A 3D computational model for understanding tuberculosis lesions dynamics in lungs

Author: Martí Català Sabaté
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisors: Ignacio Pagonabarraga and Clara Prats
(Dated: June 10, 2016)

Abstract: Tuberculosis (TB) is an infectious bacterial disease caused by *Mycobacterium tuberculosis* (*Mtb*), which most commonly affects the lungs. In healthy people, an infection with *Mtb* often causes no symptoms, remaining controlled as a non-contagious latent tuberculosis infection. World Health Organization estimates that one third of the world population is already infected by this bacillus. From those, a 10% will probably develop an active disease the next decade. Nowadays, over 1 million people die annually because of an active TB.
The mechanisms that maintain a latent infection for a few years or that make it evolving towards an active disease are not fully understood, yet. In a previous work, the dynamics of TB lesions during an active disease in mice was described by an Agent-Based Model (ABM). This model accounted for the growth, coalescence and proliferation of lesions, showing that the most important mechanism for lesions growth during the active disease was coalescence. In a later work, the dynamics of lesions during a latent infection in minipigs was tackled by implementing a revised version of the previous ABM into a computational model of the bronchial tree. The model was fed with Computed Tomography scan data from latent infection in minipigs. In this case, the model showed that the proliferation of lesions through the bronchial tree was essential for maintaining the latent infection.
In this Master thesis we propose a first approach on the evolution of a latent tuberculosis infection into an active disease. The parameter space will be explored trying to elucidate which is the role of each mechanism on the trigger for the disease.

I. INTRODUCTION

A. Tuberculosis

Tuberculosis (TB) is a infectious disease that on 2014 killed nearly 1.5 millions of humans [1]. The same year, 9.6 million people developed TB. TB is caused by the bacillus *Mycobacterium tuberculosis* (*Mtb*). In fact, World Health Organization [1] estimates that one third of worldwide population is already infected with *Mtb*, and that a 10% of these infected people will develop an active TB disease in a few years.

Natural history

TB infection starts when *Mtb* arrives at a pulmonary alveolus and it is phagocyted by an alveolar macrophage. These bacilli can resist the bactericidal mechanisms induced by the macrophage and multiply inside the phagosome [2]. Finally, they cause macrophage necrosis and thereby enter the extracellular milieu, where they are phagocyted by another macrophage which also fails to control the bacillary growth and is likewise destroyed. This cycle ideally ends once the specific immune response appears and the TB lesion is controlled and calcified. According to the Dynamic Hypothesis [3], there is a certain probability that few bacilli escape from the lesion and start a new infection in other alveolus. This is known as the endogenous reinfection process, and takes place through the bronchial tree (figure 1).
If the host is immunocompetent, in most of the cases lesions growth is controlled and they are finally encapsulated. This host is said to be infected but not ill because it does not show symptoms and can not transmit the disease. According to the Dynamic Hypothesis, there is a certain probability that the endogenous reinfection affects the pulmonary upper lobe. In this location, a process of liquefaction occurs in the intragranulomatous

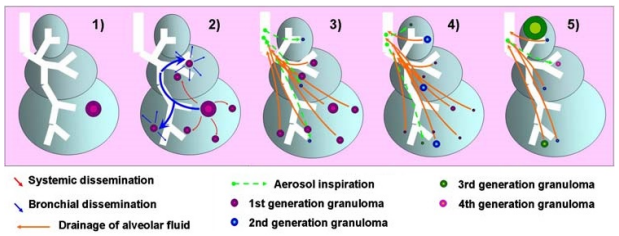

FIG. 1: Latent tuberculosis infection and the generation of active TB according to the Dynamic Hypothesis. Once the initial lesion has been generated (1), bronchial (blue arrows) and systemic (red arrows) dissemination generate new secondary granulomas (2). This process is stopped once the specific immunity has been established, which starts a constant drainage of non-replicating bacilli towards the bronchial tree (solid arrows) to which the inspired aerosols (dotted arrows) can return, thereby generating new granulomas (3, 4). This process implies finding different generations of granulomas simultaneously. In this dynamic process, if one of these reinfections takes place in the upper lobes, it will have the opportunity to induce a cavitary lesion (5). Adapted from [3]

*Electronic address: marticatalasabte@gmail.com

necrosis (or caseum) triggered by the macrophages, the bacilli, or both, which favors the extracellular growth of bacilli. This process finally results in bronchial erosion and drainage of the liquefied material, which leads to the formation of a cavity in the lung. This factor is not that crucial in immunosuppressed hosts, where the reactivation that occurs in different locations and cavitation is less frequent due to the lack of a sufficiently strong inflammatory response.
The success on the control of the lesion depends on a correct equilibrium between the inflammatory response, which promotes the growth of the lesion and the immune response, which controls and stops its growth. If one of these two responses have an incorrect behavior the infection can evolve towards disease.

B. Previous work

This research is the continuation of a previous project [4]. A computational model was built in order to simulate the evolution of TB lesions through a minipig bronchial tree fed with experimental data. We are now going to summarize the previous work done: how the experimental data were obtained and the two models that were joined (Bubble model and bronchial tree model).

Experimental data

Pig's anatomy and physiology have many features in common with humans. In particular, pigs lungs and immune system are very similar to the humans, so the evolution of TB in these animals should be similar to the dynamics in humans. In an experiment carried out at Centre de Recerca en Sanitat Animal by Unitat Tuberculosi Experimental, minipigs, a breed of pig developed and used for medical research, were used as an animal model for carrying out an experiment with three vaccine candidates.
For this study 24 minipigs were infected with a strain of *Mtb* and, after 12 weeks of infection, they were euthanized. During this 12 weeks of infection 18 minipigs received treatment to compare the efficiency of the different vaccines candidates. After being euthanized their lungs were analyzed with a Computer Tomography Scan (CT) in order to determine size, position and density of TB lesions. These data were used to fed the mathematical model. All experimental data obtained are analyzed on [4].

