

Class 2: Cancer and Causation

17 January 2019

cs4501/econ4559 Spring 2019
David Evans and Denis Nekipelov
<https://uvammm.github.io>

Plan

Course

Why I'm Teaching this Class

Causation

Definitions

Correlation

Cancer

Everyone who wants to take the class (including unregistered students) should have a teammate for Project 1. If you don't, talk to us after class today.

Why I'm Teaching this Class

1. Learn about Economics

PLEASE TYPE, OR PRINT FIRMLY WITH BALL POINT PEN

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MID-TERM FRESHMAN PERFORMANCE EVALUATION

Extra forms and envelopes may be obtained in 7-104 (X3-6771)

Information on these forms may be released outside of MIT only by the student.

ACADEMIC YEAR 1990	<input type="checkbox"/> FALL <input checked="" type="checkbox"/> SPRING
STUDENT'S NAME, TERM ADDRESS (PRINT) DAVID EVANS Soo Memorial Drive, Rm 337	

SUBJECT NUMBER 14.01	RECITATION INSTRUCTOR Xie
ADVISOR'S NAME, MIT ROOM NO. (PRINT) David DEVEAU 14 N - 217B	

STUDENT'S SELF-ASSESSMENT

COMMENT ON YOUR PERFORMANCE to date, including your experience in the subject, your class participation, etc., as suggested on the accompanying instruction sheet. Include any questions you may wish the instructor to answer.

I more or less lost faith in this subject when my answer to the question concerning the relative merits of giving money or food to the homeless, what I thought was a rather insightful political discussion, elicited the response: "No! (It's not subjective). The only thing that is important is the utility function." Economics is not a science, and should not be taught as one. It is a highly subjective matter and purchasing decisions are not made by individuals in the real world by optimizing utility functions and an endless slew of supply and demand graphs but by complex and mostly psychological factors. I realize, that as an introductory course, 14.01 attempts to present simplified models of economic principles which is justified. When these models become a science, or even a dogma, something is terribly wrong and original ideas and thought are repressed by unrealistic theories.

SIGNATURE *David Evans* DATE 19 Mar 1990

INSTRUCTOR'S EVALUATION

PLEASE RESPOND TO STUDENT'S QUESTIONS/CONCERNs in the space below by writing comments and advice and checking boxes as appropriate.

Economics is not a science. It's not taught as one. It's a model based on selfish economics agents. Regarding the comment on your ps., I am sorry if it hurt your feelings. You can come to my office to discuss the issue.

	Exceptionally Good	Good	Adequate	Not Adequate	Not Known
Test Scores and Papers	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Homework	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Class Participation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Overall Understanding of Subject	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

CHECK IF STUDENT'S ATTENDANCE HAS BEEN POOR

SIGNATURE *Huzhanghe* ROOM F52-369 EXTENSION 3-8718 DATE 4/3/90
ADVISOR - (FILE IN ADVISOR FOLDER)

PLEASE TYPE, OR PRINT FIRMLY WITH BALL POINT PEN

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MID-TERM FRESHMAN PERFORMANCE EVALUATION

Extra forms and envelopes may be obtained in 7-104 (X3-6771)

Information on these forms may be released outside of MIT only by the student.

ACADEMIC YEAR 1990	<input type="checkbox"/> FALL	<input checked="" type="checkbox"/> SPRING
STUDENT'S NAME, TERM ADDRESS (PRINT) DAVID EVANS		

SUBJECT NUMBER 14.01	RECITATION INSTRUCTOR Xie
ADVISOR'S NAME, MIT ROOM NO. (PRINT) David DEVEAU	

STUDENT'S SELF-ASSESSMENT

COMMENT ON YOUR PERFORMANCE to date, including your experience in the subject, your class participation, etc., as suggested on the accompanying instruction sheet. Include any questions you may wish the instructor to answer.

I more or less lost faith in this subject when my answer to the question concerning the relative merits of giving money or food to the homeless, what I thought was a rather insightful political discussion, elicited the response: "No! (It's not subjective) The only thing that is important is the utility function." Economics is not a science, and should not be taught as one. It is a highly subjective matter and purchasing decisions are not made by individuals in the real world by optimizing utility functions and an endless slew of supply and demand graphs but by complex and mostly psychological factors. I realize, that as an introductory course, 14.01 attempts to present simplified models of economic principles which is justified. When these models become a science, or even a dogma, something is terribly wrong and original ideas and thought are repressed by unrealistic theories.

SIGNATURE

DATE

19 Mar 1990

INSTRUCTOR'S EVALUATION

Test S

Homework

<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	-------------------------------------	--------------------------

Class Participation

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------

Overall Understanding of Subject

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------

Quality of Writing

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------

CHECK IF STUDENT'S ATTENDANCE HAS BEEN POOR

SIGNATURE

ROOM
F52-369
EXTENSION
3-8718
DATE
4/3/90

ADVISOR - (FILE IN ADVISOR FOLDER)

PLEASE TYPE, OR PRINT FIRMLY WITH BALL POINT PEN
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MID-TERM

E
Information

ACADEMIC YEAR
1990 F/A

STUDENT'S NAME, TERM

DAVID EVANS
Soo Memorial L

COMMENT ON YOUR PERFORMANCE
accompanying instruction sheet. If

I more or less lost faith in giving money or food to elicited the response: "A function." Economics subjective matter and pi optimizing utility functi and mostly psychological present simplified mode become a silence or a

SIGNATURE

David

PLEASE RESPOND TO STUDENT'S
appropriate.

Economics is
a model b
the comment
feelings.)
the issue

Test Scores and Papers
Homework
Class Participation
Overall Understanding of Subject
Quality of Writing

SIGNATURE

CHECK IF STUDENT'S ATTENDANCE HAS BEEN POOR

INSTRUCTOR'S EVALUATION

PLEASE RESPOND TO STUDENT'S QUESTIONS/CONCERNS in the space below by writing comments and advice and checking boxes as appropriate.

Economics is not a science. It's not taught as one. It's a model based on selfish economics agents. Regarding the comment on your ps., I am sorry if it hurt your feelings. You can come to my office to discuss the issue.

	Exceptionally Good	Good	Adequate	Not Adequate	Not Known
Test Scores and Papers	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Homework	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Class Participation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Overall Understanding of Subject	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

ROOM EXTENSION DATE
E52-369 *3-8718* *4/3/90*

ADVISOR - (FILE IN ADVISOR FOLDER)

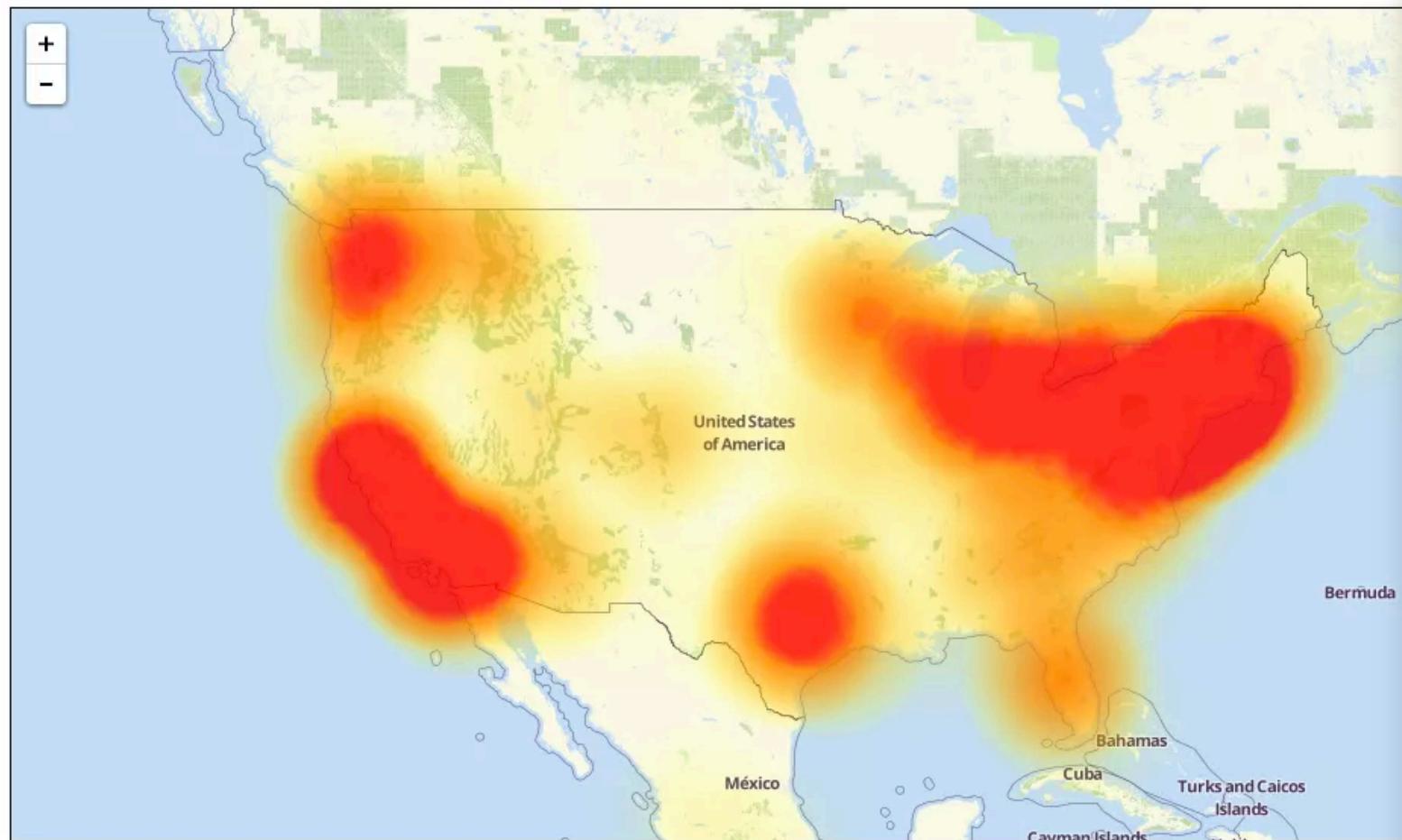
ROOM EXTENSION DATE
E52-369 *3-8718* *4/3/90*

ADVISOR - (FILE IN ADVISOR FOLDER)

2. Economics in Security and Privacy

Mirai Botnet

Level3 outage map



Level3 Communications offers telecommunications services to business customers. Level 3 services include internet connectivity and managed services such as VPN, collaboration, voice and video.



ars TECHNICA

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE STORE

SLEEPING GIANT —

Internet-paralyzing Mirai botnet comes roaring back with new strain

100,000 devices infected in 60 hours by strain that targeted ZyXEL devices.

DAN GOODIN - 11/29/2017, 12:21 PM



[Enlarge](#) / One of the modems targeted by a newly discovered strain of Mirai.

Mirai, the Internet-of-things malware that turns cameras, routers, and other household devices into potent distributed denial-of-service platforms, may be lying low, but it's certainly not dead.

```

123 // Set up passwords
124 add_auth_entry("\x50\x4D\x4D\x56", "\x5A\x41\x11\x17\x13\x13", 10); // root xc3511
125 add_auth_entry("\x50\x4D\x4D\x56", "\x54\x4B\x58\x5A\x54", 9); // root vizxv
126 add_auth_entry("\x50\x4D\x4D\x56", "\x43\x46\x4F\x4B\x4C", 8); // root admin
127 add_auth_entry("\x43\x46\x4F\x4B\x4C", "\x43\x46\x4F\x4B\x4C", 7); // admin admin
128 add_auth_entry("\x50\x4D\x4D\x56", "\x1A\x1A\x1A\x1A\x1A\x1A", 6); // root 888888
129 add_auth_entry("\x50\x4D\x4D\x56", "\x5A\x4F\x4A\x46\x4B\x52\x41", 5); // root xmhdipc
130 add_auth_entry("\x50\x4D\x4D\x56", "\x46\x47\x44\x43\x57\x4E\x56", 5); // root default
131 add_auth_entry("\x50\x4D\x4D\x56", "\x48\x57\x43\x4C\x56\x47\x41\x4A", 5); // root juantech
132 add_auth_entry("\x50\x4D\x4D\x56", "\x13\x10\x11\x16\x17\x14", 5); // root 123456
133 add_auth_entry("\x50\x4D\x4D\x56", "\x17\x16\x11\x10\x13", 5); // root 54321
134 add_auth_entry("\x51\x57\x52\x52\x4D\x50\x56", "\x51\x57\x52\x52\x4D\x50\x56", 5); // support support
135 add_auth_entry("\x50\x4D\x4D\x56", "", 4); // root (none)
136 add_auth_entry("\x43\x46\x4F\x4B\x4C", "\x52\x43\x51\x51\x55\x4D\x50\x46", 4); // admin password
137 add_auth_entry("\x50\x4D\x4D\x56", "\x50\x4D\x4D\x56", 4); // root root
138 add_auth_entry("\x50\x4D\x4D\x56", "\x13\x10\x11\x16\x17", 4); // root 12345
139 add_auth_entry("\x57\x51\x47\x50", "\x57\x51\x47\x50", 3); // user user
140 add_auth_entry("\x43\x46\x4F\x4B\x4C", "", 3); // admin (none)
141 add_auth_entry("\x50\x4D\x4D\x56", "\x52\x43\x51\x51", 3); // root pass
142 add_auth_entry("\x43\x46\x4F\x4B\x4C", "\x43\x46\x4F\x4B\x4C\x13\x10\x11\x16", 3); // admin admin1234
143 add_auth_entry("\x50\x4D\x4D\x56", "\x13\x13\x13\x13", 3); // root 1111
144 add_auth_entry("\x43\x46\x4F\x4B\x4C", "\x51\x4F\x41\x43\x46\x4F\x4B\x4C", 3); // admin smcadmin
145 add_auth_entry("\x43\x46\x4F\x4B\x4C", "\x13\x13\x13\x13", 2); // admin 1111
146 add_auth_entry("\x50\x4D\x4D\x56", "\x14\x14\x14\x14\x14\x14", 2); // root 666666
147 add_auth_entry("\x50\x4D\x4D\x56", "\x52\x43\x51\x51\x55\x4D\x50\x46", 2); // root password
148 add_auth_entry("\x50\x4D\x4D\x56", "\x13\x10\x11\x16", 2); // root 1234
149 add_auth_entry("\x50\x4D\x4D\x56", "\x49\x4E\x54\x13\x10\x11", 1); // root klv123

```

Software Vulnerabilities as *Externalities*

“According to one common view, information security comes down to technical measures. Given better access control policy models, formal proofs of cryptographic protocols, approved firewalls, better ways of detecting intrusions and malicious code, and better tools for system evaluation and assurance, the problems can be solved. In this note, I put forward a contrary view: information insecurity is at least as much due to perverse incentives. **Many, if not most, of the problems can be explained more clearly and convincingly using the language of microeconomics: network externalities, asymmetric information, moral hazard, adverse selection, liability dumping and the tragedy of the commons.**”

Why Information Security is Hard – An Economic Perspective

Ross Anderson

University of Cambridge Computer Laboratory

Ross.Anderson@ccl.cam.ac.uk

30th January 2001

How much should we spend on security?

