1

Basics Definitions and Compartmental Models

> *Some models are wrong, but most of them are useful.*
>
> – Unkown author

Models in science have two different roles: **understanding** what happens and **prediction**. There are two types of models: one more simple and one more complex. In the simplest one you just consider the minimal number of parameters and events involved: this allow to understand what are the main mechanism of a phenomena.

In this course, we gonna start with very simple models in which we firstly assume that there are no structures in the population. This is not accurate, but allows us to understand at a first glance some underlying mechanism. Then, we will consider social structures and contact network models. We will also take into account the interaction of different populations and we analyze how data move from one population to the other. At last, there is another class of models called "Agent Based" models, but we will only brief introduce it.
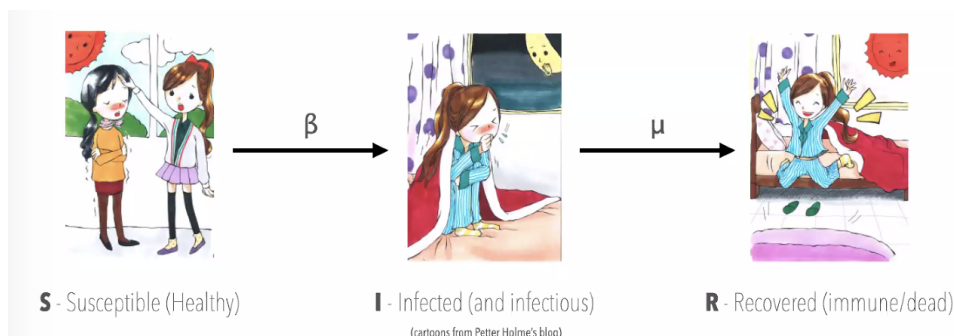
1.1 Comportamental models

We introduce the **comportamental models** in which most based epidemiologist theories are based on. In reality, there are different levels of understanding how the diseases diffuse, as for instance at a biological level or a more simplest one. It is impossible to insert all the details of a process in a models. We need to summarize all the biological process in few parameters which are the average of what you see inside the population. This is the same principle behind the statistical mechanics in which we concentrate for large scale effects.

Let us consider a population of any individuals and try to characterize them. For

S - Susceptible (Healthy)   **I** - Infected (and infectious)   **R** - Recovered (immune/dead)

(cartoons from Petter Holme's blog)

Figure 1.1: Classification of infected population in three different stage of the disease.
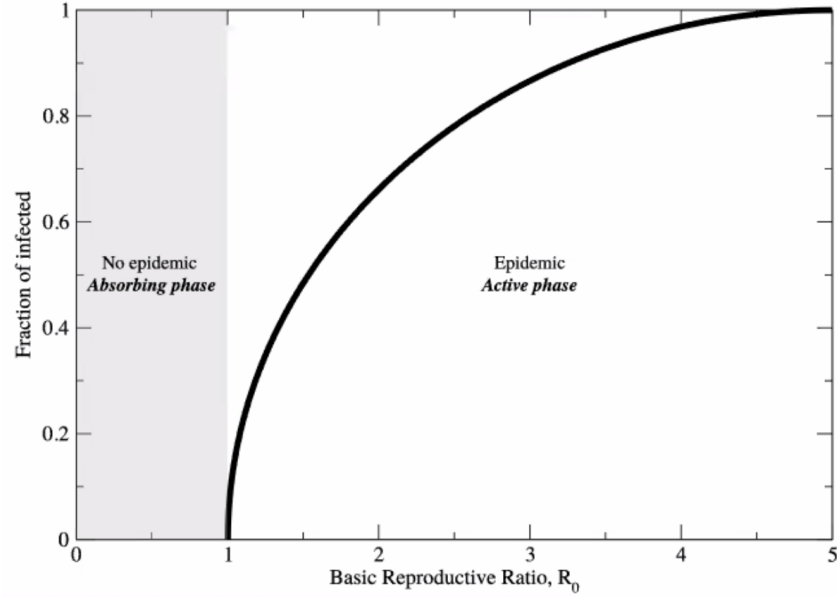
Figure 1.2:   Fraction of infected vs basic Reproductive Ratio, R_0.

instance we have three compartments and we want to classify people according to the state of their diseases, as in Fig. 1.1. We have also transition from one state to the other with rates as β and μ which describes the dynamic of the diseases The only problems it is that this approximation is quite strong, because by fixing the rates we are assuming that the process underlying is a Markovian one. In reality, you do not have a exponential distribution but a decay. This will be seen in the course when we will talk about "non-Markovian" epidemics. For instance the β can be the "per contact" infectous rate, so you are just counting the number of contacts. As examples we can consider different models which depend on the type of disease as SI, SIS, SIR and so on and so forth.

