Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Mathematical Epidemiology, SIR Models and COVID-19

by

Stephan Luckhaus

Preprint no.: 60 2020

Mathematical Epidemiology, SIR Models and COVID-19

Stephan Luckhaus

University of Leipzig, Mathematical Institute,
Augustusplatz 10, D-04109 Leipzig, Germany
Stephan.Luckhaus@math.uni-leipzig.de

May 24th, 2020

1 Introduction and Summary

This article is meant to contribute to the discussion about containment strategies for Corona. Based on the mathematical theory of infectious diseases, we discuss the stability or instability of a state/situation where a fraction s_j of each age group j in the population is still susceptible, whereas the rest of the population is already immunized. The simple meaning is, that in an unstable state, the epidemy will break out again.

In contrast to the simple theory where there is only one parameter - the so-called herd immunity s_0 - such that all states with s larger than s_0 are unstable and those below s_0 are stable, we now have a region in k-parameter space of susceptibilities s_j, $j = 1 \cdots, k$, where k is the number of age groups, that is the stability region. It is given by a matrix $(A_{jl})_{1\le j,l\le k}$ of cross infection rates, and not by one infection rate. If the dynamical system

$$[A(s)]^n(1, \cdots, 1) \ , \quad \text{with } A(s) = (s_j A_{jl})_{1\le j,l\le k}$$

is exponentially growing then s is unstable; if it is decreasing, then s is stable.

Deadliness of Corona depends dramatically on age. If you are below fifty, the probability of dying, if you contract the disease, is still in the order of 10^{-4}, similar to the risk from a vaccination. But this number grows more or less exponentially with age, with a factor 3 for 10 years of age difference. In the last section of the present paper I explain how these mortalities can be estimated. Basically I take the cases of Corona deaths in Germany as reported by the Robert Koch Institute (RKI) on May 15th, and compare them to an assumed 5% infection rate in the German population in the above age group. Here I have to thank the Rector of my University, Professor Beate Schücking, for explaining the method to me in a long phone conversation. She is not supporting my conclusions. But that is what scientific debate is all about. The 5% are a rough estimate and based on the fact that the infection rate for the icelandic population was 5 per mille on April 15th. And Iceland was able to test 13% of the whole population for coronavirus, which was sufficient to stop the further transmis-

sion of the virus almost completely already in April by quarantining the infectious individuals. Of course we are not able to do anything of that kind in Germany.

Coming back to the stable regions of susceptibility, the region that we should aim for is obviously the one where the susceptibles vector is of the form $(0, \cdots, 0, \sigma, 1, \cdots, 1)$, where $s_j = 0$ in the groups of younger ages and $s_j = 1$ in the higher age groups, because that means the minimum loss of lives.

It is of course already too late in some sense. But what we should still do is, to try to get as many infections as possible among the under forty years old, and warn the elderly to stay in splendid isolation during the course of the epidemy. You need as many people immunized under forty as possible, and not as few infections as possible in that age group. If you really want to save lives as opposed to grandstanding you should immediately lift all the lockdown restrictions and be as candid as possible about the true risk in each age group.

2 Overview

The aim of the following two sections is to explain to a non-mathematical audience the SIR model of mathematical epidemiology and also its relation with a particular stochastic model - of meanfield type - for the spread of infectious diseases. This part grew out of a heated discussion with my colleague Matthias Kreck from Bonn University, whom I want to thank here. The results can be summarized as follows. The stochastic model reduces in the limit for large populations to a deterministic age structured model, which is a model that keeps the information, $a =$ *time since infection*, for the infected. To be precise this model has the following variables for $j = 1, \cdots, k$:

- n_j is the percentage of individuals in the j-th subpopulation, not depending on time t.
- $s_j(t)$ is the percentage of susceptibles, i.e. of individuals who have not yet been infected in the subpopulation j.
- $i_j(t, a)$ is the percentage of infectious individuals who have been infected at time $t - a$ in the subpopulation j.
- $r_j(t)$ is the percentage of removed, i.e. those who are immune, dead, or in quarantine in the subpopulation j.