\$124B

Projected 2019 spending on information security [Gartner]

How much should we spend on security?

- \$1700B** Military spending worldwide (2017)
US: \$610B
- \$265B** Apple's 2018 Revenues
- \$124B** Projected 2019 worldwide spending on information security [Gartner]
- \$3.5B** University of Virginia, 2018 operating budget (~50% Medical)

Half the money I spend on advertising is wasted; the trouble is I don't know which half.

John Wanamaker

Vulnerability

ZERODIUM Payouts for Mobiles*

Up to
\$2,000,000

Up to
\$1,500,000

Up to
\$1,000,000

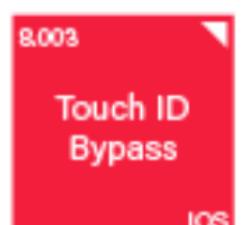
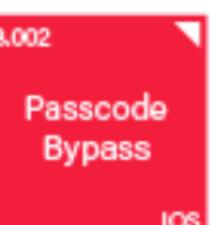
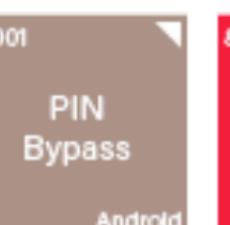
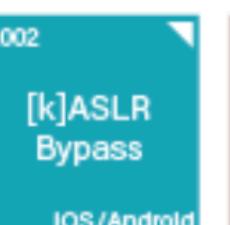
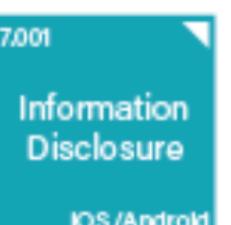
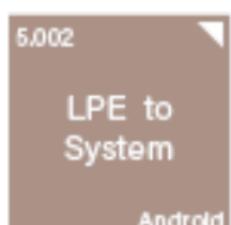
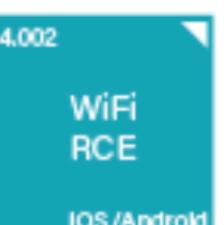
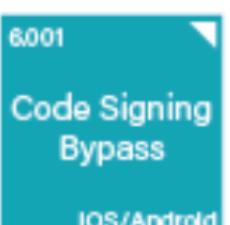
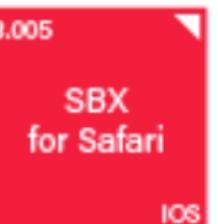
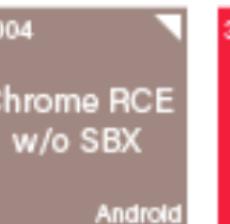
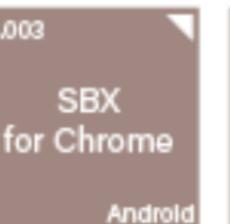
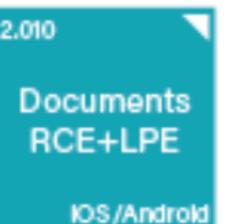
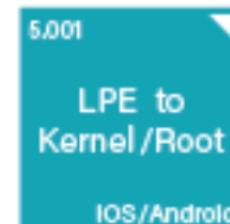
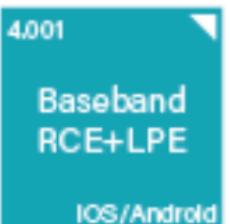
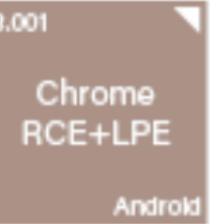
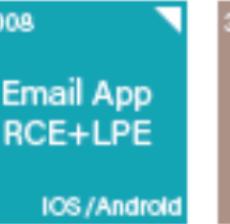
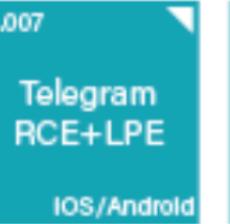
Up to
\$500,000

Up to
\$200,000

Up to
\$100,000

RJB: Remote Jailbreak with Persistence
 RCE: Remote Code Execution
 LPE: Local Privilege Escalation
 SBX: Sandbox Escape or Bypass

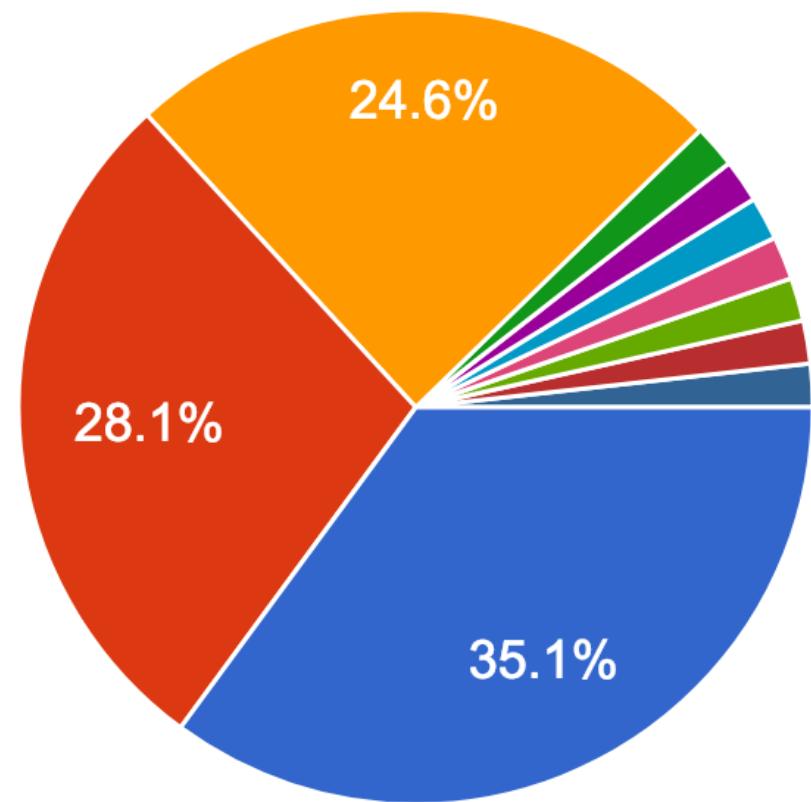
iOS
 Android
 Any OS



3. Experimental Interdisciplinary Course

What is your first major?

57 responses

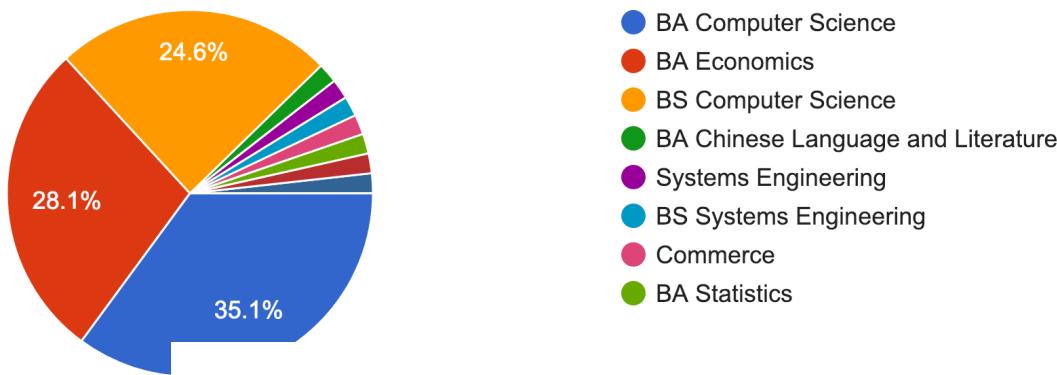


- BA Computer Science
- BA Economics
- BS Computer Science
- BA Chinese Language and Literature
- Systems Engineering
- BS Systems Engineering
- Commerce
- BA Statistics

▲ 1/2 ▼

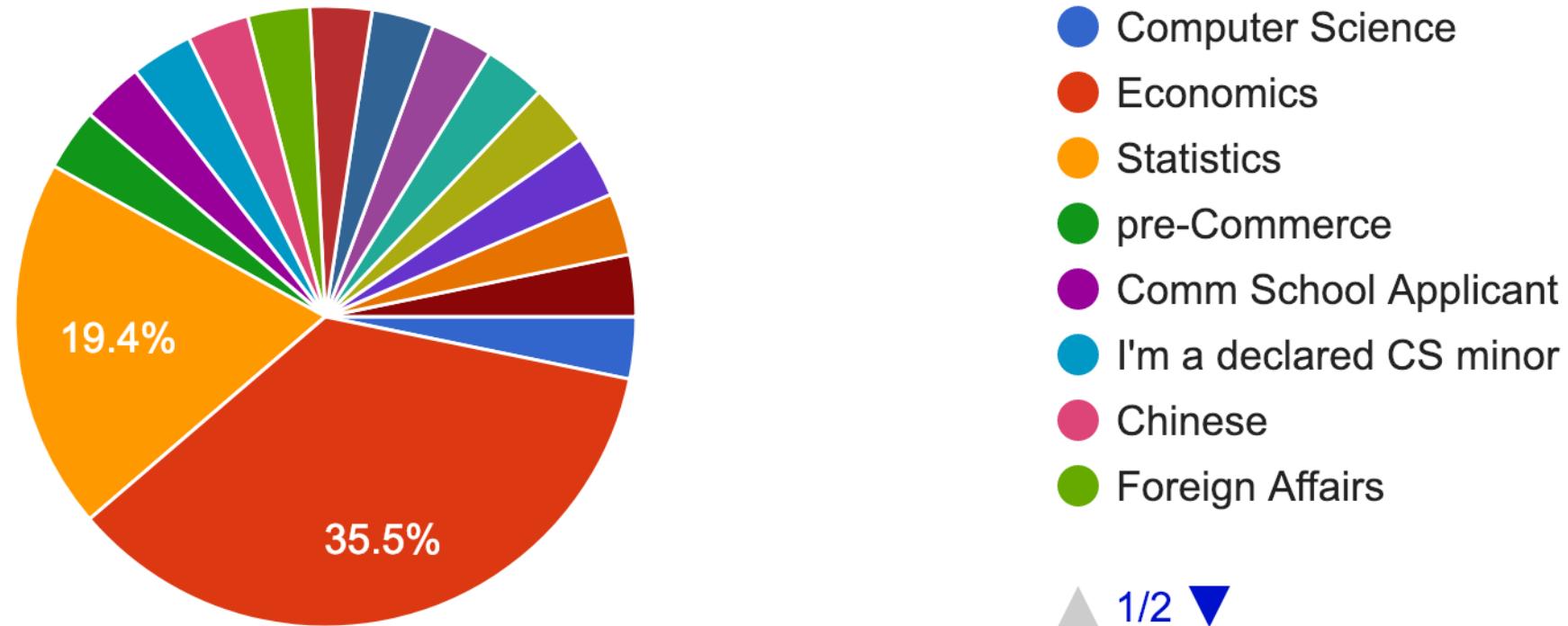
What is your first major?

57 responses



If you have a second major, what is your second major?

31 responses



Course Questions?

Causation

What is the goal of science?

find "truth")

understand natural world

finding patterns, rules

change/control
the world

models → make predictions

**Does smoking
cause cancer?**



CC: Silberio77

Federal Cigarette Labeling and Advertising Act of 1966

§ 1333. Labeling; requirements; conspicuous statement

(a) Required warnings; packages; advertisements; billboards

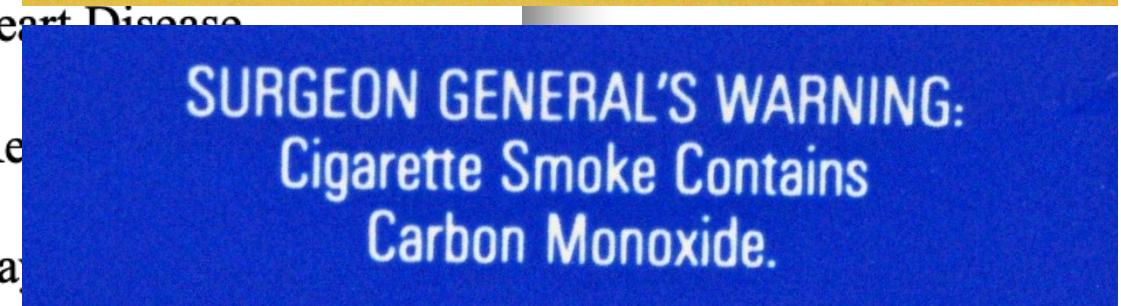
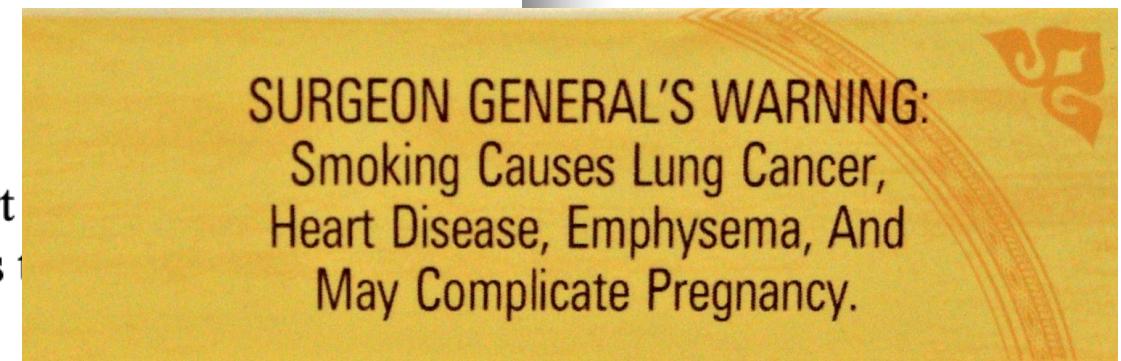
(1) It shall be unlawful for any person to manufacture, package, or import distribution within the United States any cigarettes the package of which fails in accordance with the requirements of this section, one of the following labels:

SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, And May Complicate Pregnancy.

SURGEON GENERAL'S WARNING: Quitting Smoking Now Greatly Reduces Serious Risks to Your Health.

SURGEON GENERAL'S WARNING: Smoking By Pregnant Women May Result in Injury, Premature Birth, And Low Birth Weight.

SURGEON GENERAL'S WARNING: Cigarette Smoke Contains Carbon Monoxide.