There are difference between medical status and infection status. We do not care about medical status but only about the disease and how the immuning system is reacting. We have four main stages: starting from a helthy state, the individual contract the disease and then it become infectious until it recover. The most important things is that the compartments are not the same of the medical status.

Now, let us introduce the **Basic Reproductive Number** R_0 (pronunced R naught) which is a measure of the infection of the population. We put one guy inside a group for instance for three days and at the end of these days we count the number of secondary cases that we have. This is the main idea of R_0. This number determines wheter a disease will spread or not:

$$\begin{cases} R_0 < 1 \\ R_0 = 1 \\ R_0 > 1 \end{cases} \tag{1.1}$$

Let us consider the plot of Fig. 1.2, we have a sort of second order phase transition at the point $R_0 = 1$. Note that the R_0 number for the SARS is higher thant the one of COVID-19, but we did not experienced an outbreak of this disease. For computing R_0 we are assuming that the population is totally susceptible, but this assumption is valid only at the very early stages. Then, we care ombining also epidemiological and demographical features. The conclusion is that this number could vary from one population to another. Since we are doing a coarse-graining of the dynamics, the number R_0 represents the mean of possibly different distributions. It can give the

wrong idea of similar R_0, means similar outbreaks. The distribution of infections can be quite heterogeneous, hence the mean could be quite representative in homogeneous populations. For instance, let us consider the plot in Fig. 1.3. We see that SARS is etherogeneous, while Spanish Flu was a homogeneous one. The COVID-19 is most likely somewhere in the middle.
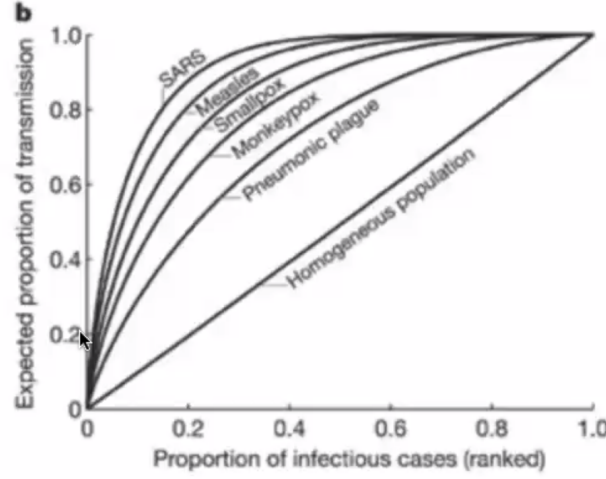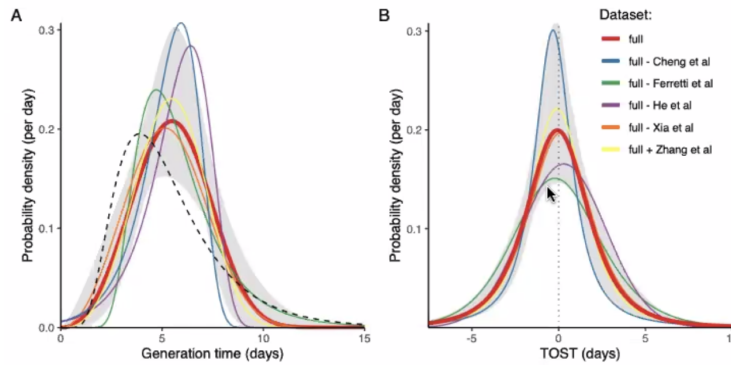

Figure 1.3: Figure from: Lloyd-Smith et al. Nature 438, 355–359 (2005).

Now, we introduce the **Effective Reproductive Number** $R(t)$, which is the same of R_0 but it varies in time. Hence, it is the average number of secondary cases a case produces in a population at time t.

Another important quantity is the **Infectious period**:

$$\tau = \frac{1}{\mu}, \quad \tau = \frac{1}{(\alpha + \mu)} \tag{1.2}$$

Other relevant quantities are the **Incubation period** (from infection to symptoms), the **Generation time** (from infection to infection), the **Serial interval** (from symptoms to symptoms) or the **TOST** (measures from symptoms to infection). The problems is that TOST in many cases can be negative (see Fig. 1.4 for more details).

Figure 1.4: Figure from: Ferretti et al. `https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v1`