If you are interested mainly in the outcome of the infection, that is the percentage $n_j r_j(\infty)$ of removed at the end - i.e. when $i_j(t)$ has returned close to zero - then there

is no need for a computer simulation. And that is because the system has invariants of the form

$$\log s_j(t) - \sum_{l=1}^{k} A_{jl}\ s_l(t) - \sum_{l=1}^{k} \int_0^d B_{jl}(a)\ i_l(t,a)\ da\ .$$

Here the A_{jl} are the effective (cross-) infection rates, and the weighted integrals in the second sum represent effective infection levels.

Moreover, the question whether the state $s_j = \sigma_j, i = 0$ is stable or unstable, i.e. such that the introduction of a tiny number of infected will start the epidemy, will be determined by the matrix $\sigma_j A_{jl}$ alone, more precisely for the physicists and the mathematicians its spectral radius $R_A(\sigma)$.
If $R_A(\sigma) < 1$, then $s_j = \sigma_j$ is stable.
If $R_A(\sigma) > 1$, then $s_j = \sigma_j$ is unstable.
So to predict $r_j(\infty)$ and with it the total number of deaths during the epidemy, you have to know the parameters of your mean field model. Section 1 explains how to calculate the parameters in the age structured model. Section 2 explains how to calculate A_{jl} and $B_{jl}(a)$ from those. Then you have to estimate $s_j(t_0)$ and $i_l(t_0, a)$ from data, where t_0 is the current time. And finally you have to find the unique solutions $s_j(\infty)$ of

$$\log(s_j(\infty)) - \sum_{l=1}^{k} A_{jl}\ s_l(\infty) = \log s_j(t_0) - \sum_{l=1}^{k} A_{jl}\ s_l(t_0) - \sum_{l=1}^{k} \int_0^d B_{jl}(a)\ i_l(t_0,a)\ da$$

among the stable σ_j, e.g. by the Newton method.

Do I trust the model?

Qualitatively certainly for even not so large numbers of infected. Below a certain threshold though the stochastic process cannot be replaced by the deterministic one, and even new phenomena, like reappearence of the virus in a removed individual - think of varicella - have to be taken into account. So the model is a model for the course of the epidemy, not for the time when the epidemy starts.

Can it help for Corona?

Yes, in two ways, because there are 60% of infected without the obvious symptoms, you can never hope to find all infected with only partial testing for the virus. The formula giving the A_{jl} from the stochastic model explains how far quarantine methods combined with sample testing for the virus can reduce these effective (cross-) infection rates.

Because Corona is not what I would call a deadly illness for the under forty years old, and the mortality of the infected becomes noticeable for the 60-70 years old and frightening for the over 80 years old, who on the other hand have a lower contact

rate, it is important to deal with several cohorts or subpopulations distinguished by age. As mentioned in the summary, the model clearly suggests, that the best strategy is to reduce the contact rate between young and old temporarily, but let the infection reach levels of immunization that are well in the stable regime, say 80%, in the subpopulation of the under 40 years old. The only constraint is to keep the level of hospitalized infected manageable.

The model also tells you that continuing something like an almost complete lockdown during the tail end of the epidemy under lockdown conditions $A_{jl}(lock)$ is unnecessary, since you are already in the stable regime for a higher effective $\hat{A}_{jl}$. But as I already mentioned, reducing contact rates irrespective of age is in any case useless.

3 Classes of models.

Mathematical epidemiology is a subfield of population dynamics. The oldest type of models are recursive sequences. Probably the most famous of these are the Fibonacci numbers proposed by Leonardo of Pisa around 1200 to model the unchecked growth of a rabbit population. (Actually he was most probably not primarily interested in rabbits but in continued fractions). In formulas these models are

$$n_{k+l+1} = f(n_k, n_{k+1}, \cdots, n_{k+l}) \ ,$$

for an algorithm f having as input the number of individuals in the preceding l generations and as output the number of individuals in the current generation, i.e. the rabbits of the l previous generations "produce" the new generation. These generational models are still popular but mostly in the form of stochastic processes in discrete time. There the final formula is

$$n_{k+l+1} = f(n_k, n_{k+1}, \cdots, n_{k+l}, \omega) \ .$$

This is shorthand for an f that has as output not a number but probabilities. Mostly the examples are of just three conditional probabilities, birth, death or keeping the status quo

$$p_i(n_k, \cdots, n_{k+l}) = P\left(n_{k+l+1} = n_{k+l} + i \mid n_k, \cdots, n_{k+l}\right) \ ,$$

where $i = 1, 0,$ or -1 and $\sum_{i=-1}^{1} p_i = 1$. As an algorithm that means that the output is no longer one number but three positive real numbers. As a dynamical system that is initialized with l natural numbers, it produces likewise a sequence not of numbers but of probabilities to observe a natural number, i.e. a sequence of sequences of length $n_k + l$ of positive real numbers.