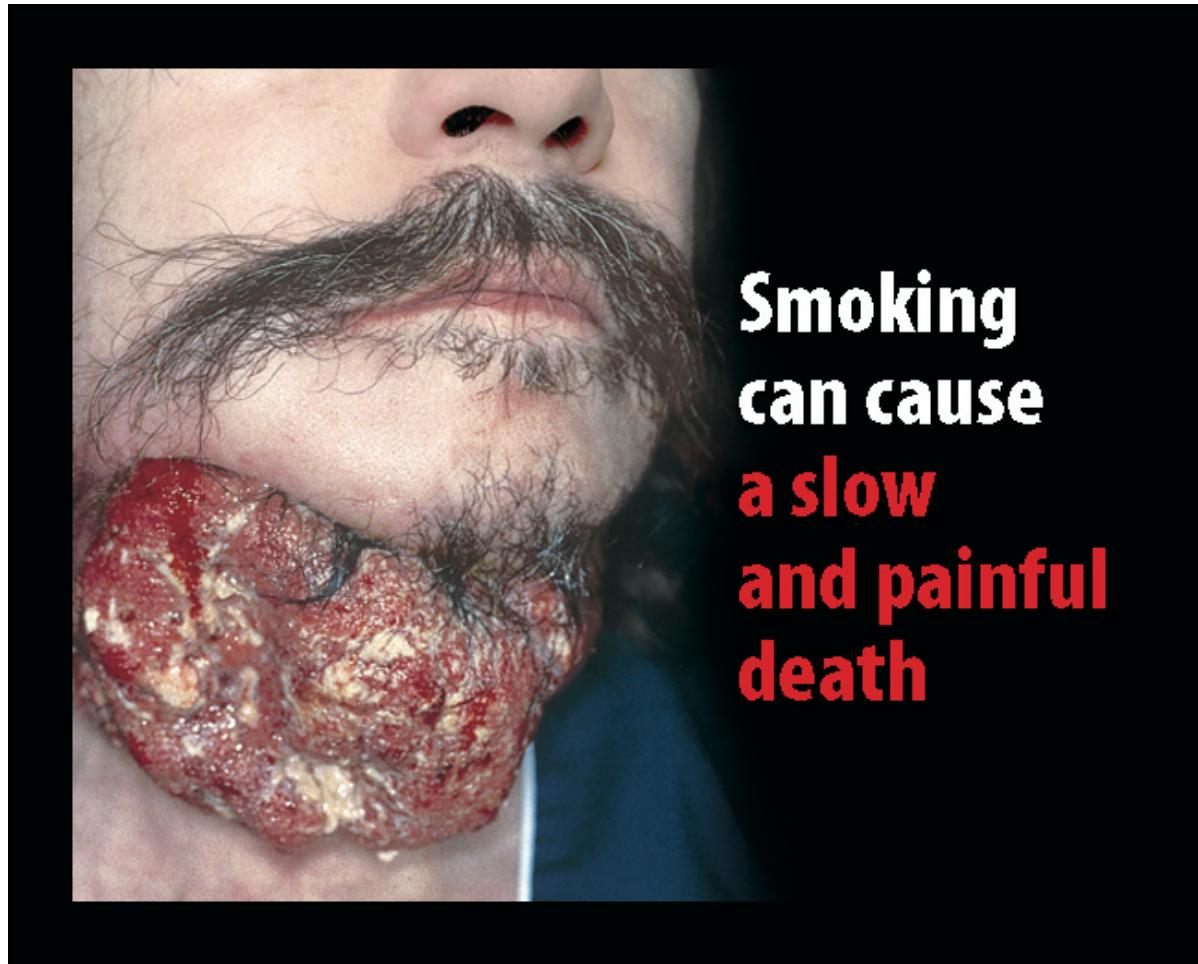
C. Recommendation

Amendment No. 1 would make little change in the label except it's more positive and shorter. Retaining the word "may" does soften its impact somewhat.



WARNING: Cigarettes
cause fatal lung disease.

FDA Proposed Warning (2011),
blocked by companies/courts



UK Warnings



Protect children: **don't make
them breathe your smoke**

Nature 182 (1958 August 30), 596.

276

CANCER AND SMOKING

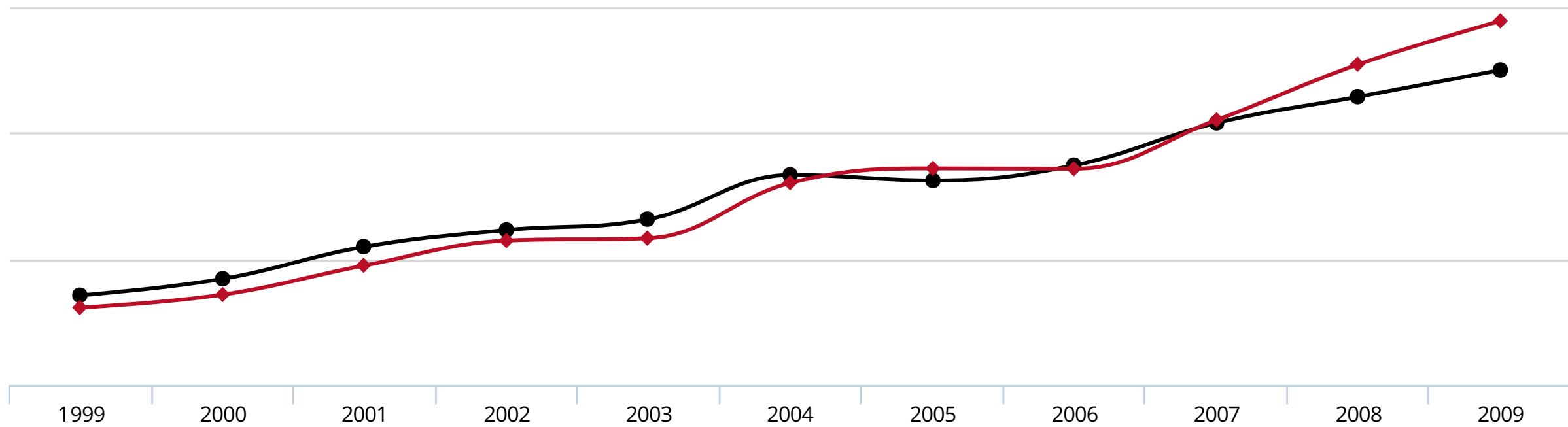
THE CURIOUS ASSOCIATIONS with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer. If, for example, it were possible to infer that inhaling cigarette smoke was a practice of considerable prophylactic value in preventing the disease, for the practice of inhaling is rarer among patients with cancer of the lung than with others.

Such results suggest that an error has been made of an old kind, in arguing from correlation to causation, and that the possibility should be explored that the different smoking classes, cigarette smokers, cigar smokers, pipe smokers, etc., have adopted their habits partly by reason of their personal temperaments and dispositions, and are not lightly to be assumed to be equivalent in their genotypic composition. Such differences in genetic make-up between those classes would naturally be associated with differences of disease incidence without the disease being causally connected with smoking. It would then seem not so paradoxical that the stronger fumes of pipes or cigars should be so much less associated with cancer than those of cigarettes, or that the practice of drawing cigarette smoke in bulk into the lung would have apparently a protective effect.

(Sir) Ronald A. Fisher
Nature, 30 August 1958



Correlation



Random Variable

Definition: a **random variable** (e.g., X) is a distribution of values that is the measured outcome of some experiment.

not
function \times : probability \rightarrow measure

number

Random Variable

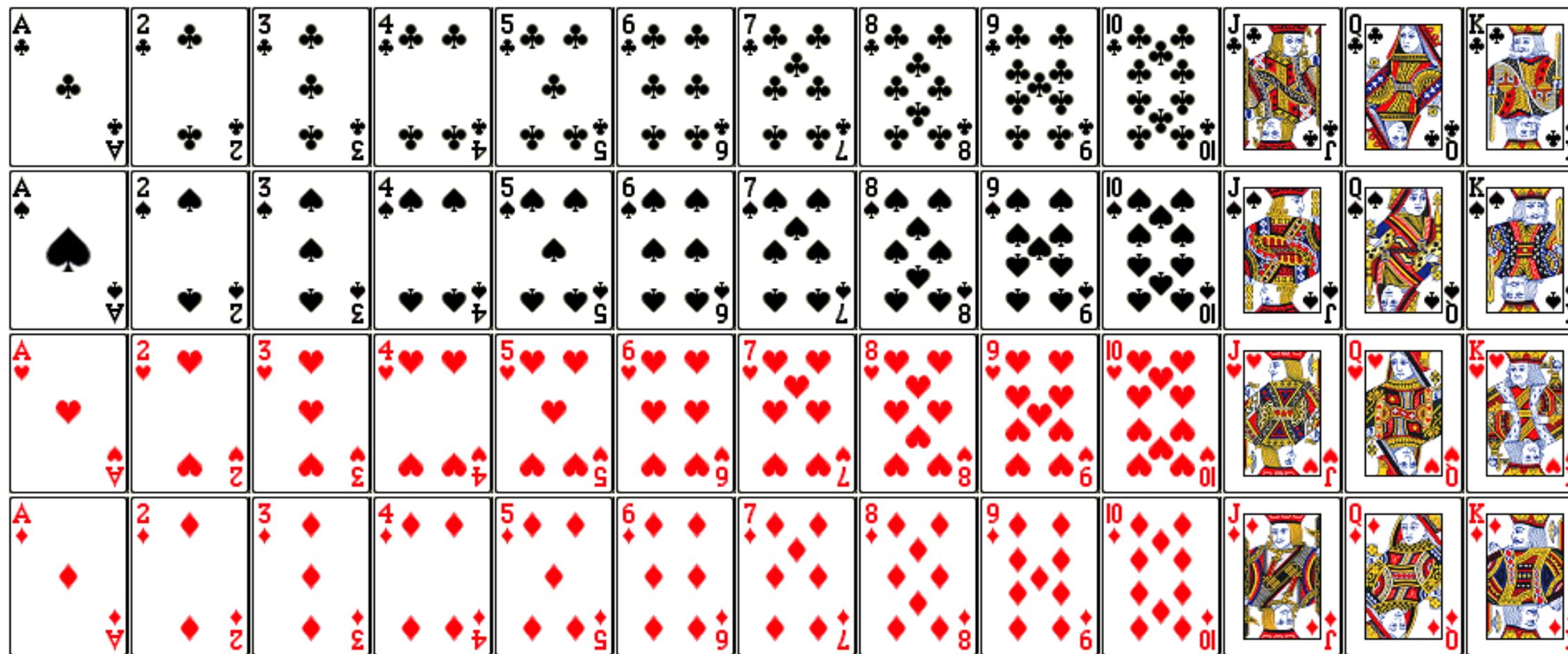
Definition: a **random variable** (e.g., X) is a distribution of values that is the measured outcome of some experiment.

$$X: \Omega \rightarrow E$$

Function from a *probability space* (set of possible outcomes) to a *measurable space* (usually a real numbers)

Example: Playing Card ^{Hindclass}

Think of a playing card



CS

Red Black

Econ

16

10

9

5

CIGARETTES, CANCER, AND STATISTICS

Sir Ronald Fisher

VIC

that if one tries to think of numbers at random, one thinks of numbers very far from at random. If one tries to think of a card of an ordinary playing deck, it's well known (perhaps it's not so well known—it is known to me, at least) that red cards are thought of more readily than black cards, that odd numbers are thought of more readily than even numbers, and that the Queen of Diamonds is a hot favorite. This proclivity of the human mind affects any consciously guided

not for our class

Covariance

Measure of joint variability of two random variables:

$$\text{covariance}(X, Y) = E[(X - \underbrace{E(X)}_{\text{expected}})(Y - \underbrace{E(Y)}_{\text{mean}})]$$

expected

mean

Independent Variables don't Covary

Theorem: If X is independent of Y , $\text{covariance}(X, Y) = 0$.

$$\text{covariance}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$E(X) + E(Y) = E(X + Y)$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

$X = [0, 0, 0, 1, 1, 0, 0, 0]$

$Y = [1, 1, 1, 0, 0, 0]$

only if independent

Independent Variables don't Covary

Theorem: If X is independent of Y , $\text{covariance}(X, Y) = 0$.

$$\begin{aligned}\text{covariance}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY - X \cdot E[Y] - Y \cdot E[X] + E[X] \cdot E[Y]] \\ &= E[XY] - E[X] \cdot E[Y] - E[Y] \cdot E[X] + E[X] \cdot E[Y] \\ &= E[XY] - E[X] \cdot E[Y] \\ &= 0\end{aligned}$$

only when X, Y are independent

Covariance with Itself

$$\text{covariance}(X, X) = ?$$

$$\begin{aligned}\text{covariance}(X, X) &= E[(X - E(X))(X - E(X))] \\ &= E[(X - \mu)^2] \\ &= \text{variance}(X)\end{aligned}$$

