MACROEPIDEMIOLOGY:

1. PUBLIC HEALTH RECORDS
2. POPULATION GENETICS AND FAMILIAL RISK
3. ENVIRONMENTAL EPIDEMIOLOGY
4. HUMAN PHYSIOLOGY AND GENETICS

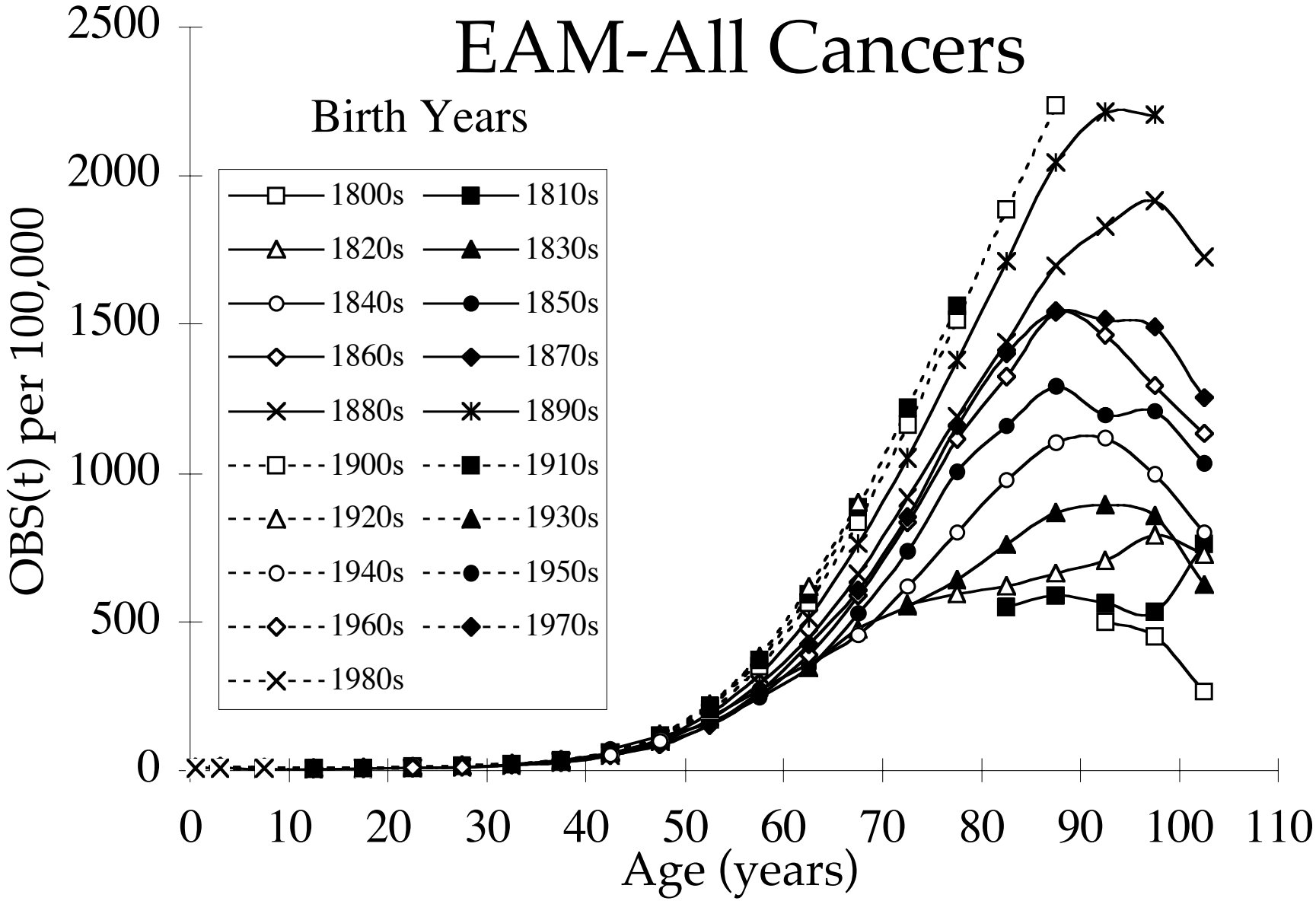
"*Eliminate the impossible, and whatever remains, however improbable, must be the truth.*"