Minipig bronchial tree model

The design of a 3D model of the minipig bronchial tree required anatomical information about lungs. Data about the size and shape of the minipigs' lungs were obtained from the reported experiments. The size of each pair of lungs was determined using the maximum coordinates obtained with CT measurements. The corresponding images were used for setting the shape of one specific pair of lungs. Regarding minipigs bronchial tree, there is still very few information, but minipig anatomy is similar to humans. Then, we considered that its bronchial tree could be modeled as a human bronchial tree having into account that their size and surfaces are different. Therefore, we built a bronchial tree inside the computed surface using a set of rules that were developed for simulating a human bronchial tree [5], with the appropriate re-dimensioning [4]. 21 bronchial trees were successfully simulated and the results were analyzed concluding that they were reasonable [4].

Bubble model

Bubble model is a mathematical model that aims to describe the evolution of the TB lesions from an initial infection. It was initially designed for studying an active TB disease in mice [6]. It is an Agent-Based Model (ABM), where each lesion is an autonomous unit that can perform some actions.
Each lesion is modeled as a sphere with a certain radius (r_i, in mm, variable), spatial coordinates ($\vec{x_i}$, in mm, constant) and age (a_i, in days, variable). The rules that drive a lesions' dynamics are:

- **Growth**: the lesion grows due to cells accumulation. The inflammatory response is the one that causes the exponential growth of the lesion and the immune one stops the lesion growth and promotes its calcification. As lesions growth was supposed to be related with its surface we modeled it as a logistic in surface:

$$\frac{dr}{dt} = vr\left(1 - \left(\frac{r}{r_{max}}\right)^2\right) \tag{1}$$

- **Coalescence**: when two lesions get close enough ($\max\{r_1, r_2\} > \|\vec{x_1} - \vec{x_2}\|$), both lesions merge into one. The volume is assumed to be conserved. Then:

$$r_{new} = \sqrt[3]{r_1^3 + r_2^3} \qquad r_{max,new} = \sqrt[3]{r_{max,1}^3 + r_{max,2}^3} \tag{2}$$

$$\omega_i = \left(\frac{r_i}{r_{new}}\right)^3 \qquad i = 1, 2 \tag{3}$$

$$a_{new} = \omega_1 a_1 + \omega_2 a_2 \qquad \vec{x}_{new} = \omega_1 \vec{x_1} + \omega_2 \vec{x_2} \tag{4}$$

- **Endogenous reinfection**: The probability that a macrophage escapes from an alveoli and arrives into another one forming a new lesion depends on two facts: the size of the mother lesion (more bacilli, more probability) and its age (the greater the calcification, the lower the probability). Therefore, we modeled the probability of a macrophage to escape and form a new lesion (P_i) as:

$$P_i(t) = \rho R_i(t) \xi_i(t) \tag{5}$$

Where ρ is a constant to be fitted, $R_i(t)$ models the dependence on lesion's size and $\xi_i(t)$ models the dependence on lesion's age. $R_i(t)$ assumes a linear relationship with $r_i(t)$:

$$R_i(t) = \frac{r_i(t)}{r_{min}} \tag{6}$$

As there is lack of dynamic information about the calcification process, simple model was assumed taking into account that most of new lesions are generated in the period 14 $days < a_i < 28$ $days$: Where $r_i(t)$ is the lesion's radius at time t and r_{min} is the minimum radius of the lesions, $r(t_{min}) = r_{min}$.

$$\xi_i(t) = \begin{cases} 2 - \frac{a_i(t)}{14} & \text{if } 14\ days < a_i(t) < 28\ days \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The alveolus where the new lesion appear depends on the distance trough the bronchial tree between the one from which the macrophage escapes (mother lesion) and the alveolus where the new lesion is formed (daughter lesion). In fact, we modeled the probability to go from alveoli i to alveli j as:

$$P_{i\to j} = \frac{e^{-\beta d_{ij}}}{\sum_k e^{-\beta d_{ik}}} \tag{8}$$

Where β is a parameter that determines the mean distance where the new lesions will appear ($\bar{d} \propto \beta^{-1}$), and d_{ik} is the distance though the bronchial tree between i and k alveoli.

C. Objectives

In the previous work, the bubble model was successfully implemented into the virtual bronchial tree. After a careful calibration, experimental results of the control group (non-treated infected minipigs) were correctly reproduced.
The objectives of this project are:

- To explore the effect of different configurations of the initial infection
- To perform a sensitivity analysis and examine dependence between parameters.
- To delimit ranges of involved parameters for a latent infection and for an active disease
- To reveal the mechanisms that may cause a latent infection to divert towards an active TB disease.

II. THE MODEL

A. Model's update

Although the previous model correctly reproduced experimental results, some parts were modified in this project in order to better account for the underlying physical and biological processes:

- Modify the endogenous reinfection probability in order that the lesions can generate new lesions at $a_i >$ 28 $days$.
- Introduce differences in new lesions spreading according to the differences in breathing amplitude.

This is a first necessary step towards a potential mechanistic model of TB lesions dynamics in human lungs.

