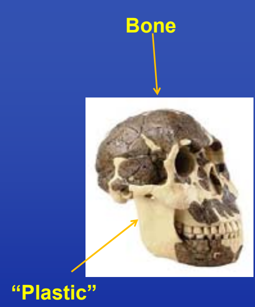
Why Bother with Modelling?

Eduardo Massad

edmassad@usp.br

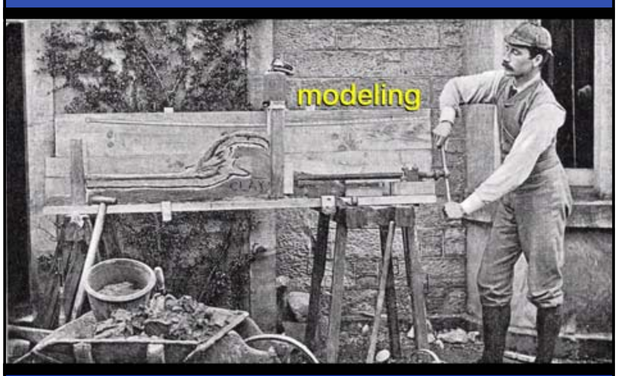
Image courtesy of the Smithsonian's Human Origins Program; John Gurche, reconstructions artist



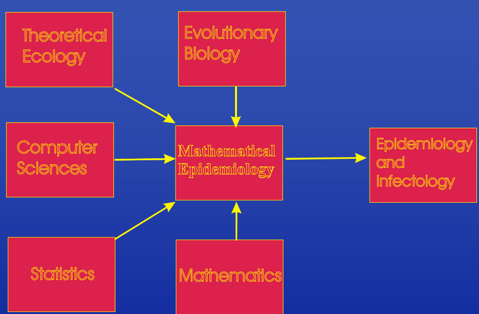
Bone

“Plastic”

A Model can be defined as *‘a convenient representation of something important’*



Diseases are complex entities!



EPIDEMIOLOGY

Disease’s character	Non-infectious	Infectious
Objective	*Causality*	*Control*
Cognitive approach	*Induction*	*Deduction*
Tools	*Statistics*	*Mathematics*
Models	*Functional*	*Structural*
Underlying aims	*Risk Factors*	*Mechanisms of the disease*

Why Use Mathematical Models?

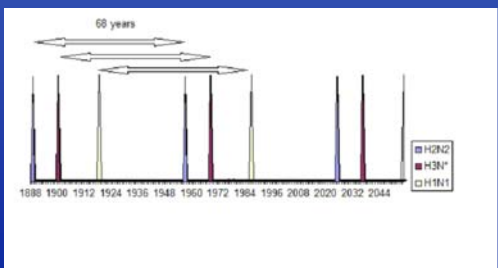
- Modeling perspective
  - Mathematical models
    - » reflect the known causal relationships of a given system.
    - » act as data integrators.
    - » take on the form of a complex hypothesis.
- Benefits of modeling
  - Provides information on knowledge gaps.
  - Provide insight into the process that can then be empirically tested.
  - Provides direction for further research activities.
  - Provides explicit description of system (mathematical vs. conceptual models)