The problem with generational models is that generations are not synchronized. Even in humans it was not so rare to find e.g. an uncle who is younger than his niece. The

second problem is that time is continuous and not discrete. The simplest type of stochastic models that avoids these types of shortcomings are mean field processes in continuous time.

In order not to waste too much paragraphs I will switch now to a model for the spread of an infectious disease of S(usceptible), I(nfected), R(emoved) type. Here we have a population of size N, $S(t)$ is the number of individuals that have not yet been infected at time t, $I(t,a)$ is the number of individuals that have been infected at time $t-a$, and $R(t)$ is the number of removed individuals, i.e. either dead or immune.

From the point of view of an individual in the population the process or course of the illness looks as follows. The individual is in the set of susceptibles up to a random time t_1, when it gets infected, then the virus starts to multiply within the individual. When the virus count reaches a certain value, the infected becomes infectious with infectivity $\alpha(a)$ at time $t = t_1 + a$, and after a certain time length d, the function α will become zero. Between t_1 and $t_1 + d$ the infected individual will be removed at a time $t_1 + a_2$, which again is random. The usual way to model these random times is as independently exponentially distributed, that is

$$\frac{P(t_1 > t + h) - P(t_1 > t)}{hP(t_1 > t)} \approx \frac{1}{N} A(t) \quad , \quad \frac{P(d_2 > a + h) - P(d_2 > a)}{hP(a_2 > a)} \approx \delta(a) \ .$$

The mean field assumption is that

$$A(t) = \alpha_0 \sum_{t=t_1+a; a<a_2} \alpha(a) \ ,$$

where α_0 is the contact rate in the population. In words, this assumption means that the probability of two individual meeting is independently identically distributed (i.i.d.) - the magic notion in stochastics.

This model I will call the stochastic SIR model. Historically the deterministic SIR model preceeds the stochastic model. It is around 100 years old and does not take effects like the incubation time of the disease into account. It is of the type of equations describing mass action kinetics in chemical reactions, i.e. a simple system of O(rdinary) D(ifferential) E(quations). The simple SIR models reads

$$\begin{aligned} \partial_t s &= -\alpha i s \\ \partial_t i &= \alpha i s - \beta i \\ \partial_t r &= \beta i \ . \end{aligned}$$

The meaning is: s, i, r are the percentages of susceptibles, infected, and removed in the population, α is the infection rate, and β is proportional to $1/d$, where d is the duration of the infection. The connection between the stochastic SIR model and this

deterministic SIR model is indirect. This will be discussed in the next section.

Here I will just point out that it is easy to show that in the large N-limit the stochastic SIR model will become deterministic. The limit is not a simple ODE, since it keeps the information of the time a since infection for the infected. Formally the equations are P(artial) D(ifferential) E(quations) in t, a, albeit of the simplest type

$$\begin{aligned} \partial_t s(t) &= -\alpha_0 \left[\int_0^d \alpha(a)\ i(t,a)\ da \right] s(t) = -i(t,0) \\ \partial_t i(t,a) + \partial_a i(t,a) &= -\delta(a)\ i(t,a) \\ \partial_t r(t) &= i(t,d) + \int_0^d \delta(a)\ i(t,a)\ da\ . \end{aligned}$$

Models of this type have been extensively studied. They are called delay differential equations, integro-differential equations or (age)structured models. They define a dynamical system but in the (infinite dimensional) space of functions $i(t,\cdot)$ in the interval $[0,d]$. So they are difficult to observe, e.g. by a virus test, which will be negative during incubation and when the individual has ceased to producing the virus, and which never will be so precise as to give you the a at the time of the test.