Population Covariance

$$cov(X, Y) = \sum_{i=1}^n p_i(x_i - E(X))(y_i - E(Y))$$

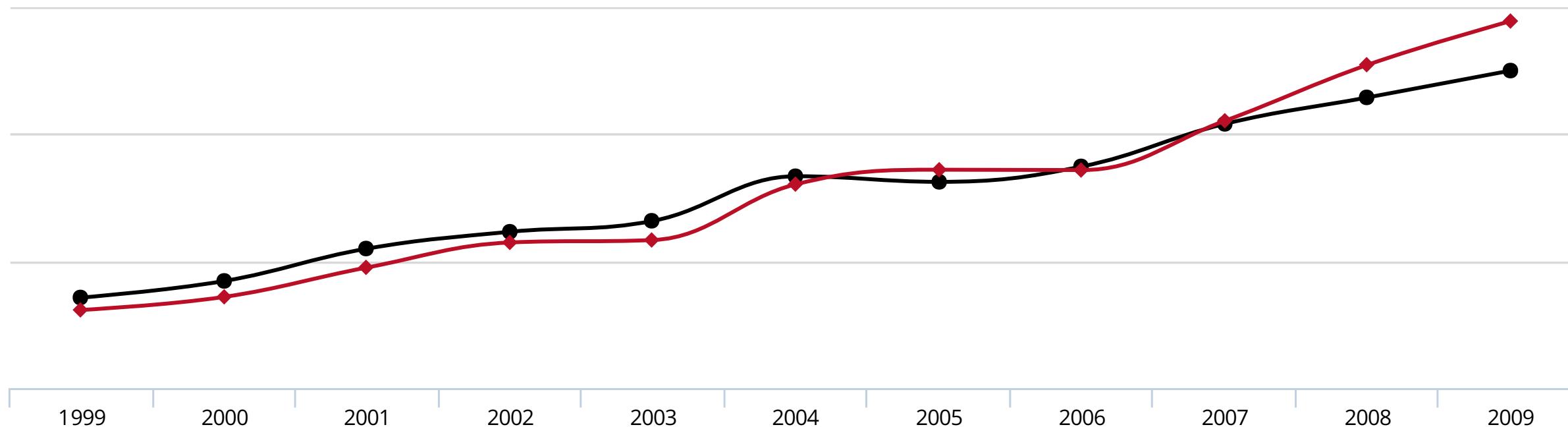
Measuring Correlation

Pearson correlation coefficient

$$\rho_{X,Y} = \frac{\text{covariance}(X, Y)}{\sigma_X \sigma_Y}$$

Correlation

$r = 0.997$



Explanations of Correlation

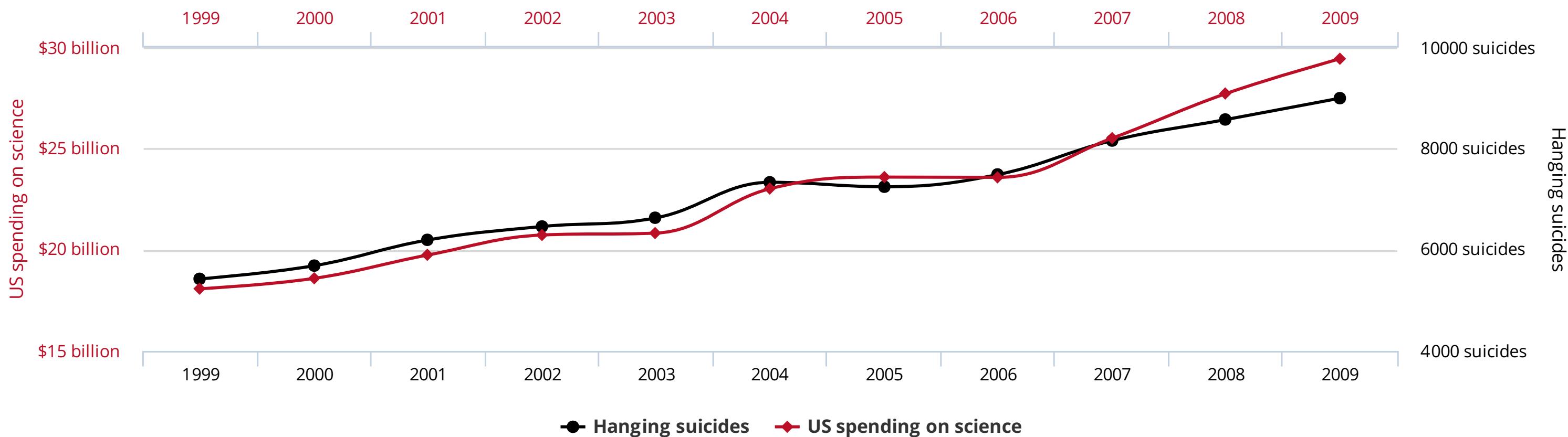
$A \rightarrow B$

$B \rightarrow A$

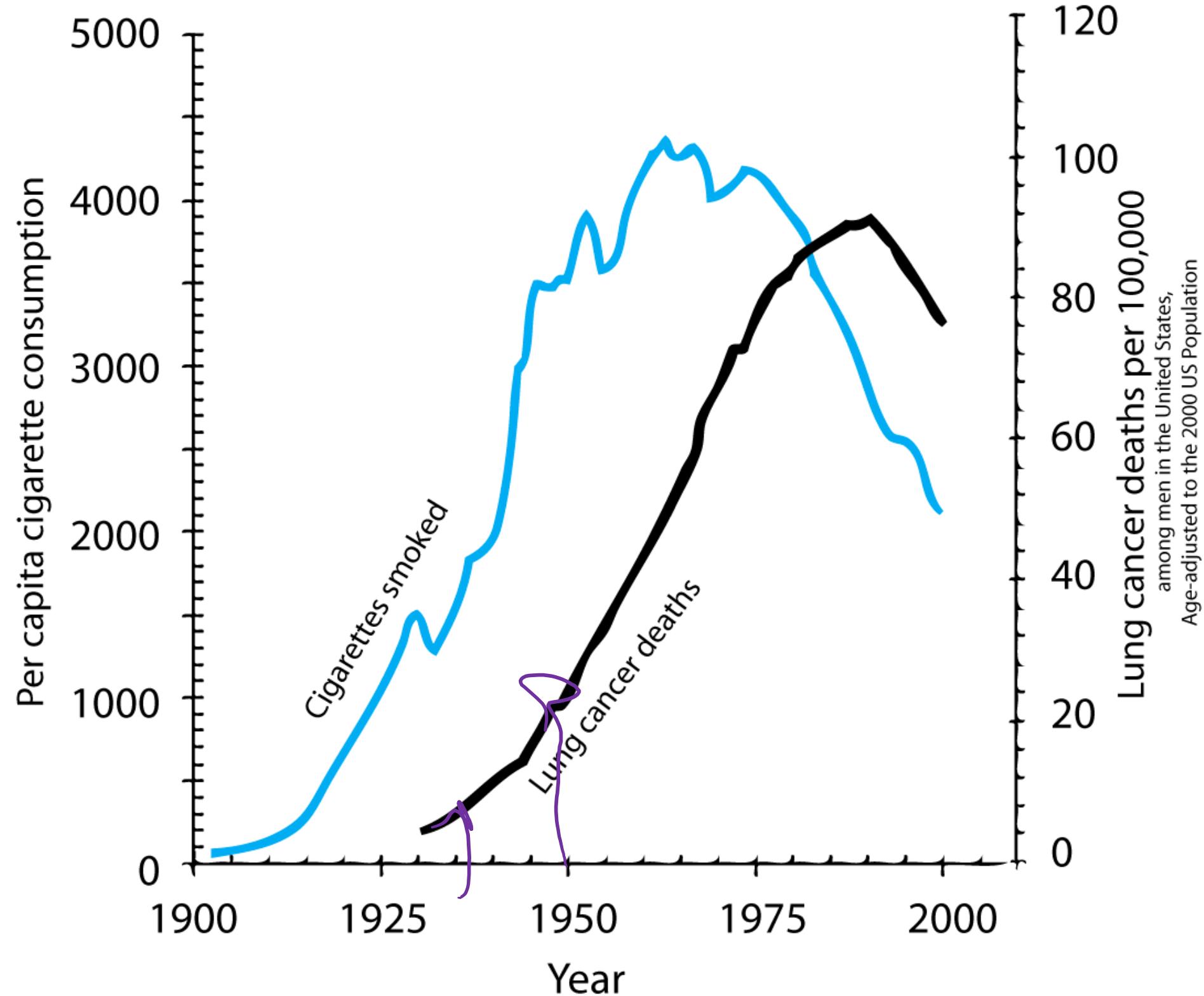
$Z \rightarrow A, B$

"luck"

US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation



<http://tylervigen.com/spurious-correlations>



Smoking and Cancer

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

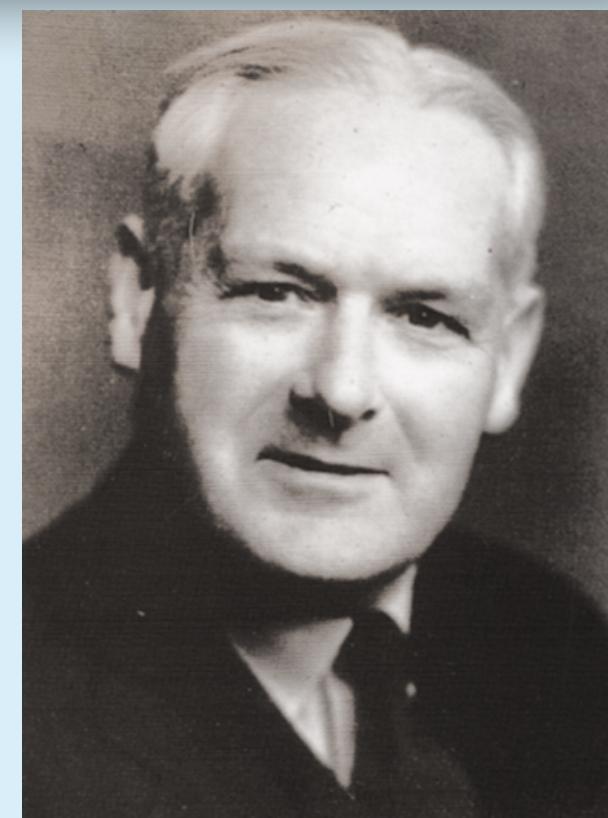
RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council



Sir Richard Doll (1912-2005)

Sir Austin Bradford Hill (1897-1991)

In England and Wales the phenomenal increase in the number of deaths attributed to cancer of the lung provides one of the most striking changes in the pattern of mortality recorded by the Registrar-General. For example, in the quarter of a century between 1922 and 1947 the annual number of deaths recorded increased from 612 to 9,287, or roughly fifteenfold. ... The rise seems to have been particularly rapid since the end of the first world war, between 1921-30 and 1940-4 the death rate of men at ages 45 and over increased sixfold and of women of the same ages approximately threefold. This increase is still continuing. It has occurred, too, in Switzerland, Denmark, the U.S.A., Canada, and Australia, and has been reported from Turkey and Japan.

The large and continued increase in the recorded deaths even within the last five years, both in the national figures and in those from teaching hospitals, also makes it hard to believe that improved diagnosis is entirely responsible. In short, there is sufficient reason to reject that factor as the whole explanation, although no one would deny that it may well have been contributory. As a corollary, it is right and proper to seek for other causes.

Possible Causes of the Increase

Two main causes have from time to time been put forward : (1) a general atmospheric pollution from the exhaust fumes of cars, from the surface dust of tarred roads, and from gas-works, industrial plants, and coal fires ; and (2) the smoking of tobacco. Some characteristics of the former have certainly become more prevalent in the last 50 years, and there is also no doubt that the smoking of cigarettes has greatly increased. Such associated changes in time can, however, be no more than suggestive, and until recently there has been singularly little more direct evidence. That evidence, based upon clinical experience and records, relates mainly to the use of tobacco. For instance,

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council
AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

Present Investigation

The present investigation was planned in 1947, to be carried out on a sufficiently large scale to determine whether patients with carcinoma of the lung differed materially from other persons in respect of their smoking habits or in some other way which might be related to the atmospheric pollution theory. Patients with carcinoma of the stomach, colon, or rectum were also incorporated in the inquiry, as one of the contrasting groups, and special attention was therefore given at the same time to factors which might bear upon the aetiology of these forms of malignant disease. A separate report will be made upon these inquiries. The present study is confined to the question of smoking in relation to carcinoma of the lung.

The method of the investigation was as follows : Twenty London hospitals were asked to co-operate by notifying all patients admitted to them with carcinoma of the lung, stomach, colon, or rectum. For the most part these hospitals were initially confined to one region of London (the north-west), to allow ease of travelling, but others were subsequently added to increase the scope of the inquiry. A list of those taking part is given at the end of the paper. The method of notification varied ; in some it was made by the admitting clerk on the basis of the admission diagnosis, in others by the house-physician when a reasonably confident clinical diagnosis had been made, and in yet others by the cancer registrar or the radiotherapy department. None of these methods is likely to have resulted in complete notification, but there is no reason to suppose that those who escaped notification were a selected group—that is, selected in such a way as to bias the inquiry—as the points of interest in the investigation were either not known or known only in broad outline by those responsible for notifying.

Study Design:

arrange for hospitals to contact investigators when a patient is admitted with lung cancer

On receipt of the notification an almoner, engaged wholly on research, visited the hospital to interview the patient, using a set questionary. During the inquiry four almoners were employed and all the patients were interviewed by one or other of them. As well, however, as interviewing the notified patients with cancer of one of the four specified sites, the almoners were required to make similar inquiries of a group of "non-cancer control" patients. These patients were not notified, but for each lung-carcinoma patient visited at a hospital the almoners were instructed to interview a patient of the same sex, within the same five-year age group, and in the same hospital at or about the same time. (Where more than one suitable patient was available the choice fell upon the first one in the ward lists considered by the ward sister to be fit for interview.)

Study Design:

arrange for hospitals to contact investigators when a patient is admitted with lung cancer

interview patient about smoking
also interview a non-cancer "control" patient

The Data

Between April, 1948, and October, 1949, the notifications of cancer cases numbered 2,370. It was not, however, possible to interview all these patients. To begin with, it had been decided beforehand that no one of 75 years of age or more should be included in the inquiry, since it was unlikely that reliable histories could be obtained from the very old. There were 150 such patients. In a further 80 cases the diagnosis was incorrect and had been changed before the almoner paid her visit. Deducting these two groups leaves 2,140 patients who should have been interviewed. Of these, 408 could not be interviewed for the following reasons : already discharged 189, too ill 116, dead 67, too deaf 24, unable to speak English clearly 11, while in one case the almoner abandoned the interview as the patient's replies appeared wholly unreliable. No patient refused to be interviewed.

TABLE I.—*Number of Patients Interviewed in Each Disease Group,
Subdivided According to Certainty of Diagnosis*

Disease Group	No. of Cases		
	Group A. Diagnosis Confirmed at Necropsy, etc.	Group B. Other Criteria of Diagnosis	Total
Carcinoma of lung	489	220	709
,, ,, stomach	178	28	206
,, ,, colon and rectum ..	412	19	431
Other malignant diseases	—	—	81
Diseases other than cancer (controls)	—	—	709
Other cases	—	—	335
Excluded	—	—	4
All cases	—	—	2,475

TABLE II.—Comparison Between Lung-carcinoma Patients and Non-cancer Patients Selected as Controls, With Regard to Sex, Age, Social Class, and Place of Residence

Age	No. of Lung-carcinoma Patients		No. of Non-cancer Control Patients		Social Class (Registrar-General's Categories. Men Only)	No. of Lung-carcinoma Patients	No. of Non-cancer Patients
	M	F	M	F			
25- ..	2	1	2	1	I and II ..	77	87
30- ..	6	0	6	0	III ..	388	396
35- ..	18	3	18	3	IV and V ..	184	166
40- ..	36	4	36	4	All classes ..	649	649
45- ..	87	10	87	10	<i>Place of residence</i>		
50- ..	130	11	130	11	County of London ..	330	377
55- ..	145	9	145	9	Outer London	203	231
60- ..	109	9	109	9	Other county borough ..	23	16
65- ..	88	9	89*	9	Urban district ..	95	54
70-74..	28	4	27*	4	Rural district ..	43	27
					Abroad or in Services ..	15	4
All ages	649	60	649	60	Total (M + F) ..	709	709

* One control patient was selected, in error, from the wrong age group.

“How much did you smoke before the onset of your present illness?”