Inductive reasoning

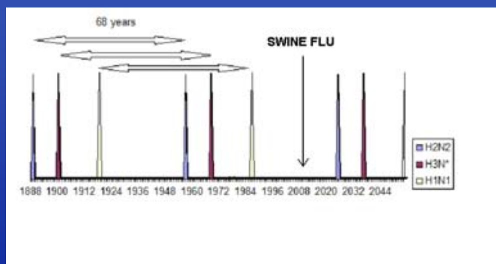
Looking for patterns

Disciplina de Informatica Medica FMUSP

The Hilleman Hypothesis



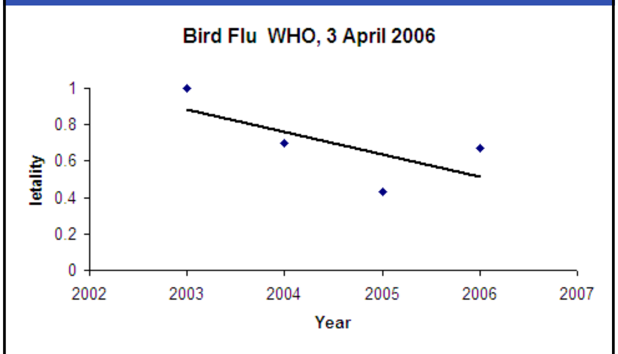
Disciplina de Informatica Medica FMUSP



Disciplina de Informatica Medica FMUSP

Bird Flu

Disciplina de Informatica Medica FMUSP



Disciplina de Informatica Medica FMUSP

Deductive reasoning

Looking for "the first principles"

Inductive reasoning is concerned with <u>conclusions that probabilistically (= with a certain likelihood) follow from their premises</u>.

Deductive reasoning is concerned with <u>conclusions that follow necessarily or with certainty</u> from their premises.

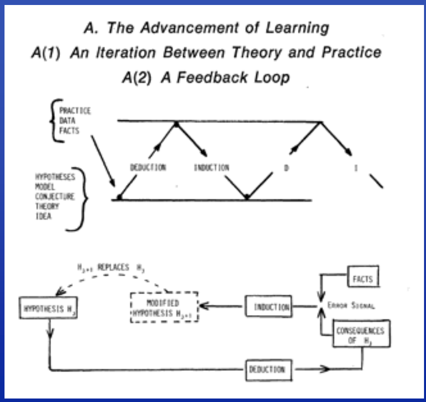
Here are two premises:

Simon is the brother of Julie
Julie is the mother of Emma

Now I'm giving you two different conclusions:

a) Simon is the uncle of Emma

b) Simon is older than Emma

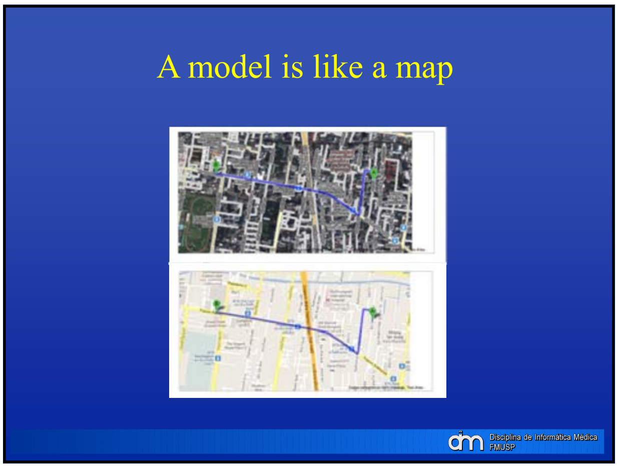


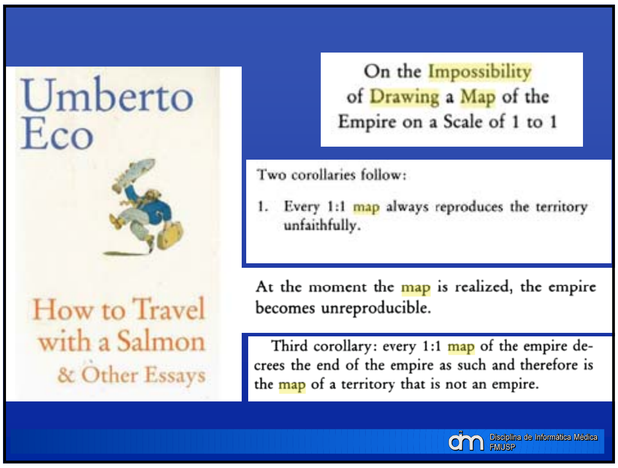
Reductionism

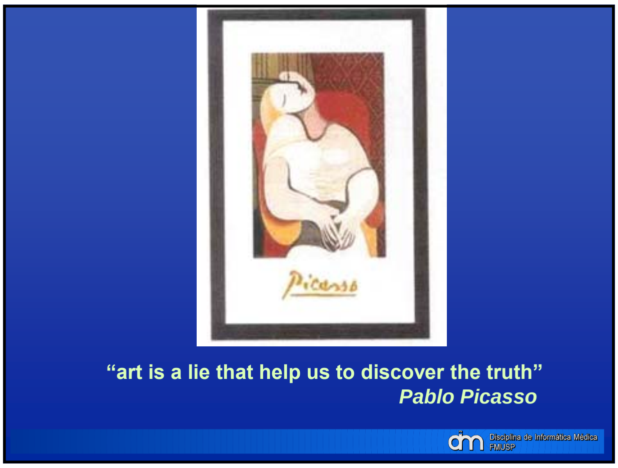
Reductionism is a philosophical concept according to which a complex system is nothing but the sum of its parts, and that an account of it can be reduced to accounts of individual constituents.

Modelling infectious diseases is quintessentially reductionist: very complex entities like diseases are reduced to a ser of mathematical equations!

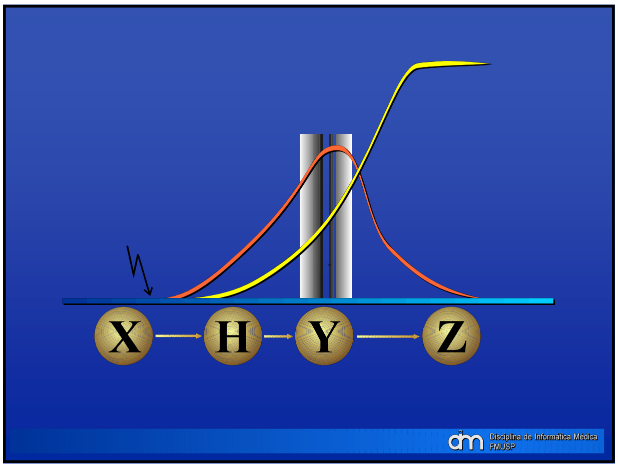
A Model can be defined as *'a convenient representation of something important'*

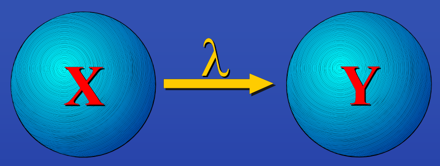

A model is like a map


Umberto Eco
How to Travel with a Salmon & Other Essays
On the Impossibility of Drawing a Map of the Empire on a Scale of 1 to 1
Two corollaries follow:
1. Every 1:1 map always reproduces the territory unfaithfully.
At the moment the map is realized, the empire becomes unreproducible.
Third corollary: every 1:1 map of the empire decrees the end of the empire as such and therefore is the map of a territory that is not an empire.


Picasso
"art is a lie that help us to discover the truth"
Pablo Picasso

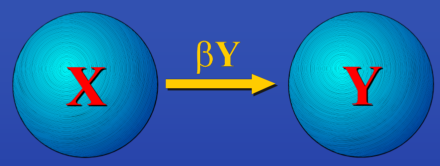

Ceci n'est pas une pipe.


X
H
Y
Z



$$\begin{cases} \dfrac{dX(t)}{dt} = -\lambda(t)X(t) \\ \dfrac{dY(t)}{dt} = \lambda(t)X(t) \end{cases}$$

$$X(t) = X(0)\exp\left[-\int_0^t \lambda(s)ds\right]$$



$$\begin{cases} \dfrac{dX(t)}{dt} = -\beta X(t)Y(t) \\ \dfrac{dY(t)}{dt} = \beta X(t)Y(t) \end{cases}$$

$$Y(t) = \frac{X(0)\exp[\beta X(0)t]}{\exp[\beta X(0)t] + KX(0)}$$

Definition of Modelling

A model can be defined as a "convenient representation of anything considered important". This is an operational definition and, when the representation consists of quantitative components, the model is called a mathematical model

The process of Modelling consists of "a set of complex activities associated with the designing of models representing a real-world system and their solution". As we shall see later on the solution may be analytical or numerical.