Endogenous reinfection

The previous model [4] was good enough to reproduce the experimental results, but it was too restrictive. In fact, the probability that a new lesion is formed after the 28th day is no zero strictly and it must be extended trough all time, since the probability that an infected macrophage escapes and forms a new lesion is always non zero. Due to the good results obtained with the previous model we looked for a similar model with a small non-zero probability for $a_i > 28$ $days$. It is modeled as a decreasing exponential:

$$\xi_{new}(t, n, \alpha) = e^{-\alpha(t-t_{min})^n} \tag{9}$$

Where α and n are positive constants that must be fitted by imposing two conditions:
-The area under the initial curve ($\xi(t)$) and the new one ($\xi_{new}(t)$) must be conserved:

$$\int_0^{+\infty} \xi(t)dt = \int_0^{+\infty} \xi_{new}(t, n, \alpha)dt \tag{10}$$

$$\alpha(n) = \left(\frac{\Gamma(1/n)}{7n}\right)^n \tag{11}$$

-The difference between both curves must be minimized: Defining the difference between both curves as:

$$Error(n) = \int_{14}^{28} dt\,[\xi(t) - \xi_{new}(t, n, \alpha(n))]^2 \tag{12}$$

We have to solve:

$$\int_{14}^{28} dt \frac{d}{dn}\,[\xi(t) - \xi_{new}(t, n, \alpha(n))]^2 = 0 \tag{13}$$

It was solved numerically using extended Simpsons rule for the integral with $h = 10^{-6}$ and Newton's method for the iterative process. On the Newton's method we used centered difference with $\Delta n = 10^{-6}$ to compute its derivative. We obtained $n = 1.63483 \pm 0.00001$ and $\alpha = 0.034638 \pm 0.000001\ day^{-n}$.
The previous model and the updated one are shown in Figure 2. The new one shows a tail that provide a non-zero probability for $a_i > 28$ $days$, but for $a_i > 35$ $days$ is

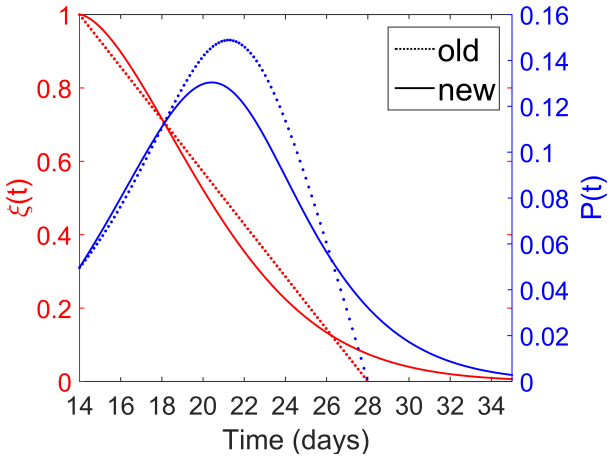

FIG. 2: Old and new endogenous reinfection probability. In red the age dependence ($\xi(a_i(t))$) and in blue the total probability $P_i t$). The old age dependence (red dashed line) was linear between 14 and 28 and the new one is the exponential $\xi(a_i(t)) = e^{-\alpha(a_i - t_{min})^n}$ with $\alpha = 0.034638$ and $n = 1.63483$.

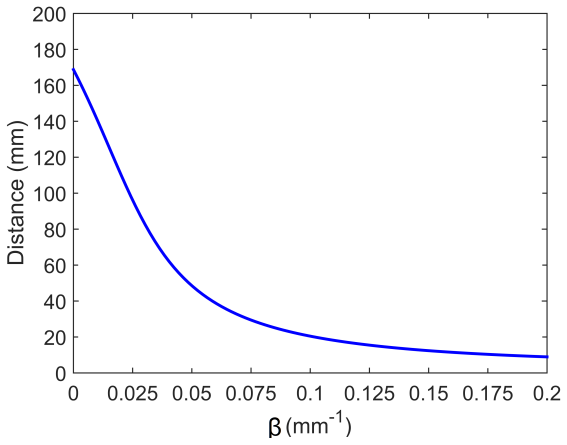

FIG. 3: Mean distance where new lesions appear depending on β. All distances are measured through the bronchial tree.

nearly zero.

β variable profile

In humans, active disease usually appears on the upper lobes. Although the model was developed for minipigs, we used it to test an hypothesis regarding this evidence. The breathing amplitude is not the same on every part of the lung. In fact, in humans we know that on the lower part of the lung the breathing amplitude is three times bigger than on the upper lobe [7].
Breathing amplitude determines the distance where new lesions may appear due to the endogenous reinfection process. At zones where the breathing amplitude is high the new lesions appear further and at zones where the breathing amplitude is low they appear closer.
Breathing amplitude is related with gravity and how the lungs are placed in the body. In humans the lungs are located vertically, so it is reasonable to say that the breathing amplitude is related with the vertical coordinate (Z). In our model the parameter that determines where new lesions appear is β (figure 3). Then, β was modeled to depend on Z coordinate as an arc-tangent profile (figure 4) that depends on four parameters:

$$\beta(z) = \frac{\beta_1 + \beta_2}{2} + \frac{\beta_1 - \beta_2}{2} \tan^{-1}\left\{\frac{\pi f}{\beta_2 - \beta_1}(z - z_0)\right\} \tag{14}$$

Where $\beta_1 = \beta_{min}$ is the minimum value of $\beta(z)$ at the lower part of the lung ($\lim_{z\to+\infty}\beta(z) = \beta_{min}$); $\beta_2 = \beta_{max}$ is the maximum value of $\beta(z)$, at the upper part of the lung ($\lim_{z\to-\infty}\beta(z) = \beta_{max}$); z_0 is the transition point where $\beta(z_0) = \frac{\beta_{min}+\beta_{max}}{2}$ and f is the absolute slope value at $z = z_0$, $\beta'(z_0) = -f$.
As said, we know that the breathing amplitude in the lower lobe is three times the breathing amplitude on the upper lobe. We assume that the mean distance where lesions appear is proportional to the breathing amplitude. Then, we want to find:

$$\frac{Breathing\ amplitude\ (upper\ lobe)}{Breathing\ amplitude\ (lower\ lobe)} = \frac{\overline{d}(\beta_{min})}{\overline{d}(\beta_{max})} = \frac{1}{3} \tag{15}$$

