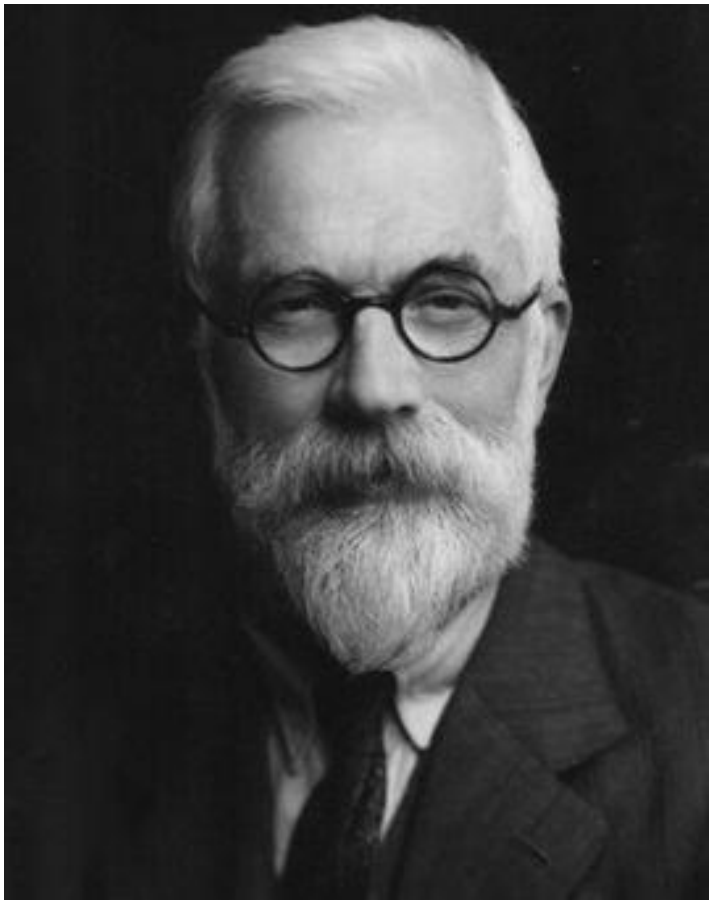
Subgroup analysis: female

smoke	case	control	total
yes	41(a)	28(b)	69
no	19(c)	32(d)	51
total	60	60	120

OR：(a/c)/(b/d)= (41/19)/(28/32)=2.47

Debates

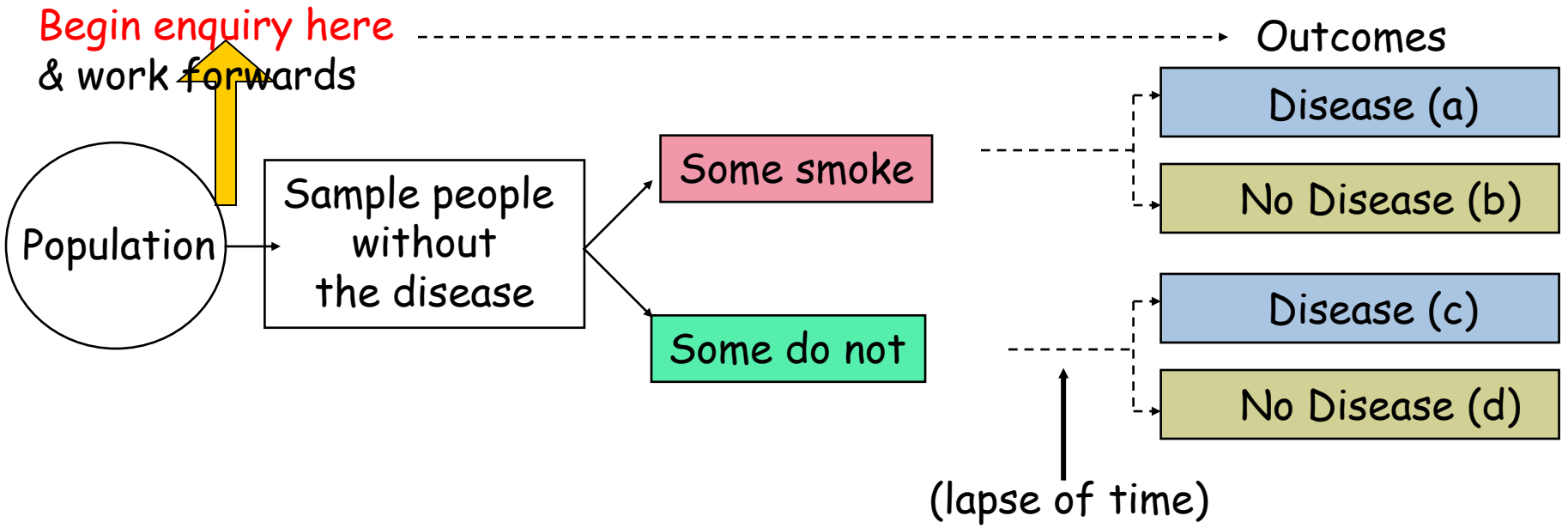
Ronald A. Fisher, the Father of Modern Statistics didn't Believe Smoking Caused Cancer: "People suffering from LC may smoke to relieve their lung irritation".



British Doctors Study

- **Cohort study--in 1951, they mailed a questionnaire about smoking habits to all the doctors resident in the UK through the British Medical Register. 59 600 men and women answered, but data from subjects younger than 35 years old and from females were excluded since lung cancer was rare in these cohorts and to number of women were too few. As a consequence, they collected 34 440 questionnaires from male doctors and started long-term observation of their mortality.**

Observational study



Statistic = Relative Risk [RR] = [a/(a+b)] divided by (c/(c+d))

This shows the ratio of incidence in exposed compared to non-exposed.