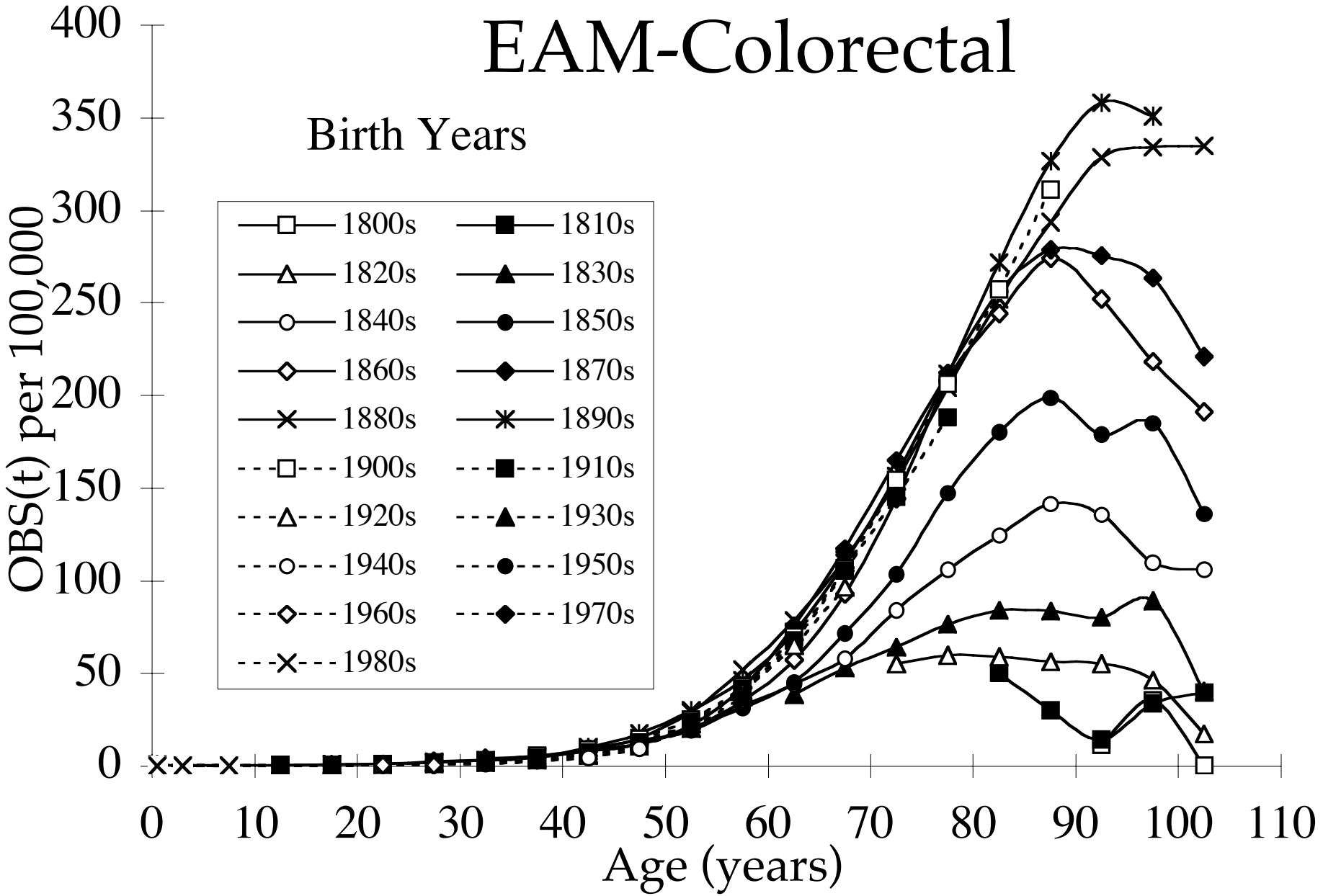
A.C. Doyle, M.D.