We impose that the mean value of β must be the one adjusted for the experimental case in minipigs (see II.B). Then, imposing that $\beta_{min} = 0.08 - \Delta\beta$ and $\beta_{max} = 0.08 + \Delta\beta$ we can find that in order to verify equation 15 $\Delta\beta = 0.03948$. We approximate $\beta_{min} = 0.04\ mm^{-1}$ and $\beta_{max} = 0.12\ mm^{-1}$.
z_0 is adjusted to be at $\sim 1/3$ of the lung. Then, we

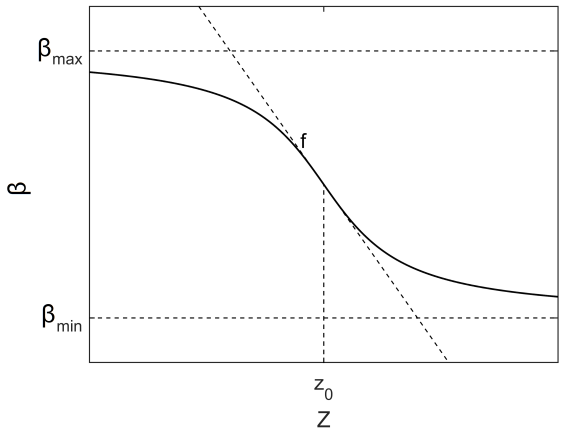

FIG. 4: β profile along the z axis of the bronchial tree. β follows an arc tangent function as seen in equation 14 that depends on four parameters: β_{min}, β_{max}, z_0 and f. β is a decreasing function that models the breathing amplitude of the lungs, the breathing amplitude is inversely proportional to β.

fit $z_0 = 25mm$. f is fitted to see a proper transition between the upper lobe and the lower lobe breathing amplitude, $f = 0.005\ mm^{-2}$

B. Model's setup

Input parameters

Our model depends on a set of input parameters, which determine the outcome of the simulations. Some of these parameters were given by the experimental data and other ones were adjusted to fit the experimental results, as described below:

Time of simulation, $\boldsymbol{T_{max}}$*:* time when the simulation ends and the lesions are analyzed as if there were a CT. It was set as $T_{max} = 84\ days$, like the experiment.

Number of initial lesions and distribution: we can choose one of the distributions that the model accepts as inputs giving the mean number of initial lesions and, in some cases, the deviation.

Mean maximum radius, $\boldsymbol{r_{max}}$*:* is the mean maximum radius that the non-merged lesions may achieve by a logistic growth. The maximum radius of a lesion follows a Gaussian distribution of mean value r_{max} and standard deviation $\sigma_{r_{max}} = 0.2\ r_{max}$. It is measured in mm. When we want to reproduce the experiment it is fitted to minimize an objective function and to accomplish, in mean, a number of desired lesions.

Natality, $\boldsymbol{\rho}$*:* a parameter that is proportional to the probability of trigger an endogenous reinfection process. It has no units. When we want to reproduce the experiment it is adjusted to accomplish, in mean, a number of desired lesions.

Growth velocity, $\boldsymbol{v}$*:* growth velocity of the lesions during the exponential part of the logistic function, it is measured in $days^{-1}$. It is adjusted to reach the 95% of the maximum radius of the lesion at around a_i=28 days, $r(28) \approx 0.95 \cdot r_{max}$ [4]. This yields to a default value of $v = 0.3\ days^{-1}$.

Calcification parameter 1, $\boldsymbol{\alpha}$*:* it is one of the two parameters used to model the growth of the coat around the lesion due to calcification process that prevent the lesions from escaping. It is measured in $days^{-n}$ and, as explained in section II.A, its default value is $\alpha = 0.034638\ days^{-n}$.

Calcification parameter 2, n: it is one of the two parameters used to model the growth of the coat around the lesion due to calcification process that prevent the lesions from escaping. It has no units and, as explained on II.A, its default value is $n = 1.63483$.

$\boldsymbol{\beta}$ *(constant model):* determines the mean distance where the new lesions appear. It is inversely proportional to the mean value where they appear. It is measured in mm^{-1}. If all traveling distances were likely probable the mean distance would be the inverse of β. When we want to reproduce the experiment it is adjusted to minimize an objective function.

$\boldsymbol{\beta_{min}}$ *(variable model):* minimum value for the β parameter, it is the value that it has on the lower part of the lung where the mean distance where new lesions appear is bigger. It is measured in mm^{-1}, its default value is set to $\beta_{min} = 0.04\ mm^{-1}$.

$\boldsymbol{\beta_{max}}$ *(variable model):* maximum value for the β parameter, it is the value that it has on the upper part of the lung where the mean distance where new lesions appear closer. It is measured in mm^{-1}, its default value is set to $\beta_{max} = 0.12\ mm^{-1}$.

$\boldsymbol{z_0}$ *(variable model):* transition point where $\beta = \frac{\beta_{min}+\beta_{max}}{2}$. It is measured in mm and by default it is set to $z_0 = 25\ mm$.

$\boldsymbol{f}$ *(variable model):* slope of the $\beta(z)$ function at the transition point, $z = z_0$. It determines how fast the β passes from its minimum value, β_{min} to its maximum value β_{max}. It is measured in mm^{-2}, its default values is set to $f = 0.005\ mm^{-2}$.