4 Qualitative behavior of the SIR model, herd immunity.

I will start the discussion with the simplest SIR model. This model is almost explicitly solvable. If you use the method of separation of variables, which most science students will remember from their calculus class, the equations become

$$\partial_t(s+i+r) = 0 \quad , \quad \partial_t(\log s + \frac{\alpha}{\beta} r) = 0 \quad , \quad \partial_t i = \alpha i s - \beta i\ .$$

So we have not one invariant or integral, but two:

$$i + s + r = 1 \quad \text{and} \quad \partial_t(\log s - \frac{\alpha}{\beta} s - \frac{\alpha}{\beta} i) = 0\ ,$$

and as a consequence everything reduces to the equation

$$\partial_t s = \beta\ \gamma(0)\ s - \beta\ s \log s + \alpha\ s^2\ ,\ \text{where}\ \beta\ \gamma(0) = \beta \log s(0) - \alpha\ s(0) - \alpha\ i(0)\ ,$$

and the asymptotic limit for large times will be given just by the identities

$$i(\infty) = 0\ \text{ and }\ \beta \log s(\infty) - \alpha\ s(\infty) = \beta\ \log s(0) - \alpha\ s(0) - \alpha\ i(0)\ ,$$

with the additional information $\partial_s(\log s - \frac{\alpha}{\beta} s)(\infty) \geq 0$, since s and i will both eventually decrease. So the discussion of the simple curve $\beta \log s - \alpha s$ tells everything

about the final outcome of the epidemy.

If you start with any $s(0), i(0) > 0$ for large times, the fraction of removed will approach the unique solution of $\beta \log s - \alpha s = \beta \log s(0) - \alpha s(0) - \alpha\, i(0)$ with $s < \frac{\alpha}{\beta}$. This value, $\frac{\alpha}{\beta}$ is called herd immunity. Actually it has two interpretations. The first is what I just explained. It is the maximal possible ratio of the population, which has escaped the infection during the whole course of the epidemy. The second is: It is the value of the ratio of susceptibles in the population where the number of infected starts to decrease.

If you believe in this simple model, the message for disease control is equally simple. Suppose the epidemy starts with small $i(0)$ and $s(0) > \frac{\beta}{\alpha}$ and suppose you are able, but only temporarily, to decrease α, how far should you decrease α? Well, you know in any case, that $s(\infty) \leq \frac{\beta}{\alpha}$ eventually. So obviously to get there quickly and not to overshoot the optimal choice is

$$\tilde{\alpha} = \alpha\beta\, \frac{\log\beta - \log\alpha - \log s(0)}{\beta - \alpha[s(0) + i(0)]}\ .$$

If you reduce the contact ratio more, that will get you only to an $s(T), i(T)$ at the time T of lifting the temporary restrictions, which is the starting point of a new epidemy.

Let me also briefly discuss what we call singular perturbation of the simple SIR model. The SIR model is very untypical for an ODE in two variables, s and i, since it has the whole interval $\{(s,i) \mid 0 < s < 1, i = 0\}$ as stationary points. If you perturb it a bit (ε, as usual, means a small number) e.g.

$$\partial_t s = -\alpha i\ s + \varepsilon(1-s) \quad , \quad \partial_t i = \alpha i s - \beta i\ .$$

the situation changes. The interpretation of the added term is that on a slower time scale either removed individuals loose their immunity or the population changes by the natural birth death process. For this singularly perturbed system there are only two stationary states $(i_0, s_0) = (0,1)$ and $(i,s) = \left(\varepsilon\left(\frac{1}{\beta} - \frac{1}{\alpha}\right), \frac{\beta}{\alpha}\right)$, and every solution which starts with positive $i(0)$ will spiral into (i_1, s_1). The behavior of the solutions will be that of so-called relaxation oscillations, they will move fast from unstable $s > \frac{\beta}{\alpha}$ and small i to stable $s < \frac{\beta}{\alpha}$ and small i, but then on the slow time scale s, will increase again, become unstable and so on. But the width of the oscillations will eventually decrease exponentially.

Now let us move to the state of the art s, i, r models. These are age structured models of the type discussed in the previous section but for several subpopulations: $s_j, i_j, r_j,\ j = 1, \cdots, k$.