A Model is composed by the following items:

Variables: the quantities of interest that varies with time or age, like the number (or proportion) of susceptibles to a given infection;

Parameters: quantities that determine the dynamical behavior of the systems, like the incidence rate;

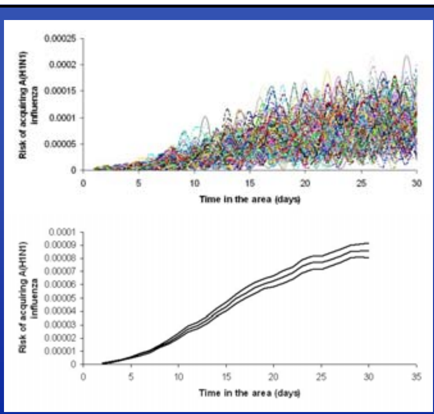
Initial and boundary conditions: the initial values of the variables with time (initial conditions) or age (boundary conditions).

Models classification:

Stochastic – include probability elements on its dynamics;

Deterministic – once defined the value of the parameters and initial conditions, all the course of its dynamics is determined.

Deterministic	Stochastic
Relatively little details	Incorporate a lot of details
Population based	Individual-based



The three qualities of models:

1. Parsimony;
2. Generality;
3. Prediction.

Purposes of Modelling

- to help the scientific understanding and precision in the expression of current theories and concepts;

- identification of areas in which epidemiological data is required;

- prediction.

Forecasting vs Projection Models

- Forecasting: prediction before the happening;

- Projection: what would have happened if…

588 E. MASSAD AND OTHERS

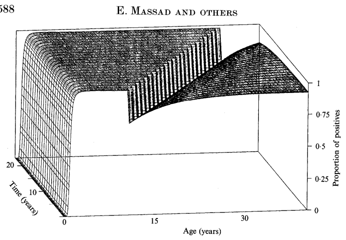


Fig. 7. Time evolution of the proportion of seropositives, stratified by age. It can be seen that the figures have a valley beginning at 10 years of age in time 0, which travels upwardly in age with time. This corresponds to a susceptibility 'window' and it divides the population into those who are positive due to natural infection (plane right to the window) and those who are positive due to vaccination (plane left to the window).

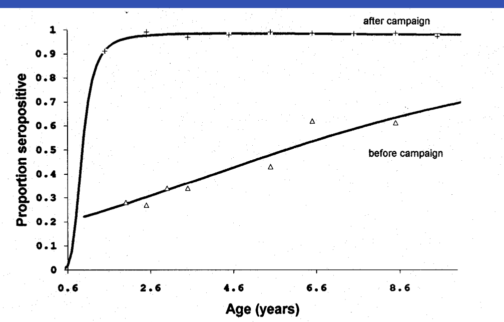


FIGURE 2 *Comparison of rubella-specific antibodies prevalence before (triangles) and after (crosses) mass vaccination campaign, with respective fitting curves*