Initial infection

From experimental data we could see the visible lesions (diameter higher than 0.9 mm) after the 84 days of infection. We know that the initial infection was caused by 10^3 Colony-forming units injected trough the respiratory track but we do not know how many initial lesions were produced by these bacilli.

In order to determine which one of the observed lesions was the original one, we studied all the lesions and we compared them with the available data about the natural history. Nevertheless, we could not determine, which the initial lesions were [4].

The previous model had been calibrated assuming an initial single lesion at the mass center (mc) of the observed lesions. Then, we designed a set of virtual experiments to explore other possibilities, as detailed in Table I.We evaluated the error between each simulation and the experimental results as:

$$Error\,\{f_{sim}\} = \frac{\sum_{x,y,z} \frac{1}{n_{sep}} \sum_{i=1}^{n_{sep}} \left[f_{sim}(i) - f_{exp}(i)\right]^2}{\sum_{x,y,z} \frac{1}{n_{sep}} \sum_{i=1}^{n_{sep}} \left[f_{cm}(i) - f_{exp}(i)\right]^2} \quad (16)$$

Where f is the frequency of lesions observed on each bin. Unfortunately, none of these initial configurations satisfied our expectations (table I), so we established a new protocol for the initial infection.

We assumed that each minipig initial lesion(s) is one (or more) of the final lesions observed on it. We associate each observed lesion with a bronchial terminal. Then, we move backwards through the bronchial tree up to generation 5 (*ThresGen*), assigning the corresponding branch of this generation to each lesion. Typically all lesions of a minipig are assigned to a single 5th generation branch or to a couple of them. In the later case, a probability weight is assigned to each of the branches according to the volume of the associated lesions. At each simulation run initial lesion(s) will be randomly thrown from

these branches (taking into account the weight probability when appropriate) into possible associated terminal branches.
We had not any information about the number of lesions that formed the initial infection. Then, we incorporated to our model the possibility of having different number of initial lesions. The number of lesions considered for each minipig was tested as follows:
Delta: all minipigs have the same number of initial lesions, N.
Gaussian: each minipig can have a different number of initial lesions. The number of lesions follows a truncated Gaussian distribution with mean M and standard deviation S.
Poisson distribution: each minipig can have a different number of initial lesions. The number of lesions follows a Poisson distribution:

$$P(k|k>0) = e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \tag{17}$$

Outcome variables

The outcome variables for each simulation we computed were:
Number of lesions: number of observable ($d_i \geq 0.9\ mm$) lesions.
Mean diameter: mean diameter of the observable lesions, measured in mm.
Diameter standard deviation: standard deviation of the observable lesions, measured in mm.
Lesions volume: sum of the volume of all the observable lesions (the lesions are considered as spheres), measured in mm^3.
Occupied volume: volume occupied for all the lesions inside the lung. It is computed using a MATLAB function called boundary, it is measured in mm^3.
Geometric center: geometric center of all the observable lesions, it is computed for all three spatial coordinates, measured in mm.
Dispersion: mean square distance between the lesions and their mass center, measured in mm.
Illness Indicator: variable that indicates the state of each minipig. We considered that a minipig has a latent infection if the maximum diameter of the observed lesions is lower than 20 mm (II=0). A minipig has an active infection if the maximum diameter observed is higher than 20 mm but lower than 120 mm (II=1), and we consider that a minipig is dead if it has lesions larger than the lung dimensions 120 mm (II=2).
Time till illness: time that the minipig least to pass from a latent infection into an active one, measured in days.
Coalescences: number of merging processes occurred during the simulation.
Lesions in contact: number of lesions that are in touch with other lesions but that are not close enough to merge.
Number of sick lesions: number of lesions with diameter higher than 20 mm.
Coordinates of the sick lesions: mean coordinates of the lesions with a diameter higher than 20 mm, measured for the three spatial coordinates in mm.
Diameter histogram: histogram of the diameter observed on all the minpigs.
Positions histogram: histogram of the positions of the observable lesions.

Initial infection	X-Error	Y-Error	Z-Error	Total-Error
Coordinate center	0.812	2.055	1.000	1.277
Biggest lesion	2.423	2.826	1.477	2.241
Two biggest lesions	1.335	1.810	1.470	1.534
30% bigger lesions	0.944	2.004	1.232	1.383
Densest lesion	1.822	2.675	2.048	2.173
Two densest lesions	1.222	1.945	1.468	1.538
30% densest lesions	0.933	1.955	1.225	1.361
Density > 150HU	0.897	1.754	1.376	1.333
One random lesion	0.605	1.437	0.220	0.747
Two random lesions	0.630	1.091	1.839	1.179

TABLE I: Error for the different initial lesions configuration computed as equation 16.

Fitting r_{max}, ρ and β

In order to fit r_{max}, ρ and β we need three objective functions:

- Mean number of final lesions.
- Minimization of the error with the diameter histogram.
- Minimization of the error with the histogram of the lesions positions.