The equations read

$$\begin{aligned}
\partial_t s_j(t) &= -i_j(t,0) \\
(\partial_t + \partial_a) i_j(t,a) &= -\delta_j(a) i_j(t,a)\ ,\ 0 < a < d \\
i_j(t,0) &= \left[\int_0^d \sum_{l=1}^k \alpha_{jl}(a)\ i_l(t,a)\ da \right] s_j(t) \\
\partial_t r_j(t) &= i_j(t,d) + \int_0^d \delta_j(a) i_j(t,a)\ da\ .
\end{aligned}$$

Again, there is an almost explicit formula with $\Delta_j(a) = \int_0^a \delta_j(\sigma)\ d\sigma$, namely

$$\begin{aligned}
\partial_a \left(e^{\Delta_j(a)}\ i_j(t,a) \right) &= -e^{\Delta_j(a)}\ \partial_t i_j(t,a)\ ,\ \text{or} \\
i_j(t,a) &= e^{-\Delta_j(a)}\ i_j(t,0) - \int_0^a e^{\Delta_j(\sigma) - \Delta_j(a)}\ \partial_t i_j(t,\sigma)\ d\sigma \\
&= -\partial_t \left(e^{-\Delta_j(a)} s_j(t) + \int_0^a e^{\Delta_j(\sigma) - \Delta_j(a)}\ i_j(t,\sigma)\ d\sigma \right)\ .
\end{aligned}$$

So apart from the k obvious invariants $n_j(t) = s_j(t) + \int_0^d i_j(t,a)\ da + r_j(t)$, we have again the k additional invariants of the form

$$\begin{aligned}
\log s_j - \sum_{l=1}^k A_{jl}\ s_l - \sum_{l=1}^k \left[\int_0^d B_{jl}(a)\ i_l(a)\ da \right] &= f_j(s,i) \\
A_{jl} = \int_0^d \alpha_{jl}(a) \exp(-\Delta_l(a))\ da\ ,\ B_{jl}(a) &= \exp(\Delta_l(a)) \int_a^d \alpha_{jl} \exp(-\Delta_l(\sigma))\ d\sigma\ .
\end{aligned}$$

And again we know that all s_j are decreasing. Stability or instability of a point $s_j(0)$ is determined by the matrix $s_j(0) A_{jl}$. If $s_j(0)$ is exponentially unstable, then the linearized system has an exponentially growing solution $\hat{i}$ with positive $\hat{i}_j$, or a solution of the linear system of integral equations

$$\hat{s}_j(0) = s_j(0) \sum_{l=1}^k \int_0^d \alpha_{jl}(a) \exp(-\Delta_l(a)) \exp(-\lambda a)\ \hat{s}_l(0)\ da\ ,$$

with $\lambda > 0$ and all $\hat{s}_l(0)$ of one sign. This is equivalent to $R(s(0)) > 1$, where $R(\sigma)$ is the spectral radius of the matrix $\sigma_j A_{jl}$. On the other hand, if $R(s(0)) < 1$ then $s(0)$ is stable. So stability of the age structured system is the same as for the system without age structure. It is also not difficult to show, that if you start a Newton interation to calculate

$$\log s_j - \sum_{l=1}^k A_{jl}\ s_l = \log \sigma_j - \sum_{l=1}^k A_{jl}\ \sigma_l + b_j\ ,\ \text{ with } b_j < 0$$

in a stable point σ, this will converge (monotonically in the sense of Krasnozelsky). So you have your choice how to calculate the unique stable solution of

$$\log s_j - \sum_{l=1}^{k} A_{jl}\ s_l = \log s_j(t_0) - \sum_{l=1}^{k} A_{jl}\ s_l(t_0) - \sum_{l=1}^{k} \left[\int_0^d B_{jl}(a)\ i_l(t_0, a)\ da \right] .$$

Let me now turn to the calculation of overshooting in a recursive SIR model. The SIR model for k subpopulations without delay is of the form

$$\begin{aligned} \partial_t s_j &= -\sum_{l=1}^{k} A_{jl}\ s_j\ i_l \\ \partial_t i_j &= \sum_{l=1}^{k} A_{jl}\ s_j\ i_l - i_j\ , \end{aligned}$$

where I scaled β to 1 by rescaling time. Examples of A_{jl} are

$$A_{jl} = \alpha_j \alpha_l n_l / \bar{\alpha}\ , \quad \text{where } \bar{\alpha} = \sum_{l=1}^{k} n_l \alpha_l .$$