1. PUBLIC HEALTH RECORDS

http://epidemiology.mit.edu

http://epidemiology.mit.edu




EAM-Colorectal
Birth Years
1800s
1810s
1820s
1830s
1840s
1850s
1860s
1870s
1880s
1890s
1900s
1910s
1920s
1930s
1940s
1950s
1960s
1970s
1980s
OBS(t) per 100,000
0
50
100
150
200
250
300
350
400
0
10
20
30
40
50
60
70
80
90
100
110
Age (years)

We want to be able to relate the physiological processes and events required for a late onset disease in quantitative terms to the actual experience of the human populations studied.

But we have to consider random errors and bias the data before we use the data.

i. Uncertainty in mortality records?
ii. Uncertainty in census data?
iii. Uncertainty in the derived variable OBS(h,t)?

$OBS(h,t) = DEAD_{OBS}(h,t) / POP(h,t) =$

RECORDED number of deaths from OBServed cause for cohort's birth year(s) h at age interval t.

RECORDED number of persons alive in cohort of
birth year (s) h and age interval t.

Errors= biases + random errors

Inspect data set for breast cancer : EAF, NEAF,JF.

VARIANCE OF DERIVED VARIABLES

X(a,b,c...) where a,b,c... are independent variables.

$$\text{VARIANCE}(X) = V(X) = V(a)\left(\partial X/\partial a\right)^2 + V(b)\left(\partial X/\partial b\right)^2 + \ldots.$$

$$X = a/b$$

$$V(X) = V((a)\,(1/b)^2 + V(b)\,(a/b^2)^2$$

Using the Poisson distribution, variance = mean and V(a)=a, V(b) =b) etc.

$$V(X) = a/b^2 + a^2/b^3$$

Letting OBS(h,t) = X= a/b
We see that the V(x) = a/b^2 where a is the numb
of recorded deaths and b is the number of
recorded persons
alive for a particular cohort, h and t.

Example OBS(breast cancer, EAF,1890, 70-74)

Death data recorded in 1962, a = 2497

Population recorded in 1962, b= 2,742,481

OBS(h,t) = 2497/2,742,481 = 91.04×10^{-5}

V(OBS) = $2497/(2,742,481)^2$ =

Standard Deviation = $V^{1/2} = 1.82 \times 10^{-5}$

EXAMPLE: OBS(breast cancer, EAF,1890, 100-104

Death data recorded in 1992, a = 80

Population recorded in 192, b= 28,218

OBS(h,t) = = 283.5×10^{-5}

V(OBS) = $80/(28,218)^2 = 1.0 \times 10^{-7}$

Standard Deviation = $V^{1/2} = 31.7 \times 10^{-5}$

OBS(h,t) = 283 +/- 63 x 10^{-5} (=/- 2 St.Dev. ~ 95%

N.B. In most of our summaries we use multiple
such as a decade to increase the values of a an

OBS (h,t) =

RECORDED number of deaths from OBServed cause for cohort's birth year(s) h at age int

---

RECORDED number of persons alive in cohort birth year (s) h and age interval t.

Errors= biases + random errors

We want to define a function that describes the probability
that a person born at h will get the disease at age t given that they are alive at the beginning of the interval t.
This we might designate as the INCidence(h,t), INC(h,t).

OBS(h,t) the observed rate of death doesn't quite do that.

A person might:
- die before reaching age interval t ----> cannot die in t (!)
- survive the disease ---->SUR(h,t)
- not be recorded as dying with the disease----> REC(h,t)

BIASES TO ESTIMATES OF DISEASE RATE

SURVIVAL = S(h,t) =

$$= \frac{\text{(recorded surviving cases at t + 5)}}{\text{(recorded diagnoses at t) x (survival of all forms of death at t+5)}}$$

Go to Herrero-Jimenez et al. (2000)

S(h,t) for EAF colon cancer: year of diagnosis and year of birth.