After some runs with different sets of parameters we saw that for low values of r_{max} the coalescence process is marginal $\left(\frac{1}{N}\sum_i N_{f,i} < 1\right)$. Then, the bronchial tree structure is not relevant for the mean number of lesions nor for the diameter distribution.
The dependence on the number of lesions (N_{obj}) can be considered as:

$$N_{obj} = N_i \cdot f\left(r_{max}, \rho\right) \tag{18}$$

Where N_i is the mean number of initial lesions. To check this approximation we performed simulations with $r_{max} = 0.5 : 0.05 : 1\ mm$, $\rho = 0.05 : 0.01 : 0.15$ and $N_i = 1 : 1 : 5$ and 1000 simulations for each of the cases. We obtained that the approximation was reasonably good because the mean error was 1.1%, the maximum error was 2.6% for $(r_{max}, \rho) = (0.7\ mm, 0.05)$ and the minimum error was 0.3% for $(r_{max}, \rho) = (1\ mm, 0.15)$. We also checked that we could not observe any correlation between the error and N_i, r_{max} and ρ.
In order to find $f(r_{max}, \rho)$ we performed simulations with $r_{max} = 0.5 : 0.05 : 1\ mm$ and $\rho = 0.05 : 0.01 : 0.15$ with 10000 run per each set. Then we fitted a function f with the form:

$$f\left(r_{max}, \rho\right) = a_{ij} \cdot \rho^i \cdot {r_{max}}^j \qquad i, j = 1 : 5 \tag{19}$$

The resulting surface obtained with Matlab fit function,

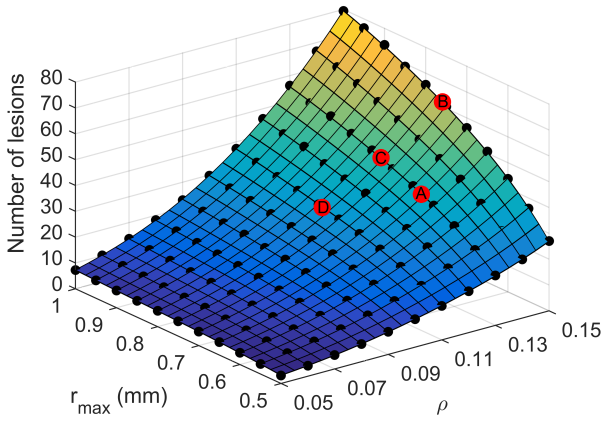

FIG. 5: Fitted surface for the number of observed lesions as a function of r_{max} and ρ for one initial lesion. The black points correspond to points obtained with 1000 simulations for different sets of ρ and r_{max}. The colored surface is the fitted surface (equation 19), with $a_{00} = 123$, $a_{10} = -1710$, $a_{01} = -719\ mm^{-1}$, $a_{20} = 390$, $a_{11} = 8540\ mm^{-1}$, $a_{02} = 1650\ mm^{-2}$, $a_{30} = 14500$, $a_{21} = -13200\ mm^{-1}$, $a_{12} = -13900\ mm^{-2}$, $a_{03} = -1900\ mm^{-3}$, $a_{40} = -492000$, $a_{31} = 218000\ mm^{-1}$, $a_{22} = -9800\ mm^{-2}$, $a_{13} = 11700\ mm^{-3}$, $a_{04} = 1060\ mm^{-4}$, $a_{50} = 1360000$, $a_{41} = -216000\ mm^{-1}$, $a_{32} = -76600\ mm^{-2}$, $a_{23} = 10300\ mm^{-3}$, $a_{14} = -4030\ mm^{-4}$ and $a_{05} = -223\ mm^{-5}$. The error is $R - square = 0.9999$ and $RMSE = 0.2064$. In red the points that correspond to the one lesion fitting for the different groups of minipigs (A,B,C,D) can be observed, table II.

is shown in Figure 5. The fitting error was: $R-square = 0.9999$ and $RMSE = 0.2064$.
Then, imposing a number of observed lesions, N_{obj}, and a number of initial lesions, N_i, we can obtain a relationship between r_{max} and ρ such as: $\rho = f(r_{max})$.
After performing some simulations we observed that the histogram of diameter distribution is not affected by variations of β, it only depends on r_{max} and ρ. Then, minimizing the error between the experimental histograms and the ones obtained from simulations and using the relationship obtained before ($\rho = f(r_{max})$) we can obtain a set of (r_{max}, ρ) that fulfills both conditions (number of observed lesions and minimization of the diameter histogram error). We defined the error between diameter histograms as:

$$Error\left\{D_{exp}, D_{sim}\right\} = \frac{1}{n_{sep} - 1} \sum_{i=1}^{n_{sep}} \left(D_{exp}(i) - D_{sim}(i)\right)^2 \tag{20}$$

Where $D_{exp}(i)$ is the number of lesions measured on the i bin of the experimental data and $D_{sim}(i)$ is the mean number of lesions measured on the i bin for all the simulations.
We tried different sets of (r_{max}, ρ) with r_{max} inside the measured experimentally radius. We could observe that $Error\left\{D_{exp}, D_{sim}(r_{max})\right\}$ had a parabolic shape near the minimum, then obtained this parabola and found its minimum. We computed the values of (ρ, r_{max}) that fulfilled the desired number of lesions and that minimized the diameters histogram.
With this methodology we could adjust the results of the different experimental groups (A,B,C,D), which are shown on figure 5 and on table II.
ρ and r_{max} were fitted using the first two constrains. Then, we fitted β by minimizing the error with the histogram of lesions positions. The objective function to minimize was defined as:

$$Error(\beta) = \sum_{x,y,z} \frac{1}{n_{sep}} \sum_{i=1}^{n_{sep}} \left(f_{sim}(i) - f_{exp}(i)\right)^2 \tag{21}$$

Where $f_{exp}(i)$ is the number of lesions measured on the i bin of the experimental data and f_{sim} is the mean number of lesions measured for all the simulations.
Simulations between $\beta = 0.01\ mm^{-1}$ and $\beta = 0.20\ mm^{-1}$ were done and the minimum error was observed at around $\beta = 0.08\ mm^{-1}$. In fact, it was observed that for $\beta \in [0.07, 0.09]\ mm^{-1}$ $Error(\beta) \approx constant$.