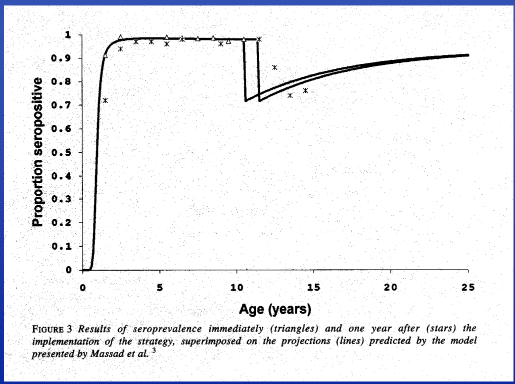
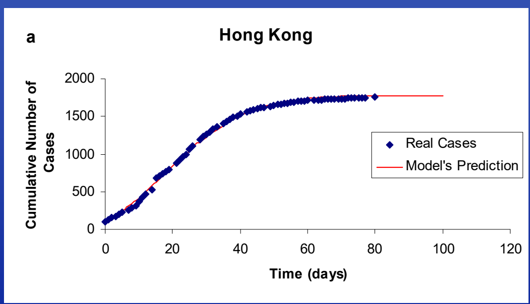
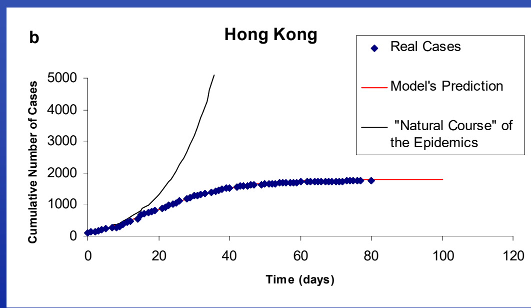
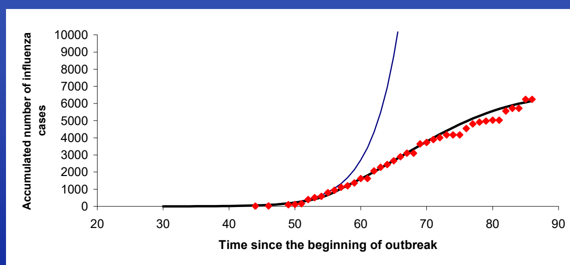


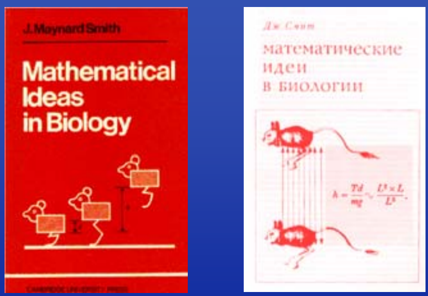
FIGURE 3 *Results of seroprevalence immediately (triangles) and one year after (stars) the implementation of the strategy, superimposed on the projections (lines) predicted by the model presented by Massad et al.* [3]

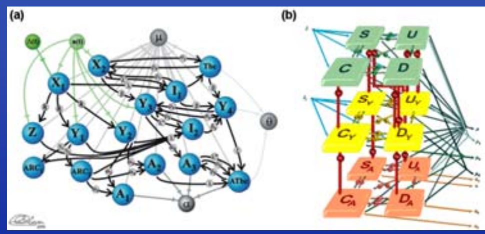


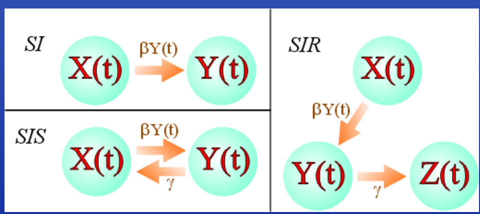


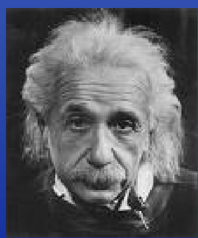
SWINE FLU











Models must be kept simple, but not simpler...

Albert Einstein

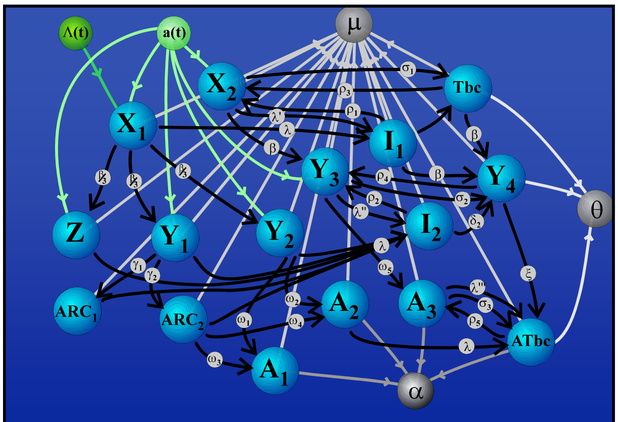
With four parameter and I fit an elephant;
with five and I make it wiggle its trunk!