Graph removed for copyright reasons. See Herrero-Jimenez, P. et al., "Population risk and physiological rate parameters for colon cancer. The union of an explicit model for carcinogenesis with the public health records of the United States." Mutat Res. 2000 Jan 17;447(1):73-116.

Table removed for copyright reasons. See Herrero-Jimenez, P. et al., "Population risk and physiological rate parameters for colon cancer. The union of an explicit model for carcinogenesis with the public health records of the United States." Mutat Res. 2000 Jan 17;447(1):73-116.

BIASES TO ESTIMATES OF DISEASE RATE

REPORTING = R(h,t) =

$$= \frac{\text{(recorded deaths from h at t.)}}{\text{(all recorded deaths from h at t)}}$$

From Herrero-Jimenez et al. (2000)

R(h,t) for EAM

Graph removed for copyright reasons. See Herrero-Jimenez, P. et al., "Population risk and physiological rate parameters for colon cancer. The union of an explicit model for carcinogenesis with the public health records of the United States." Mutat Res. 2000 Jan 17;447(1):73-116.

$$TOT(h,t) =$$

RECORDED number of deaths from all (TOTal
causes for cohort's birth year(s) h at age i

RECORDED number of persons alive in coho
birth year (s) h and age interval t.

"Two-(Rate-Limiting)-Stage" Model
Armitage & Doll, 1957

INITIATION **PROMOTION**

NORMAL -"**n**" events-> PRENEOPLASIA- "**m**" events->NEOPLASIA CELLS

Models before Herrero-Jimenez et al. (1998, 2000) treated OBS(h,t) as if it were INC(h,t) and assumed that all persons in the population were at equal risk of cancer.

Thus mortality rates were treated as the probability of experiencing the disease within the interval t, $P_{OBS}(h,t)$,
for any individual in the surviving population.

DIRECT "MEASUREMENT OF INCIDENCE (h,t)

http://www.seer.cancer.gov

1. You must register for use individually.

2. It has been difficult to access.

3. SEER data are based on several cities with specific reporting hospitals.

4. There are several examples where incidences are reported to rise rapidly without any rise in mortality rates.

We're from the government. We're here to help you.

SUMMARY

$$OBS(h,t) = DEAD_{OBS}(h,t) / POP(h,t)$$

Random errors: inspect any data set you use and
discover for yourself what kind of dispersion may be
expected using a Poisson approximation.

Note that estimates if bias, e.g.REC(h,t), SUR(h,t) and TOT(h,t) have random errors and biases, too.

<u>The variance of derived variables is an important logical tool:</u>

Where X(a,b,c...) where a,b,c... are independent variables,

$$VARIANCE(u) = V(u) = V(a)\,(\partial u/\partial a)^2 +$$

Cancer Data--->Carcinogenesis Model

$$\mathbf{INC\ (h,t) \approx OBS(h,t)/[(REC(h,t)\ 1\text{-}SUR(h,t))\ (1\text{-}TOT(h,t))] \approx}$$

$$\frac{\mathbf{F\ P_{OBS}(h,t)}}{\mathbf{F + (1\text{-}F)\ e^{1/f \int (0,\ t) POBS(h,t)\ dt}}}$$

where $\mathbf{P_{OBS}(h,t) = [1 - e^{-V_{OBS}(h,t)}]}$

$$V_{OBS}(h,t) = \mathbf{Cinit}\,(n)\int_{0}^{t} a\,N_a \frac{d\left(1 - e^{\left(\mathbf{Cprom(m)}\ 2^{(\mu)(t-a)}\right)}\right)}{d\,(t - a)}\,da$$

INITIATION PROMOTION

NORMAL CELLS ->“n” events-> **PRENEOPLASIA**-> “m” events->**NEOPLASIA**

“zero” = adult growth rate

μ = preneoplastic growth rate