III. RESULTS

A. Simulation scheduling

Firstly, we carried out a study of the variability of our model because we aimed to know how many simulations should we perform to obtain, on average, reliable results. Then, we designed a simulation series on the parameters space with: $\rho = 0.125$, $r_{max} = 0.68mm$ and $\beta = 0.08mm^{-1}$.
It can be observed that the error of the mean of a given outcome, defined as:

$$Error(N) = 100 \frac{\sum_{i=1}^{N_T} \left(m_i(N) - \frac{1}{N_T} \sum_{i=1}^{N_T} m_i(N)\right)^2}{\sum_{i=1}^{N_T} m_i(N)} \tag{22}$$

is inversely proportional to the square root of the number of simulations $\left(Error(N) = \frac{A}{\sqrt{N}}\right)$. $m_i(N)$ is

N_i	param	Group			
		A	B	C	D
1	r_{max}	0.68 mm	0.76 mm	0.78 mm	0.79 mm
	ρ	0.13	0.15	0.13	0.11
2	r_{max}	0.67 mm	0.76 mm	0.76 mm	0.79 mm
	ρ	0.09	0.12	0.10	0.08
5	r_{max}	0.66 mm	0.74 mm	0.75 mm	0.78 mm
	ρ	0.06	0.08	0.06	0.05

TABLE II: Fittings for the different groups of minipigs. Fittings for the pair r_{max} and ρ that minimize the objective function (equation 20) for a given number of initial lesions (N_i), the other parameters were set to their default value.

the given outcome's mean value of a random sample of N simulations. Performing 5000 runs (i.e., 25000 simulations, since each run consisted of simulating the 5 control minipigs) we can compute the error of the mean of this outcome averaging over $N_T = 10000$ random samples of length N and compute the error for $N = 1, 2, 3, 5, 8, 13, 22, 36, 60, 100$. We can evaluate the value of the constant A using a linear regression technique. In table III we can observe the value of A for the different outcomes of the code, how many simulations are needed to have a mean deviation of the mean of around 5% and 1% and the estimated precision for 500 runs (2500 simulations). Ideally, we should perform a large number of runs to have an accurate estimation for the mean, but we must take also into account the computational cost. Therefore, we decided the number of simulations to be done trying to have a good enough precision and a reasonable computation time depending on the virtual experiment.

B. Parameter's space exploration

With the parameters adjusted to reproduce the experimental data, we could not observe minipigs with an active disease. Then, we designed a set of virtual experiments aimed to determine which are the parameters that may cause the development of an active disease. To do so, we performed a lot of simulations with different sets of parameters and we determined that the two parameters that mostly contributed to the ill lesions formation were r_{max} and β.
In order to observe the different zones with respect to latent infection or active disease, we performed several series of runs with different sets of parameters: $r_{max} = 1 : 1 : 10\ mm$, $\beta = 0.01 : 0.01 : 0.15\ mm^{-1}$ and $\rho = 0.045$. Other parameters were set to their default values. 1000 runs for each set were performed with the model of β constant along all the bronchial tree. On figure 6 we

	A	Number of simulations			Error(%)
		E=10%	E=5%	E=1%	
Number of Lesions	62.2	39	155	3872	1.2%
Mean diameter	10.3	1	4	106	0.2%
Diameters STD	29.4	9	35	886	0.6%
Volume of lesions	66.3	44	176	4400	1.3%
Occupied volume	92.1	85	354	8489	1.8%
Dispersion	59.6	36	142	3558	1.2%
Coalescences	118.2	140	560	13964	2.4%
Lesions in contact	225.6	506	2025	50637	4.5%

TABLE III: Error of the mean for the different outcomes. The value of A constant for each mean value can be seen as well as the number of simulations needed to achieve a certain error (10%, 5% and 1%) for the different outcomes. In the last column we can observe the error in % for 500 runs (2500 simulations). The error was computed using equation 22.

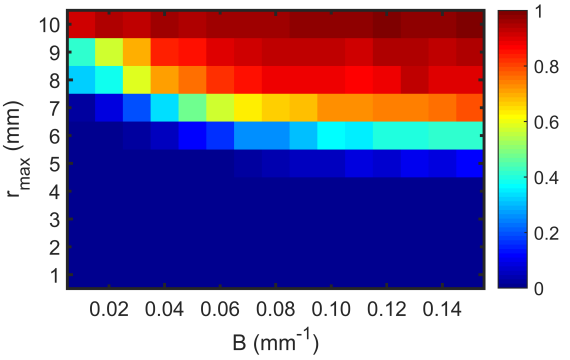

FIG. 6: Parameter space exploration to find the trigger of active disease. The illness appearance depends mainly on two parameters r_{max} (inflammatory response) and β (breathing amplitude). We quantified its dependence by performing simulations with different sets of parameters (r_{max},β) with $r_{max} = 1 : 1 : 10\ mm$, $\beta = 0.01 : 0.01 : 0.15\ mm^{-1}$ and $\rho = 0.045$. For each set 1000 simulations were done. The color of each square is proportional to the % of observed active disease's cases, see color bar at the figure's right.

can observe the obtained results. What we can conclude from them is that high inflammatory response is needed to cause a TB disease (high r_{max}) and that places with low breathing amplitude (high β) contribute to the illness appearance. In fact, what we observed was that a minimum r_{max} is needed to cause disease ($r_{max} \approx 5mm$) but it is not enough to cause illness by itself. Then, when the inflammatory response is not high enough to cause disease the trigger is the coalescence process that is favored by a low breathing amplitude (high β).