John von Neumann

Milestone of Modeling Studies

The importance of simple models stems not from realism or the accuracy of their predictions but rather from the simple and fundamental principles that they set forth.

Complicated, Complex and “NP” Models



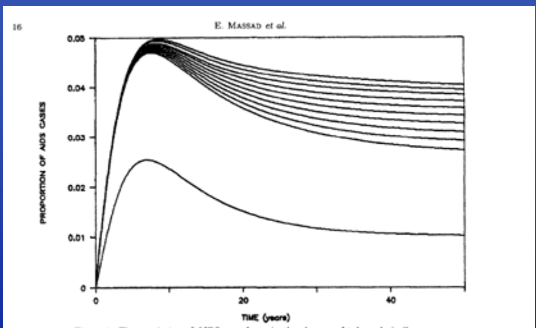
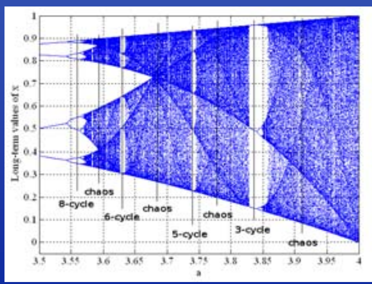


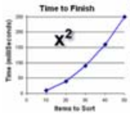
Figure 4. Time variation of AIDS prevalence in the absence of tuberculosis (lower curve) and crescent incidence of tuberculosis (upper set of curves).

$$x_{t+1} = ax_t(1 - x_t), \quad (\forall\, x \in R, \quad 0 \le x \le 1)$$



Its All About "Time to Solve"

If you measure how long a program takes to run when given more and more difficult problems, such as sorting a list of 10 items, 20 items, 30 items etc, you can then plot the times and come up with a function.



For example the time might increase by x^2, so a problem that is twice as hard takes 4 times as long.

That program would be in "P", as it is solvable in "Polynomial" time. In this case the polynomial is:

$$t = x^2$$

But if the time went up exponentially or factorially, or something that exceeds what a polynomial can do, it would not be in "P". It is not solvable in "Polynomial" time.

Now, the "N" in "NP" refers to the fact that you are not bound by the normal way a computer works, which is step-by-step. The "N" actually stands for "Non-deterministic". This means that you are dealing with an amazing kind of computer that can run things simultaneously or could somehow guess the right way to do things, or something like that.

Travelling Salesman Problem

The classic example of "NP-Complete" problems is the Travelling Salesman Problem.

Imagine you need to visit 5 cities on your sales tour. You know all the distances. Which is the shortest round-trip to follow? ABCEDA? ADECBA?

An obvious solution is to check all possibilities.

But this only works for small problems. If you add a new city it needs to be tried out in every previous combination.

So this method takes "factorial time": $t = n!$

(Actually $t = (n-1)!$ but it is still factorial.)

Lets say the program could solve a 20-city problem in 1 second, then

- a 21-city problem would take about 21 seconds to solve.
- And a 22-city problem would take about 462 seconds (over 7 minutes),
- and a 30-city problem would take 3 Million Years. Ouch!

Luckily, there are special ways to break the problem into sub-problems (called "dynamic programming", but the best still take "exponential time": $t = 2^n$

So a program that solved 20 cities in 1 second would take about 10 minutes to solve a 30-city problem and a 60-city problem would take 35.000 Years.

All models are wrong but some models are useful

George Box

THRESHOLD CONDITIONS

Disciplina de Informática Médica FMUSP

The Reproductive Rate of Infections

The Basic Reproductive Number, R_0, is the number of secondary infections produced by a single infectee during his/hers entire infectiousness period in an entirely susceptible population.

