After World War II, evidence began mounting that there was a link between cigarette smoking and pulmonary carcinoma (lung cancer). In the 1950s, two now classic articles were published on the subject. One of these studies was conducted in the United States by Wynder and Graham ("Tobacco Smoking as a Possible Etiologic Factor in Bronchiogenic Cancer," *Journal of the American Medical Association, 143*(4)). They found records from a large number (684) of patients with proven bronchiogenic carcinoma (a specific form of lung cancer) in hospitals in California, Colorado, Missouri, New Jersey, New York, Ohio, Pennsylvania, and Utah. They personally interviewed 634 of the subjects to identify their smoking habits, occupation, exposure to dust and fumes, alcohol intake, education, and cause of death of parents and siblings. Thirty-three subjects completed mailed questionnaires, and information for the other 17 was obtained from family members or close acquaintances. Of those in the study, the researchers focused on 605 male patients with the same form of lung cancer. Another 1332 hospital patients with similar age and economic distribution (including 780 males) without lung cancer were interviewed by these researchers in St. Louis and by other researchers in Boston, Cleveland, and Hines, Illinois.[Investigation 1.2.1: Smoking & Lung Cancer, ISCAM]

The following two-way table replicates the counts for the 605 male patients with the same form of cancer and for the "control-group" of 780 males.

	none < 1/day	light 1-9/day	mod heavy 10-15/day	heavy 16-20/day	excessive 21-34/day	chain 35+/day
patients	8	14	61	213	187	122
controls	114	90	148	278	90	60

Given the results of the study, do you think we can generalize from the sample to the population? Explain (what is the difference between the sample and the population here?).

	no smoking	chain smoking	
cancer	8	122	130
healthy	114	60	174
	122	182	304

Group A	Group B
expl = smoking status	expl = lung cancer
resp = lung cancer	resp = smoking status

- If lung cancer is considered a success and no smoking is baseline:

$$RR \quad =$$
$$OR \quad =$$

- If chain smoking is considered a success and healthy is baseline:

$$RR \quad =$$
$$OR \quad =$$

Case-control study: identify observational units by response

Cohort study: identify observational units by explanatory variable

Conclusion: Can we conclude causation here?

Conclusion2: What can you conclude about baseline rates?