RR > 1 implies a hazard;

RR < 1 implies a protective factor

95% CI are usually presented:

e.g., RR = 1.9 (95% CI 1.5, 2.3)

Note: as you begin with people who do not have the disease, you can calculate incidence but not prevalence. (Prevalence would be underestimated as you omitted existing cases)

data

Data managing of cohort study

	case	control	total
exposure	a	b	a+b
Non-exposure	c	d	c+d
total	a+c	b+d	a+b+c+d

interpret

- RR(Risk ratio), RR= **[a/(a+b)]/[c/(c+d)]**
- **RR also reflects the association between exposure and outcome。**
- **RR＞1 indicates a positive association, a risk factor。**
- **RR＜1 implies a negative association, a protective factor 。**

Characters, advantage and disadvantage of cohort study

- Characters:
  - Chronological: exposure occurs before outcome
- Advantage:
  - provides clarity of the temporal sequence
  - A calculation of incidence
  - Facilitate the study of rare exposure
  - Allows for multiple exposures/outcomes
  - Prevent selection bias

Characters, advantage and disadvantage of cohort study

Advantage:

- provides clarity of the temporal sequence
- A calculation of incidence
- Facilitate the study of rare exposure
- Allows for multiple effects
- Prevent selection bias
- provides a significant recall structure to reduce error rates.
- allows for data collection from several sources.

Characters, advantage and disadvantage of cohort study

- Disadvantage:
  - requires researchers to follow a large population for a significant amount of time, especially for rare disease.
  - The time and cost of a prospective cohort study may not be manageable.
  - **Prone to confounding**
  - Participants can change their classification.

Started in oct.1951, the 1st follow-up was performed in March 1954

- **smoker： 36 died because of LC**
- **Non-smoker ： 0**

The Mortality of Doctors in Relation to their Smoking Habits: A Preliminary Report

1956, the 2nd follow-up

TABLE V.—*Standardized Death Rates Per Year Per* 1,000 *Men Aged* 35 *Years or More, in Relation to the Most Recent Amount Smoked**

Cause of Death	No. of Deaths	Death Rate Among:					
		All Men	Non-smokers	All Smokers	Men Smoking a Daily Average of		
					1–14 g.	15–24 g.	25 g. or More
Lung cancer ..	84†	0·81	0·07	0·90	0·47	0·86	1·66
Other cancer ..	220	2·02	2·04	2·02	2·01	1·56	2·63
Other respiratory diseases ..	126	1·10	0·81	1·13	1·00	1·11	1·41
Coronary thrombosis ..	508	4·78	4·22	4·87	4·64	4·60	5·99
Other causes ..	779	6·79	6·11	6·89	6·82	6·38	7·19
All causes ..	1,714	15·48	13·25	15·78	14·92	14·49	18·84

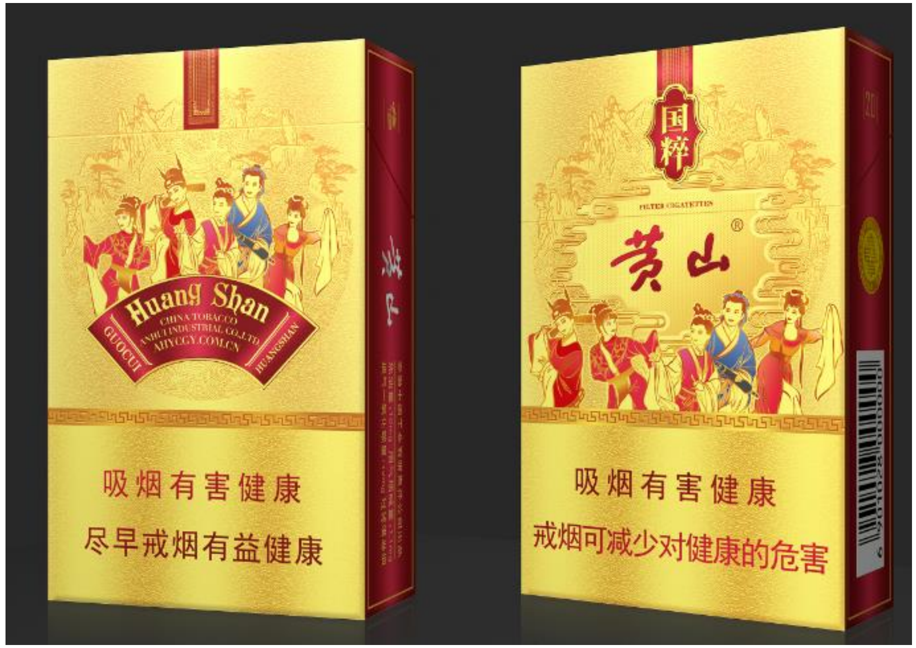
From 1951 to 1973

- Mortality in Relation to Smoking: 22 Years' Observations on Female British Doctors

FDA NEWS RELEASE

FDA requires new health warnings for cigarette packages and advertisements

New warnings with color images promote greater public understanding of negative health consequences of smoking


Huang Shan
GUOCUI
HUANGSHAN
CHINA TOBACCO
ANHUI INDUSTRIAL CO.,LTD
AHYCGY.COM.CN
吸烟有害健康
尽早戒烟有益健康
国粹
黄山
吸烟有害健康
戒烟可减少对健康的危害


YunYan
小熊猫
石林
玉溪
HONGHE
Stone Forest
红塔山
小熊猫
www.nipic.com

Question

- The difference between case-control study and cohort study?
- If a man said: ”my grandpa has been smoking for over 3o years now. but he dose not have lung cancer now”. How to reply him?