Disciplina de Informática Médica FMUSP

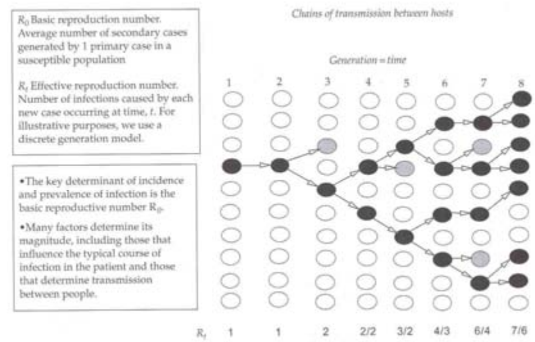


Figure 10.6 Diagrammatic representation of chains of transmission. The speed of spread is determined by the case reproductive number R_0.

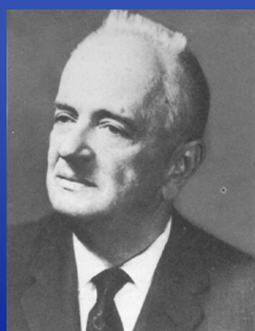
Disciplina de Informática Médica FMUSP

Infection	Geographical Location	Time Period	R_0
Measles	England	1947-1950	13-14
	USA	1918-1921	5-6
Pertussis	England	1944-1978	16-18
Chicken Pox	USA	1912-1921	7-8
Diphtheria	USA	1918-1919	4-5
Mumps	England	1960-1980	7-8
Rubella	England	1960-1970	6-7
Poliomyelitis	USA	1955	5-6
Malaria	Nigeria	1972	80-200
HIV	England	1981-1985	2-5
	Kenya	1981-1985	11-12
	USA	1981-1984	5-6
	Brazil	1991	90

Disciplina de Informática Médica FMUSP

"Nothing in biology makes sense except in the light of evolution."

T. Dobzhansky, 1973.



Disciplina de Informática Médica FMUSP

The Red Queen Hypothesis

For microparasites, in a homogeneously mixing population, the reproductive value, R, is a function of the product of the number of potentially infective contacts, β, the proportion of susceptible hosts, *x*, and the time of permanence in the infective condition, T:

$$R(t) = \beta T x(t)$$

For Vector-borne Infections

$$R_0 = ma^2 b\exp(-\mu n)/r\mu > 1$$

$$m_{th} = r\mu/a^2 b\exp(-\mu n)$$

$$\beta \longrightarrow I(t) \longrightarrow \gamma$$

$\beta > \gamma \longrightarrow$ epidemic

$dY(t)/dt = \beta X(t)Y(t) - \gamma Y(t) > 0 \Rightarrow$ epidemics

$[\beta X(t) - \gamma] > 1$

$t = 0, X(0) = 1$

$\beta > \gamma \Rightarrow dY(t)/dt > 0 \Rightarrow$ epidemics

The critical proportion to vaccinate

Let us assume that a fraction *p* of the susceptible is vaccinated against a certain disease, as soon in life as possible. Then:

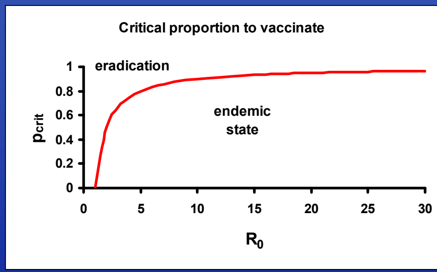
$$X(0) = 1 - p$$

$$dY(t)/dt = [\beta(1 - p) - \gamma]Y(t)$$

$$[\beta(1 - p)/\gamma - 1] = R_0(1 - p) - 1$$

$$R_0(1 - p) < 1 \rightarrow \text{eradication}$$

$$P_{crit} = 1 - 1/R_0$$



"As a matter of fact, all epidemiology must be considered mathematically, if it is to be considered scientifically at all. To say that a disease depends on certain factors is not to say much…"

Ronald Ross, 1911