TABLE III.—Amount of Tobacco Smoked Daily Before Present Illness as Recorded at Two Interviews With the Same Patients at an Interval of Six Months or More

First Interview No. of Persons Smoking	Second Interview. No. of Persons Smoking						Total
	0	1 cig.-	5 cigs.-	15 cigs.-	25 cigs.-	50 cigs. +	
0	8	1					9
1 cig.- ..		4	1				5
5 cigs.- ..		1	13				17
15 cigs - ..			4				14
25 cigs.- ..					3		4
50 cigs. + ..				9	1	0	1
Total ..	8	6	18	13	5	0	50

TABLE IV.—Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males:			
Lung-carcinoma patients (649)	2 (0·3%)	647	P (exact method) = 0·00000064
Control patients with diseases other than cancer (649) ..	27 (4·2%)	622	<u> </u>
Females:			
Lung-carcinoma patients (60)	19 (31·7%)	41	$\chi^2 = 5\cdot76$; n = 1 $0\cdot01 < P < 0\cdot02$
Control patients with diseases other than cancer (60) ..	32 (53·3%)	28	

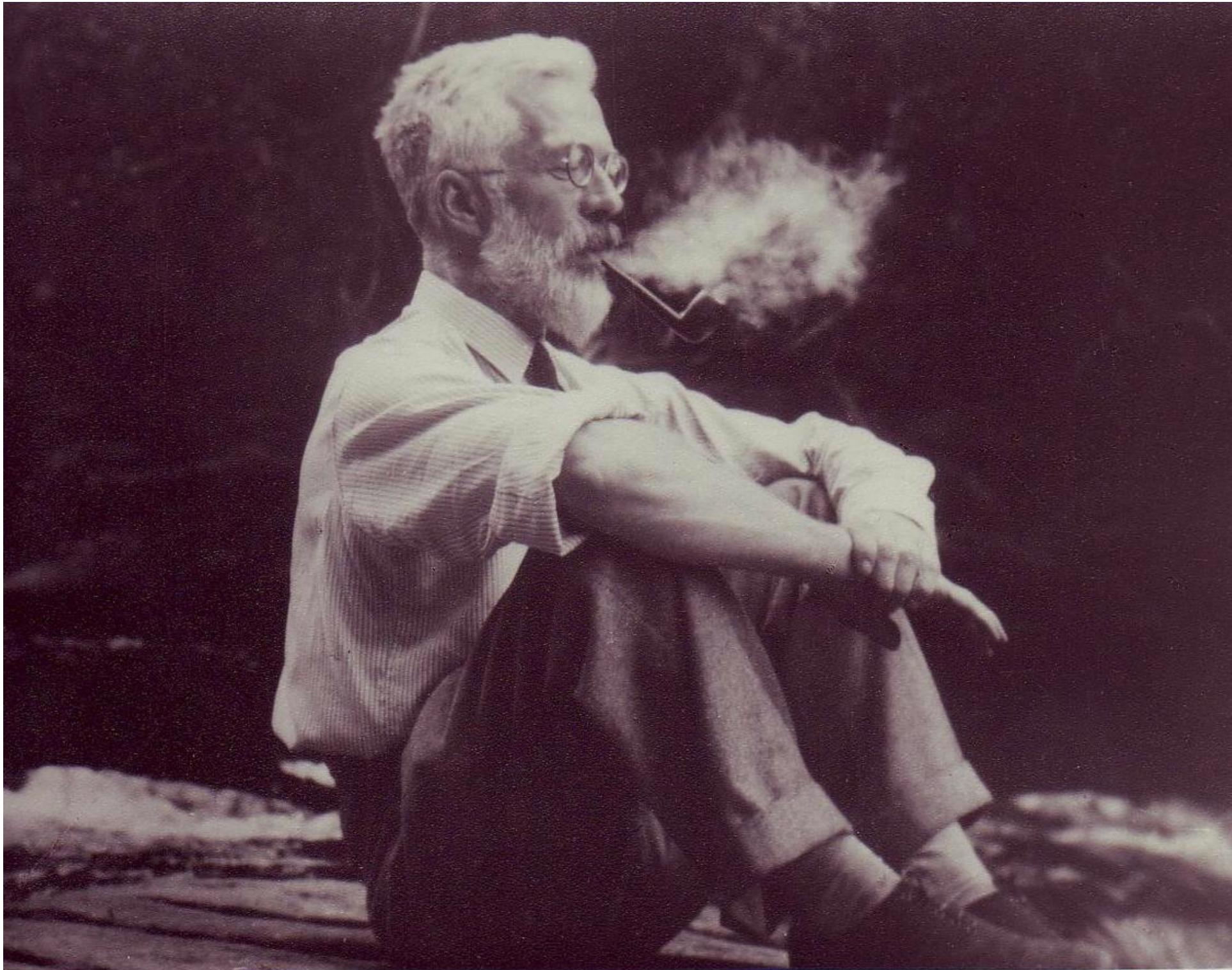
“I gave up smoking two-thirds of the way though the study.”
 – Richard Doll

Probability Test

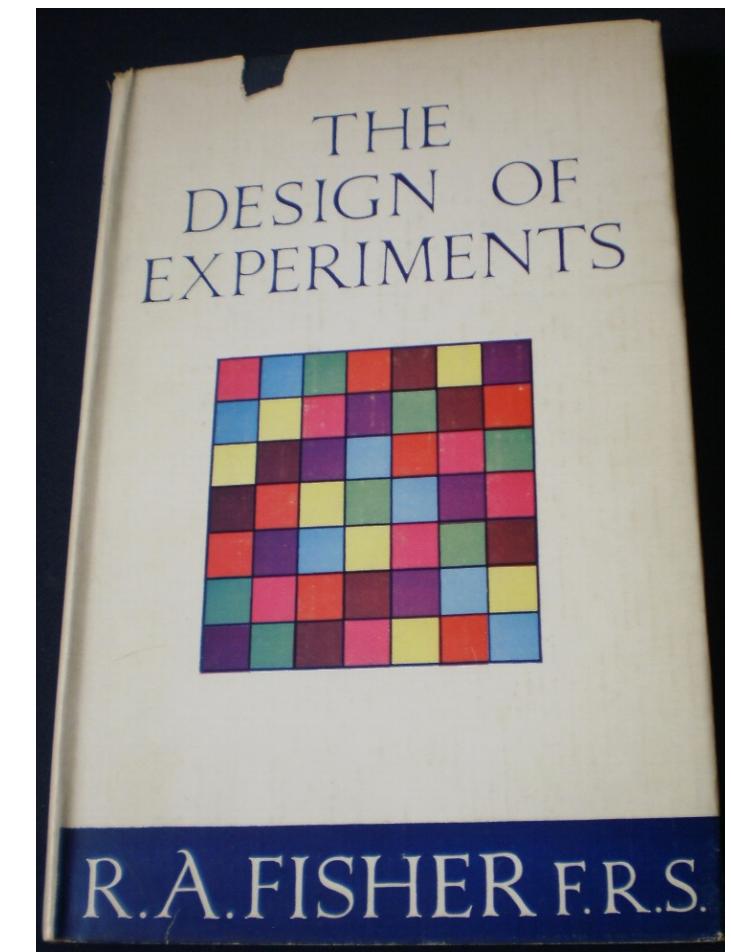
TABLE IV.—*Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer*

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males:			
Lung-carcinoma patients (649)	2 (0.3%)	647	P (exact method) = 0.00000064
Control patients with diseases other than cancer (649) ..	27 (4.2%)	622	
Females:			
Lung-carcinoma patients (60)	19 (31.7%)	41	$\chi^2 = 5.76$; n = 1 0.01 < P < 0.02
Control patients with diseases other than cancer (60) ..	32 (53.3%)	28	

Probability that if the *null hypothesis* were true, the measured correlation would be higher than observed.



Sir Ronald Fisher (1890-1962)



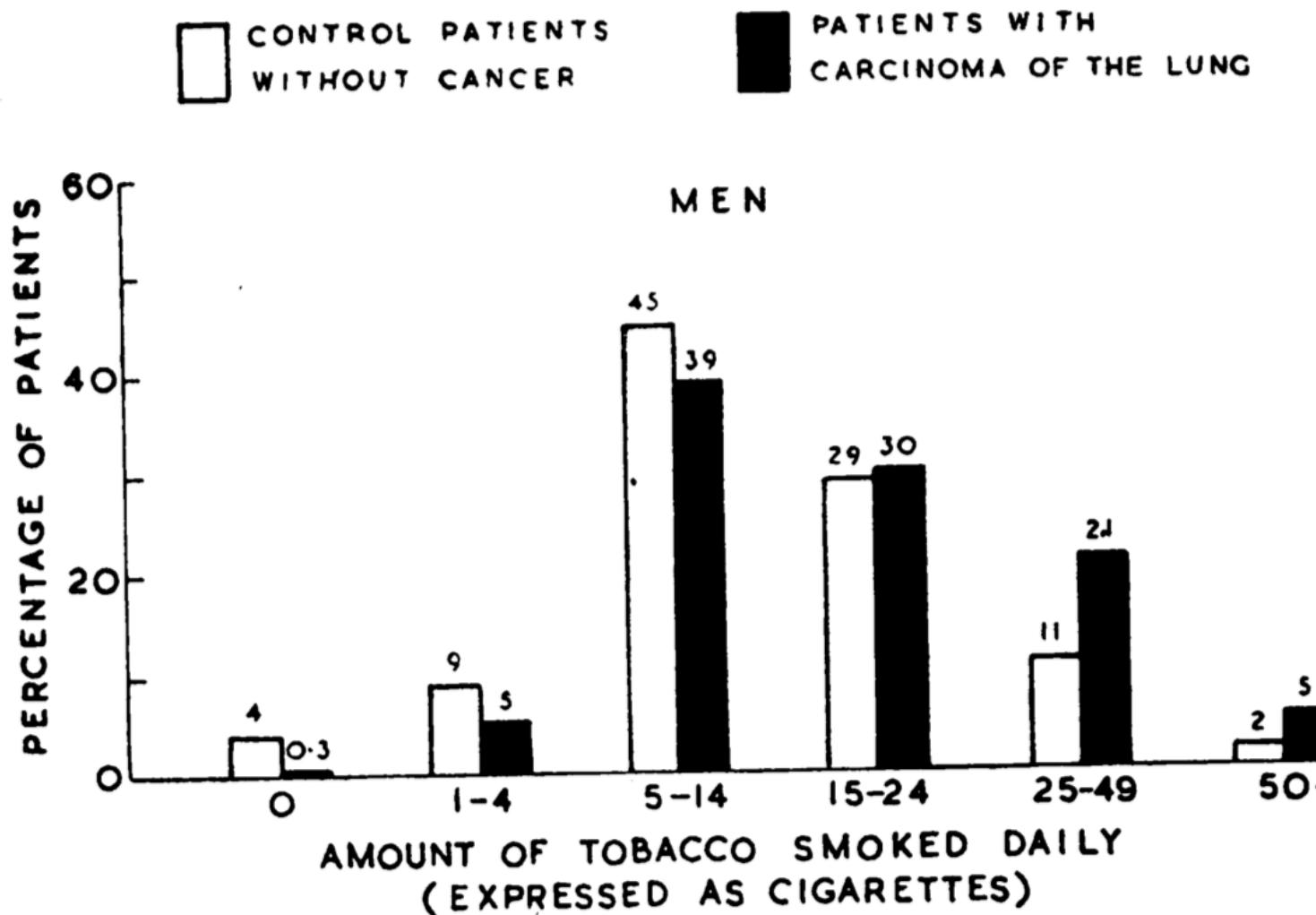
"...the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis."

More Evidence

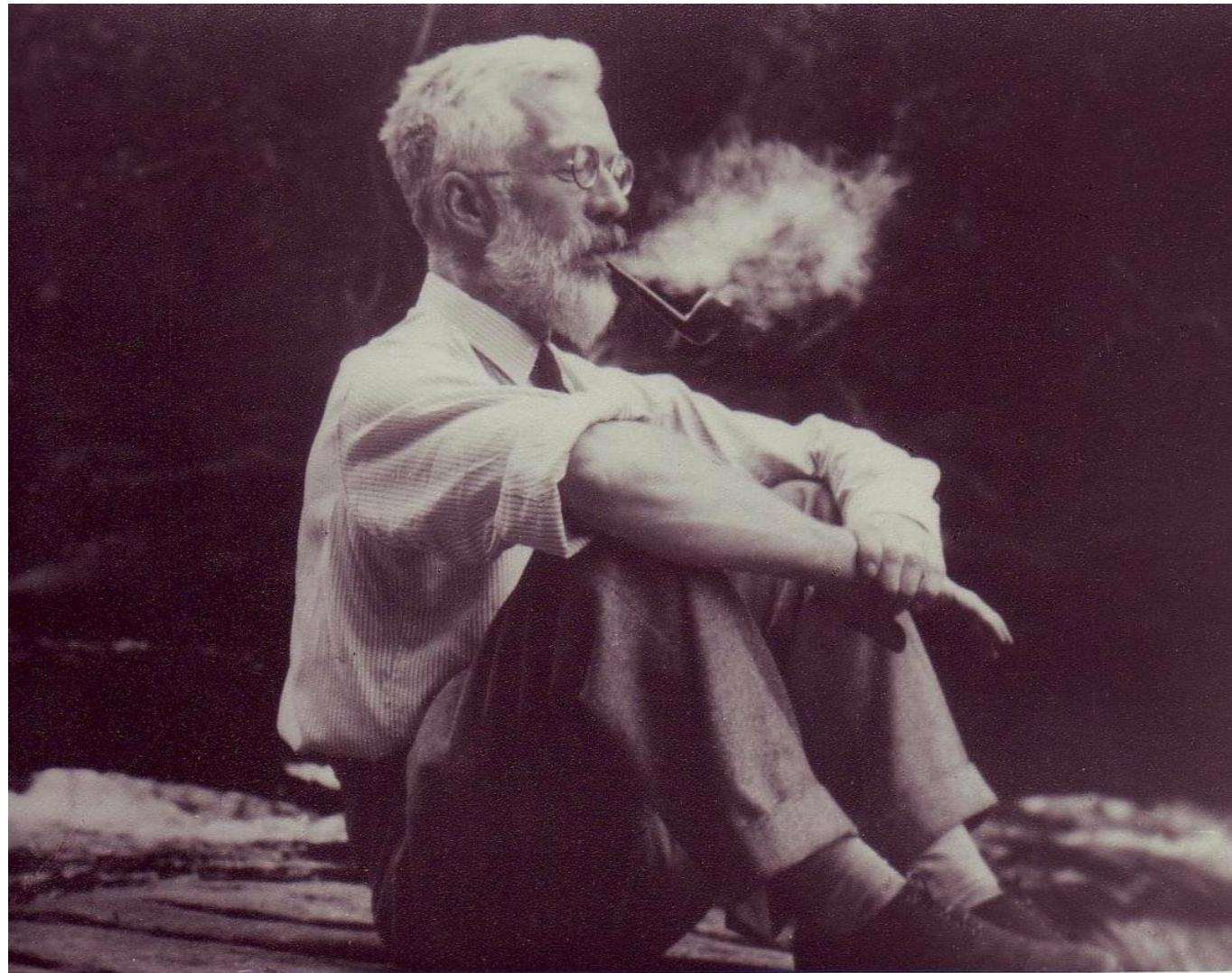
TABLE V.—*Most Recent Amount of Tobacco* Consumed Regularly by Smokers Before the Onset of Present Illness; Lung-carcinoma Patients and Control Patients with Diseases Other Than Cancer*

Disease Group	No. Smoking Daily					Probability Test
	1 Cig.-*	5 Cigs.-	15 Cigs.-	25 Cigs.-	50 Cigs. +	
Males: Lung-carcinoma patients (647)	33 (5.1%)	250 (38.6%)	196 (30.3%)	136 (21.0%)	32 (5.0%)	$\chi^2=36.95$; $n=4$; $P<0.001$
Control patients with diseases other than cancer (622) ..	55 (8.8%)	293 (47.1%)	190 (30.5%)	71 (11.4%)	13 (2.1%)	
Females: Lung-carcinoma patients (41) ..	7 (17.1%)	19 (46.3%)	9 (22.0%)	6 (14.6%)	0 (0.0%)	$\chi^2=5.72$; $n=2$; $0.05 < P < 0.10$ (Women smoking 15 or more cigarettes a day grouped together)
Control patients with diseases other than cancer (28) ..	12 (42.9%)	10 (35.7%)	6 (21.4%)	0 (0.0%)	0 (0.0%)	

* Ounces of tobacco have been expressed as being equivalent to so many cigarettes. There is 1 oz. of tobacco in 26.5 normal-size cigarettes, so that the conversion factor has been taken as: 1 oz. of tobacco a week = 4 cigarettes a day.