Here α_j represents the contact rate of the respective subpopulation and n_j the fraction of that subpopulation. If for $l > j$ one has that $\alpha_l n_l$ is much smaller than $\alpha_j n_j$, the system has nearly a recursive structure. If $A_{jl} = 0$ for $l > j$, then we know that the stable equilibria of the system are $s_j > A_{jj}^{-1}$ for each j. So each subpopulation has its own 'herd immunity'. But the problem is that the invariants are

$$\log s_j - \sum_{l \le j} A_{jl}\ s_l - \sum_{l \le j} A_{jl}\ i_l\ .$$

So even if $s_j(0) < A_{jj}^{-1}$ but $s_l(0) >> A_{ll}^{-1}$ for $l < j$, one will produce a potentially huge overshooting. If you want to avoid this, by temporarily reducing α_j, the only way is to make sure, that the subpopulation l reaches its 'herd immunity' before the epidemy starts again in the subpopulation j.

5 What to infer from the analysis of SIR models for a COVID-19 like infection

There are two specifica of COVID-19 that the SIR models can deal with. The first is the large number of infected with no or not easily detectable symptoms. Around 60% across all age groups apparently have neither fever nor a strong cough. So they will not be discovered by temperature screening or even their GP. They will be detected if you are able to test whole communities repeatedly with the virus test. Otherwise

there is no hope to stabilize the zero infection rate below the herd immunity of those 60%. This is the conclusion of the analysis of the SIR model for two subpopulations.

The second specific feature is the huge dependence of the mortality rate of the infected on the age group. The cumulative date of the Robert Koch Institute (RKI) in Germany gave on May 15th, less than 1.1 per mille for the age group of $20-50$ years old. These numbers are not corrected for the undetected. So the probability of death from Corona in that age group is very small. Indeed the reported cases up to May 24th are just 2.5 per mille of the population in that age group. In the age group of $50-60$ years old the reported ratio of deaths by infection is a bit less than 1%. In my own age group, $60-70$ years old, it is already 3%. But again, the number of reported infections per population is 2.5 per mille. So it is not unlikely that the true mortality is lower by the factor of 20. In the age group 70+ the observed mortality is around 20% and that is also the value of reported deaths versus cases in nursing homes and homes for seniors. But the ratio of reported cases versus population is vastly different. It is 1.4 per mille in the group of age larger than 70 not living in nursing homes or homes for seniors, and 1.5 percent for people living in these institutions. So probably the ratio of observed infections to all infections is higher in the age group 70+ but still lower than in nursing homes.

What can we infer, coming back to what I said about overshooting in the last paragraph.

The immunizaton we have reached so far is not more than a few percent - epidemiologically irrelevant. A stable situation with a half way normal life requires more than 50%, probably $70\% - 80\%$ in the more active groups of the population, but also the risk there is much lower.

The only sensible strategy is to get to the basically inevitable 'herd immunity' rate in the lower age groups quickly, and try to keep people informed, where the number of infected is high. Cannot e.g. a professor over 60 then decide for him/herself whether to stay in splendid isolation in the home office or not, while Corona rages among the students? Otherwise trying to get infection rate to an unrealistically - unstable - low value, one risks overshooting above the 'natural herd immunity' of the less active part of the population, driven by the slow climbing of the immunity of the young.

Another aspect which I also would like to mention, even though I did not put it into the models, is the seasonal dependence of infectivity. My opinion is that we have to get to high levels of immunization for the age group below 60 before the 'flu' months November - March. And we don't have much time.

Personally I think that a careful comparison between icelandic and german data will get even higher rates of undetected infections in the age group below 50. So for these 'young' individuals getting the Corona infection is equivalent to a vast vaccination program, with the important difference that they will be all infectious. So the elderly

should be made very aware about their risk of getting infected and their own mortality risk.

References

Odo Diekmann, Hans Heesterbeek, Tom Britton (2012): *Mathematical Tools for Understanding Infectious Disease Dynamics.* Princeton University Press.

Stephan Luckhaus (2020): *Corona: Herd-Immunity und Eindämmungsstrategien.* https://www.youtube.com/channel/UC63wZKfxbUMyb8XkU48mxMw/ Youtube video in German.

Etienne Pardoux (2016): *Probabilistic models of population evolution.* Springer.

Horst R. Thieme (2003): *Mathematics in Population Biology.* Princeton University Press.

Hans Zinsser (1934): *Rats, Lice & History: The Biography of a Bacillus.* Little, Brown and Company, USA.