Fisher's Response



Fisher's Response

There is nothing to stop those who greatly desire it from believing that lung cancer is caused by smoking cigarettes. They should also believe that inhaling cigarette smoke is a protection. To believe this is, however, to run the risk of failing to recognize, and therefore failing to prevent, other and more genuine causes.

smoking → cancer

cancer → smoking

Z → cancer, smoking

But in the case where randomization has not been possible, these other possibilities lie wide open and should be excluded, or at least every effort should be made to exclude them, before we can assert that causation has been established. When I spoke to Bradford Hill in the early days of this affair, he was entirely unwilling to claim that causation had been proved. He said he didn't see what else it could be, but he was certainly unwilling to make the claim which is being made vociferously during the last year or two by committees reporting to the Medical Research Council in England, and to the American Cancer Society. Now, randomization is totally impossible, so far as I can judge, in an inquiry of this kind. It is not the fault of the medical investigators. It is not the fault of Hill or Doll or Hammond that they cannot produce evidence in which a thousand children of teen age have been laid under a ban that they shall never smoke, and a thousand more chosen at random from the same age group have been under compulsion to smoke at least thirty cigarettes a day. If that type of experiment could be done, there would be no difficulty.

CIGARETTES, CANCER, AND STATISTICS

Sir Ronald Fisher

thing quite unexplored, something quite hypothetical, in the tobacco smoke which would be capable of producing lung cancer. It is also known that the tobacco used as pipe tobacco and for cigars is more thoroughly fermented before use than is that used in cigarettes, or at least in the predominant source of cigarette tobacco, in Virginia. I think those who prepare the tobacco produced in Virginia are rather acutely aware, that the price per pound is high, there is loss of weight in fermentation, and it is as well not to lose 10 per cent more weight than is necessary. And so, on the whole, the Virginia tobacco is rather lightly fermented. You could imagine—you could claim even—as a special pleading, that it was the unfermented condition of the Virginia tobacco, largely used in cigarettes, that was responsible for the supposedly noxious fumes which the burning of such tobacco produces. Discussion is full of such things.

smoking → cancer

Virginia → cancer
pipes ↛ cancer

cancer → smoking

Is it possible, then, that lung cancer—that is to say, the pre-cancerous condition which must exist and is known to exist for years in those who are going to show overt lung cancer—is one of the causes of smoking cigarettes? I don't think it can be excluded. I don't think we know enough to say that it is such a cause. But the pre-cancerous condition is one involving a certain amount of slight chronic inflammation. The causes of smoking cigarettes may be studied among your friends, to some extent, and I think you will agree that a slight cause of irritation—a slight disappointment, an unexpected delay, some sort of a mild rebuff, a frustration—are commonly accompanied by pulling out a cigarette and getting a little compensation for life's minor ills in that way. And so, anyone suffering from a chronic inflammation in part of the body (something that does not give rise to conscious pain) is not unlikely to be associated with smoking more frequently, or smoking rather than not smoking. It is the kind of comfort that might be a real solace to anyone in the fifteen years of

"...*ppier person a little more unhappy than he need be.*

For my part, I think it is more likely that a common cause supplies the explanation. Again, we do not know. I do not put forth any explanation as proved, but as requiring investigation. The obvious common cause to think of is the genotype. We are all different genotypes. I suppose in this nation there must be well over 150 million different genotypes. If one studies cancer in mice (and I suppose about half the mice of the world are kept to study cancer with), if one examines any of the many (and there are thousands) of inbred lines of mice (where we can get a hundred or two hundred individuals of the same genotype to study)—if you take, then, any two such lines of differing genotypes, they will, I believe, invariably be found to differ in the frequency, in the age incidence, and in the type of cancer which those mice suffer from. Consequently if there is any genotypic difference between the different smoking classes, we may expect differences in the type or frequency of cancer that they display.

genotype → cancer, smoking

smoke anything but a pipe. Why? Because they are made that way. They are the sort of men who take to the pipe and don't take to cigarettes, just as there are other men who would never take to a pipe but constantly feel the need of cigarettes. It is not, then, a very long shot to guess that there is a genetic component which distinguishes the different smoking classes. And that is the second piece of research which I think is extremely urgent.

Ronald A. Fisher
Galton Professor of Eugenics
at University College London

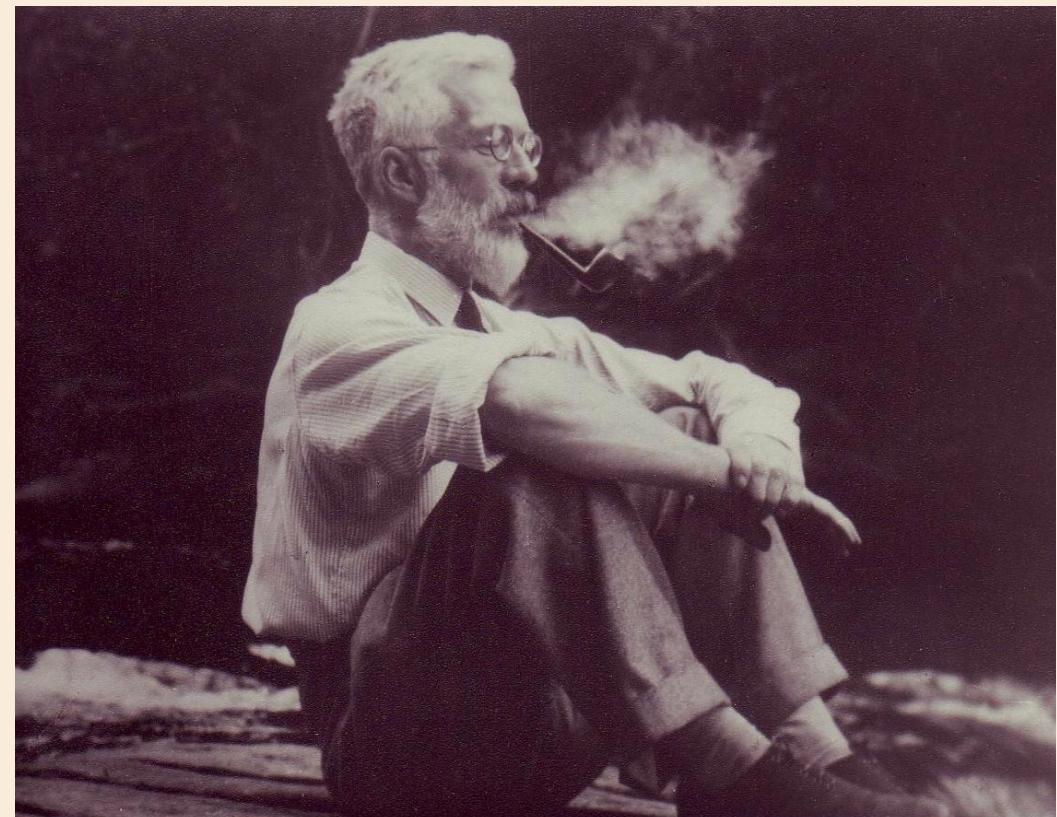
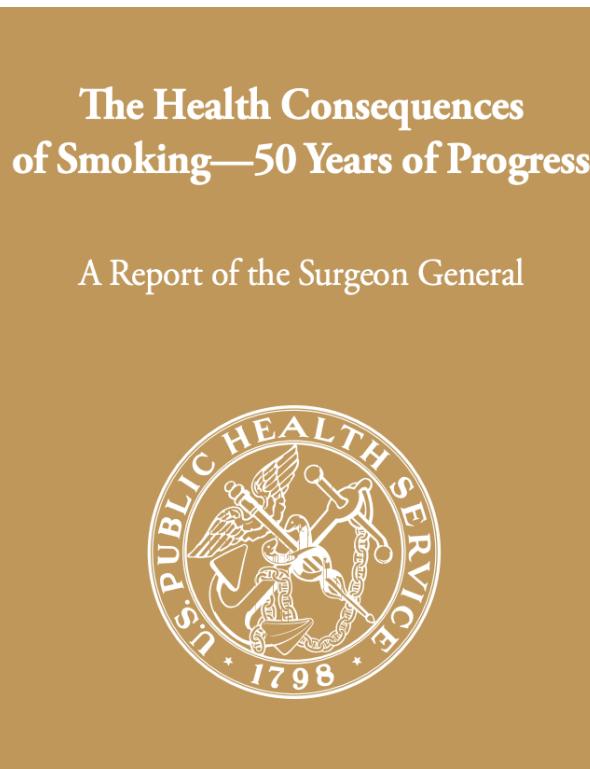
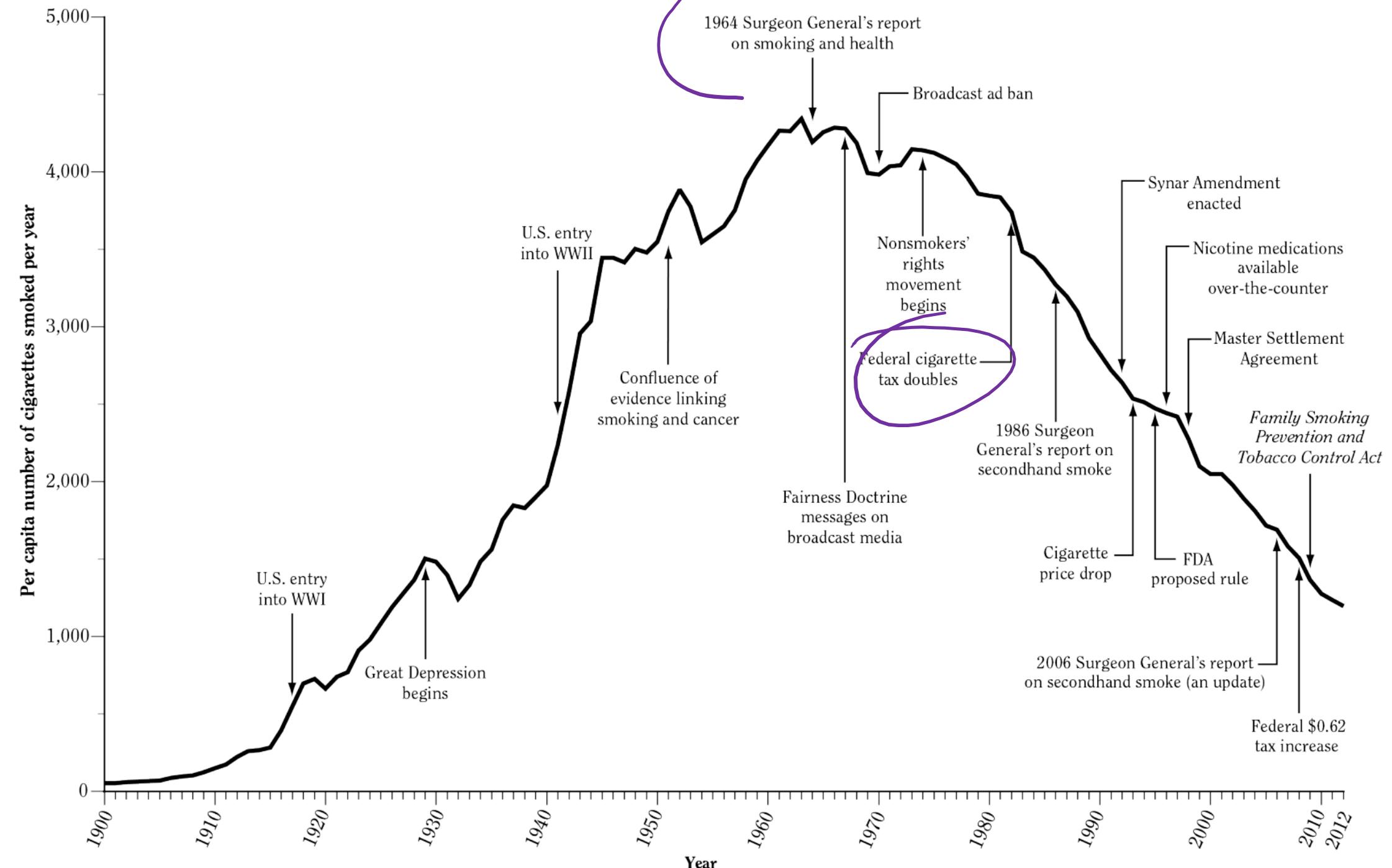


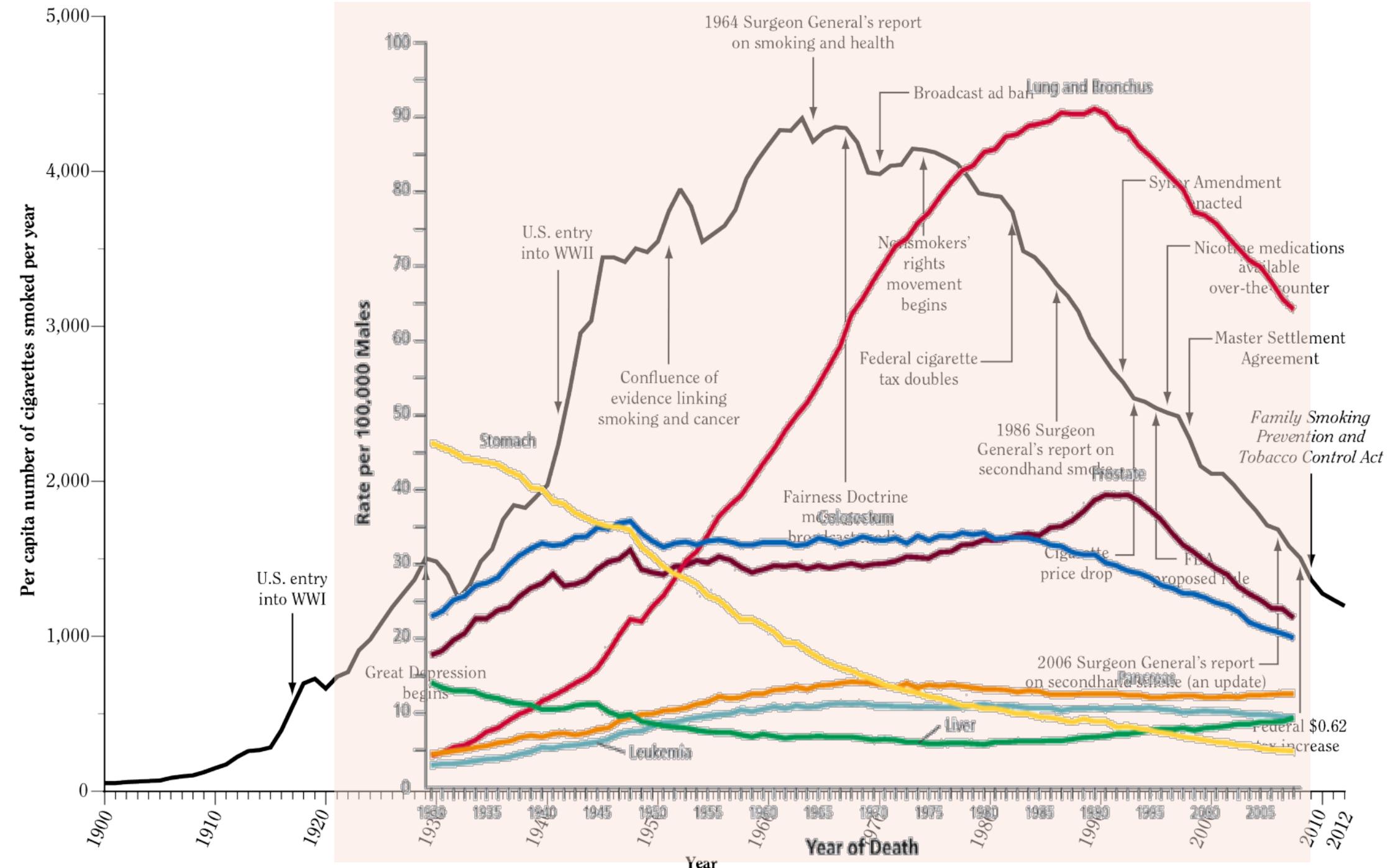
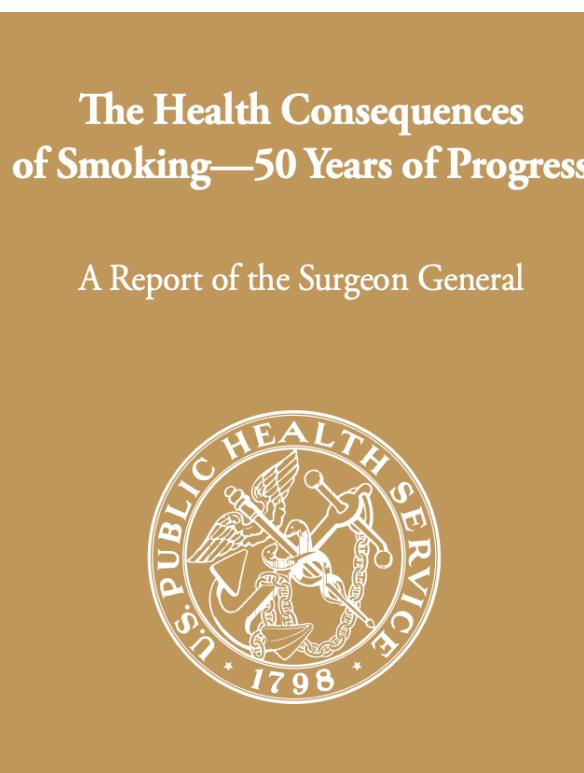
Figure 2.1 Adult* per capita cigarette consumption and major smoking and health events, United States, 1900–2012

2014



Sources: Adapted from Warner 1985 with permission from Massachusetts Medical Society, ©1985; U.S. Department of Health and Human Services 1989; Creek et al. 1994; U.S. Department of Agriculture 2000; U.S. Census Bureau 2013; U.S. Department of the Treasury 2013.

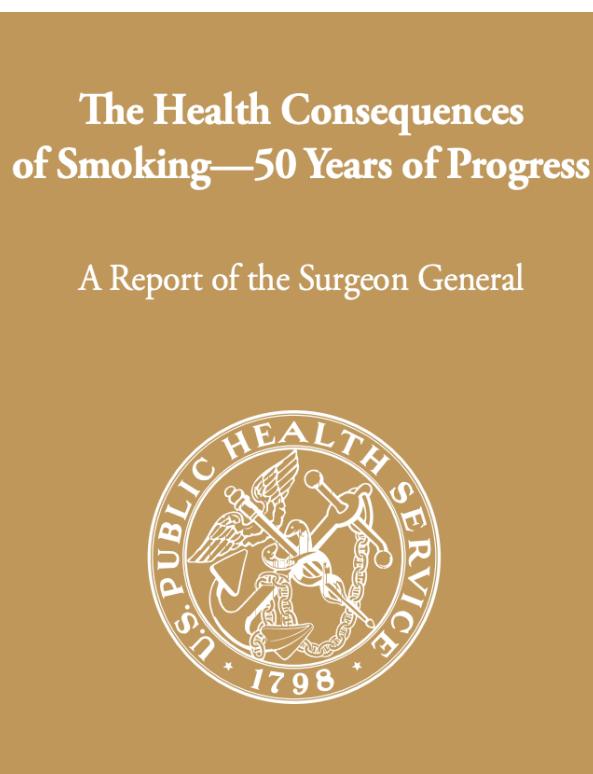
*Adults ≥18 years of age as reported annually by the Census Bureau.

Figure 2.1 Adult* per capita cigarette consumption and major smoking and health events, United States, 1900–2012

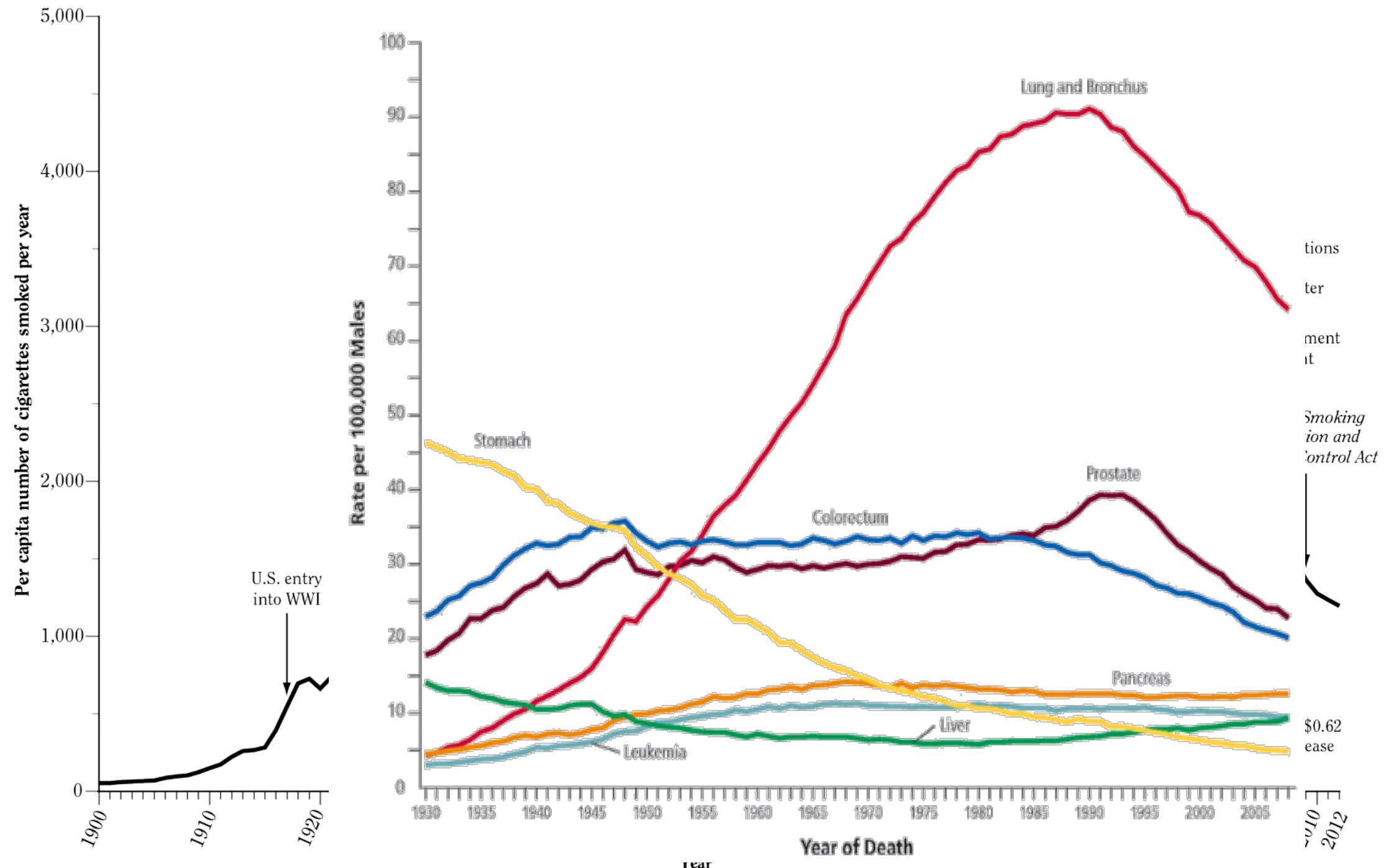
Sources: Adapted from Warner 1985 with permission from Massachusetts Medical Society, ©1985; U.S. Department of Health and Human Services 1989; Creek et al. 1994; U.S. Department of Agriculture 2000; U.S. Census Bureau 2013; U.S. Department of the Treasury 2013.

*Adults ≥18 years of age as reported annually by the Census Bureau.

Figure 2.1 Adult* per capita cigarette consumption and major smoking and health events, United States, 1900–2012



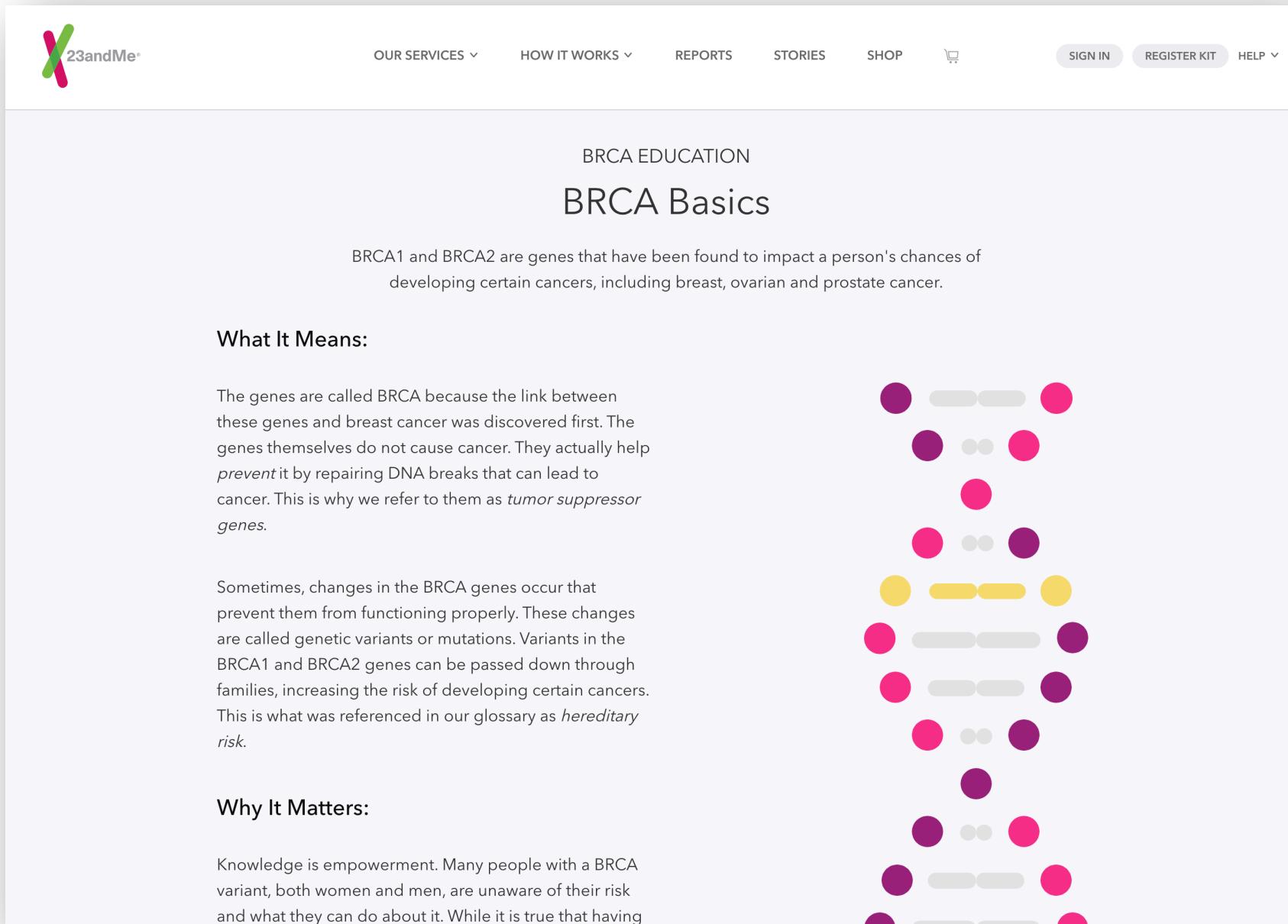
2014



Sources: Adapted from Warner 1985 with permission from Massachusetts Medical Society, ©1985; U.S. Department of Health and Human Services 1989; Creek et al. 1994; U.S. Department of Agriculture 2000; U.S. Census Bureau 2013; U.S. Department of the Treasury 2013.

*Adults ≥18 years of age as reported annually by the Census Bureau.

What causes breast cancer?



The image shows a screenshot of the 23andMe website. At the top, there is a navigation bar with links for "OUR SERVICES", "HOW IT WORKS", "REPORTS", "STORIES", "SHOP", "SIGN IN", "REGISTER KIT", and "HELP". Below the navigation bar, the page title "BRCA EDUCATION" and the specific article title "BRCA Basics" are displayed. The main content of the article discusses the BRCA1 and BRCA2 genes and their role in cancer risk. It includes sections titled "What It Means:" and "Why It Matters:". The "What It Means:" section explains that the genes are called BRCA because the link between them and breast cancer was discovered first, and they are tumor suppressor genes that help prevent cancer by repairing DNA breaks. It also mentions genetic variants or mutations that can lead to hereditary risk. The "Why It Matters:" section emphasizes the importance of knowledge and empowerment, noting that many people with a BRCA variant are unaware of their risk and what they can do about it. The page features a decorative graphic of colored dots (purple, pink, yellow) arranged in a grid pattern.

BRCA EDUCATION

BRCA Basics

BRCA1 and BRCA2 are genes that have been found to impact a person's chances of developing certain cancers, including breast, ovarian and prostate cancer.

What It Means:

The genes are called BRCA because the link between these genes and breast cancer was discovered first. The genes themselves do not cause cancer. They actually help *prevent* it by repairing DNA breaks that can lead to cancer. This is why we refer to them as *tumor suppressor genes*.

Sometimes, changes in the BRCA genes occur that prevent them from functioning properly. These changes are called genetic variants or mutations. Variants in the BRCA1 and BRCA2 genes can be passed down through families, increasing the risk of developing certain cancers. This is what was referenced in our glossary as *hereditary risk*.

Why It Matters:

Knowledge is empowerment. Many people with a BRCA variant, both women and men, are unaware of their risk and what they can do about it. While it is true that having

<http://www.cbcrp.org/causes/index.php>

A Model of Breast Cancer Causation

[Definitions](#)
[References](#)

Visualizing the many factors and relationships influencing breast cancer incidence in postmenopausal women

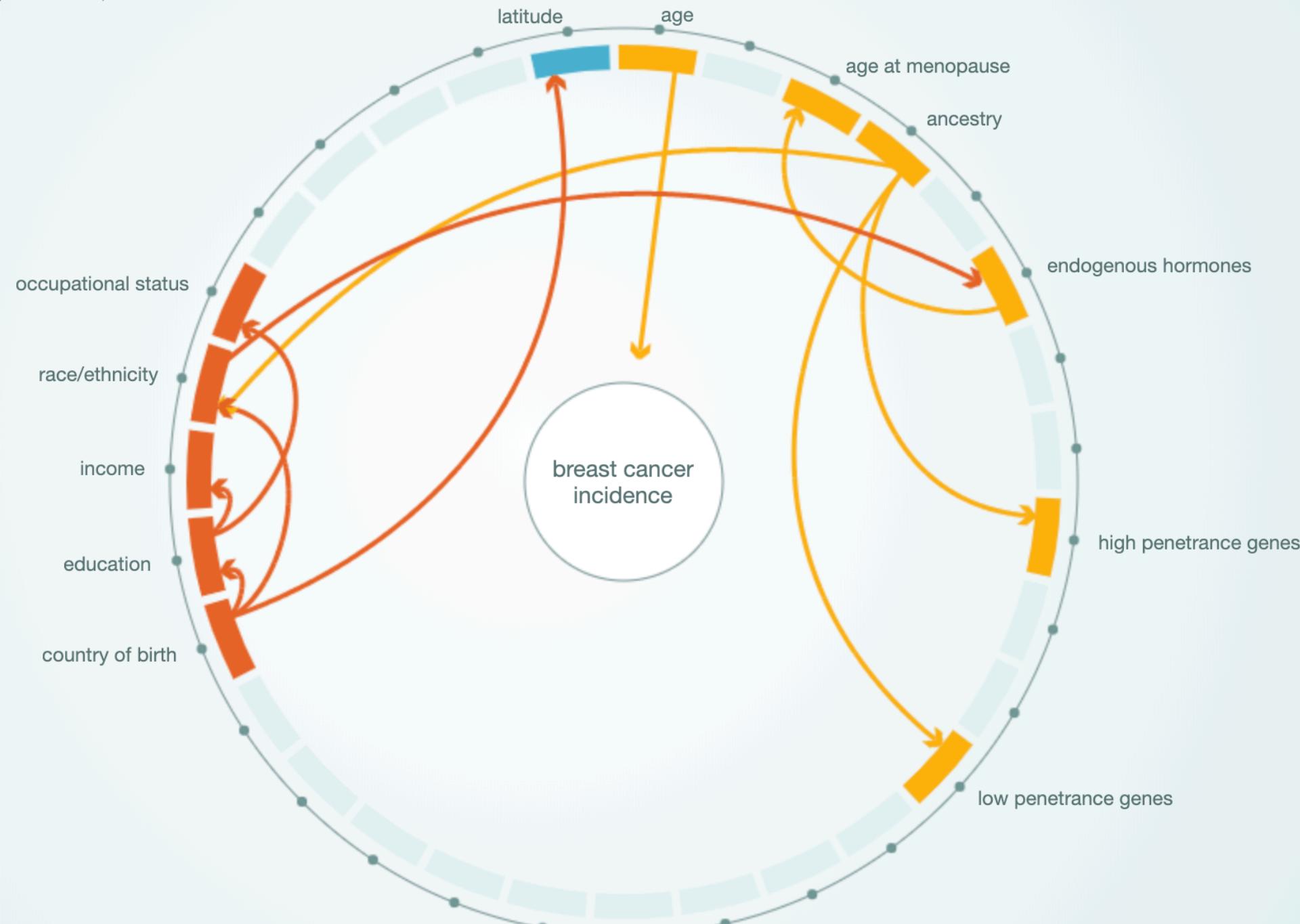
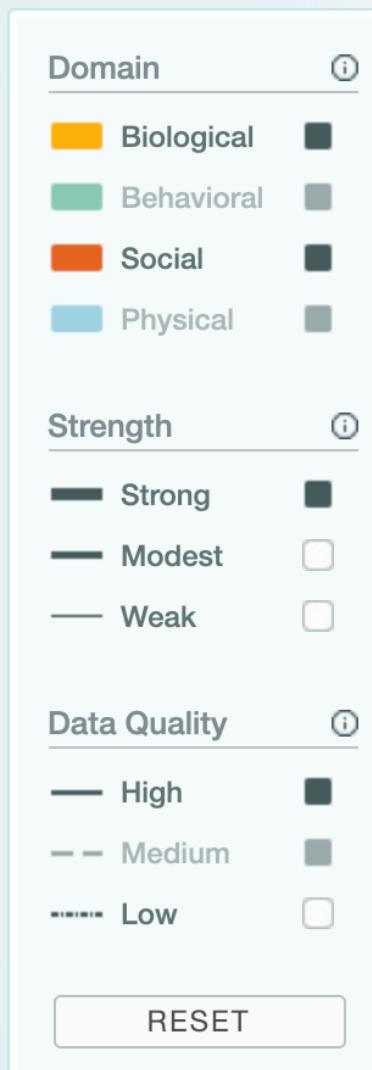


A Model of Breast Cancer Causation

Visualizing the many factors and relationships influencing breast cancer incidence in postmenopausal women

Definitions

References

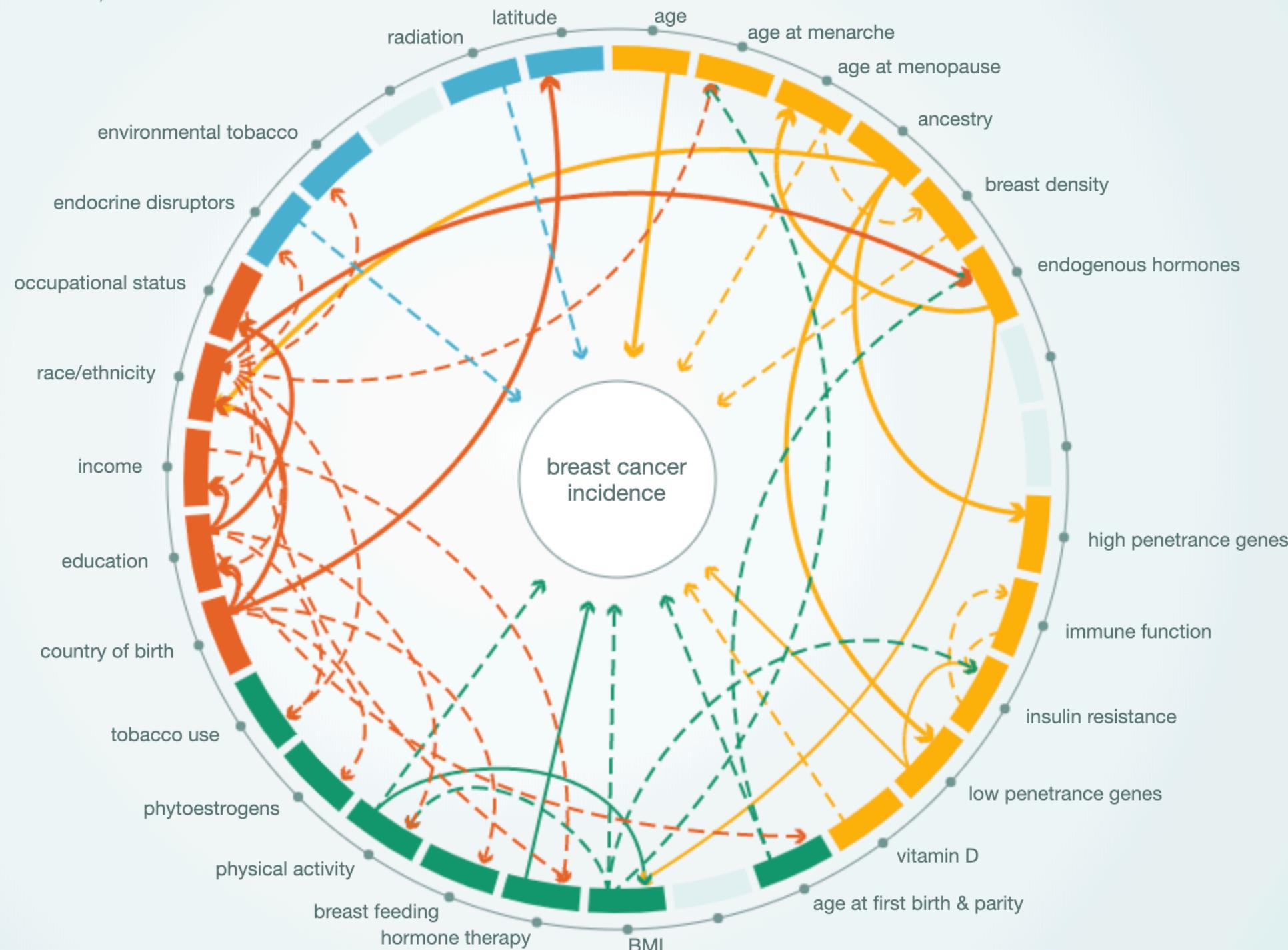
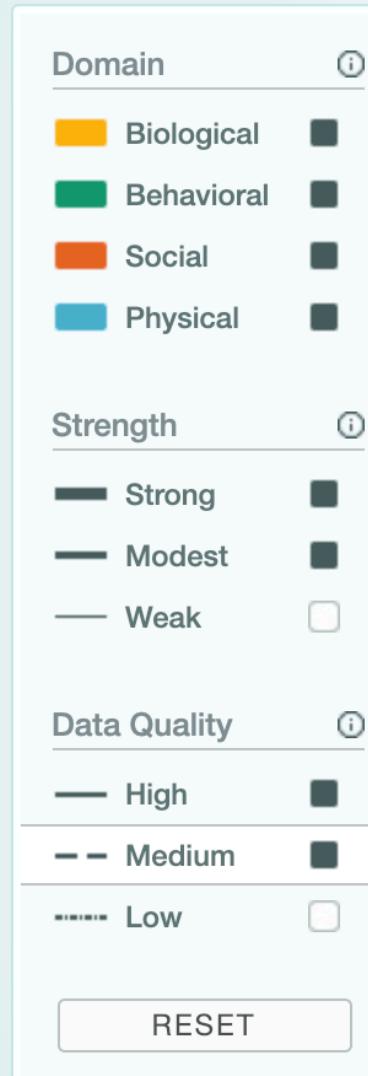


A Model of Breast Cancer Causation

Visualizing the many factors and relationships influencing breast cancer incidence in postmenopausal women

Definitions

References



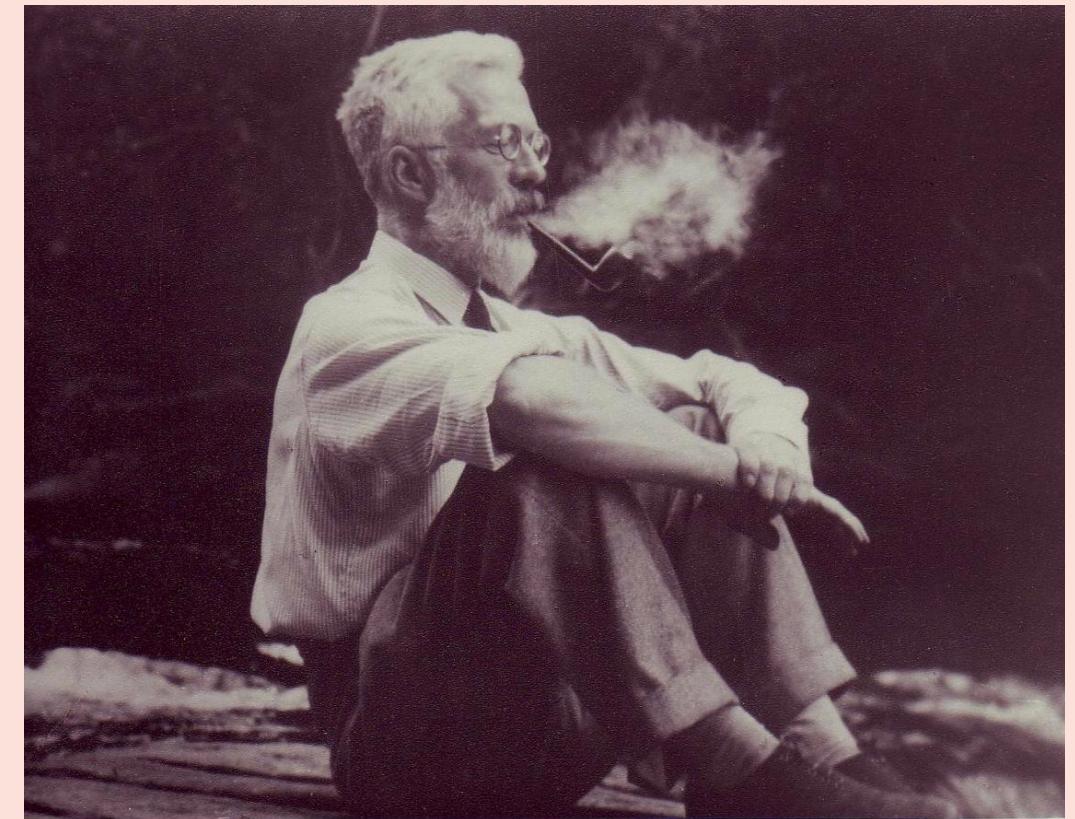
Lessons Learned?

Lessons Learned?

“Statistics has gained a place of modest usefulness in medical research. It can derive and retain this only by complete impartiality, which is not unattainable by rational minds. We should not be content to be “not so unfair”, for without fairness the statistician is in danger of scientific errors through his moral fault. ...”

Lessons Learned?

“Statistics has gained a place of modest usefulness in medical research. It can derive and retain this only by complete impartiality, which is not unattainable by rational minds. We should not be content to be “not so unfair”, for without fairness the statistician is in danger of scientific errors through his moral fault. ...”



Ronald A. Fisher
Alleged Dangers of Cigarette-Smoking,
British Medical Journal 1957

Hill's Lessons

The Environment and Disease: Association or Causation?

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(*Professor Emeritus of Medical Statistics,
University of London*)

Section of Occupational Medicine

Meeting January 14 1965

President's Address

Here then are nine different viewpoints from all of which we should study association before we cry causation. What I do not believe – and this has been suggested – is that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?

Hill's Lessons

1. Strength of association
2. Consistency – repeated observation
3. Specificity
4. Temporality
5. Gradient: more smoking → more cancer
6. Plausibility (not required)
7. Coherence (not conflict with known facts)
8. Experiment
9. Analogy

Here then are nine different viewpoints from all of which we should study association before we cry causation. What I do not believe – and this has been suggested – is that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?

The Case for Action

Finally, in passing from association to causation I believe in ‘real life’ we shall have to consider what flows from that decision. On scientific grounds we should do no such thing. The evidence is there to be judged on its merits and the judgment (in that sense) should be utterly independent of what hangs upon it – or who hangs because of it. But in another and more practical sense we may surely ask what is involved in our decision. In occupational medicine our object is usually to take action. If this be operative cause and that be deleterious effect, then we shall wish to intervene to abolish or reduce death or disease.

But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like. In asking for very strong evidence I would, however, repeat emphatically that this does not imply crossing every ‘t’, and swords with every critic, before we act.

All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8.30 next day.

Charge

Next class:
Statistical Learning Theory

**Project 1: due 9:29am,
Tuesday (Jan 22)**



Australian Cigarette Packaging