C. Sensitivity analyses

After the parameter space exploration we analyzed the effect of each input parameter on the outcome variables by means of a sensitivity analysis.
In fact, after analyzing figure 6 we designed two sensitivity analyzes. First, in the called **latent** infection zone where the probability to observe a minipig that developed an active TB (big lesions) is zero because the inflammatory response is not enough. This first set of parameters is necessary to observe the relative importance of parameters in the range of the experimental data. Secondly, a sensitivity analysis in the zone where the probability of a minipig to develop an **active** TB is about 50% (also called "'transition zone"'). This second set is necessary to observe the input parameters that determine the trigger of an active TB.
Latent set: $\rho = 0.125$, $r_{max} = 0.68\ mm$ and $\beta = 0.08\ mm^{-1}$.

	Outcome	Value	INPUT PARAMETER									
			T_{max}		r_{max}		ρ		v		β	
			+10%	-10%	+10%	-10%	+10%	-10%	+10%	-10%	+10%	-10%
LATENT	Mean number of lesions	28,6	+50%	-41%	+13%	-16%	+25%	-20%	+25%	-20%	-4%	+2%
	Mean diameter (mm)	1,37	+2%	+1%	+8%	-7%	+1%	0%	+1%	0%	+1%	-1%
	Mean volume lesions (mm^3)	45,9	+62%	-41%	+40%	-32%	+27%	-21%	+28%	-21%	0%	0%
	Mean occupied volume (cm^3)	18,0	+42%	-42%	+12%	-17%	+23%	-21%	+24%	-21%	-24%	28%
	Mean square dispersion (mm^2)	367	+7%	-12%	+2%	+6%	+6%	-8%	+3%	-7%	-16%	18%
	Mean number of coalescence	5,1	+103%	-53%	+20%	-17%	+53%	-34%	+40%	-27%	+22%	-18%
TRNSITION	Mean number of lesions	6,7	+13%	-11%	-11%	+13%	+8%	-10%	+18%	-19%	-14%	+14%
	Mean diameter (mm)	14,0	+2%	-9%	+15%	-13%	+2%	-1%	+5%	-3%	+5%	-4%
	Mean volume lesions (cm^3)	14,0	+29%	-28%	+33%	-27%	+18%	-17%	+42%	-32%	-4%	+2%
	Mean occupied volume (cm^3)	20,2	+25%	-25%	+11%	-10%	+15%	-16%	+37%	-31%	-18%	+22%
	Mean square dispersion (mm^2)	318	+9%	-8%	+1%	+1%	+1%	+1%	+1%	+1%	+1%	+1%
	Mean number of coalescence	7,1	+42%	-32%	+8%	-8%	+26%	-23%	+58%	-40%	+2%	-5%
	% of ill minipigs	60%	+17%	-21%	+30%	-42%	+15%	-17%	+28%	-33%	+5%	-8%

TABLE IV: Sensitivity analysis results. The effect of the increase or decrease (±10%) in different input parameters (T_{max}, r_{max}, ρ, v and β) for two different sets of parameters (latent and transition zone) is shown. The value obtained for the simulations with the default parameters is also shown.

Transition set: $\rho = 0.045$, $r_{max} = 6.50\ mm$ and $\beta = 0.08\ mm^{-1}$.

The sensitivity analysis was performed doing simulations with different set of parameters. For each simulation set, one of the input parameters was increased or decreased a 10% and we analyzed the output. We performed it for all the input parameters of the model. The results were analyzed by determining the variation (in %) of the outcome variables and by performing an analysis of variance (ANOVA) between the outcomes obtained from the simulation without changing any parameter and the one with a modified parameter. Table IV shows the variations of the mean in % of the latent sensitivity analysis. From the ANOVA test (not shown) and variation results we obtained that, for the latent infection, the number of lesions increases with time, with inflammatory response, with growth velocity and with natality. The breathing amplitude does not affect it. The mean diameter is mainly driven by the inflammatory response, but the coalescence process is not important. The lesions volume is, as expected, proportional to the mean diameter and the number of lesions. The occupied volume, dispersion and number of coalescences are increased when the number of lesions increases. The breathing amplitude affects the occupied volume, dispersion and coalescence process frequency. It does not affect the number of lesions or the mean diameter.

For the transition case, we obtained that when time increases, as seen on the latent case, the number of lesions, lesions volume, occupied volume, dispersion and number of coalescences also increase. In contrast to the latent infection case, the lesions diameter is also increased with the time because now the coalescence process is very important. The r_{max} effect is also very similar to the latent case excluding that it does not cause an increase in lesions. The reason is that bigger lesions coalescence is more probable and when the lesions merge the total number of lesions is reduced. It can also be seen that an increase in diameter is greater for the active case than for the latent one. Again, ρ and v have a very similar behavior as both increase the endogenous reinfection probability, what means an increase on most of the outcome variables as in the latent case. In addition the mean diameter is now also increased. The most relevant parameter when comparing the latent and the transition case is β. A decrease in it causes the lesions to be more separated; then, the merging process is not as frequent and the mean lesions diameter is reduced, as well as the number of ill minipigs. An increase in β causes the opposite effect. One thing that is a little bit surprising is that the square dispersion is not affected by variations of β. This is due to two effects that counter each other: on the one hand, an increase in β reduces the distance where new lesions appear and the dispersion is reduced; on the other hand, closer lesions increase the number of coalescences and reduce the number of lesions, which increases the mean dispersion.

D. Latent infection vs active disease

In order to compare the differences between the minipigs that developed an active TB and the ones that remained on a latent infection, we performed a large number of simulations using the set of parameters in the transition zone. On table V the differences between the minipigs that developed an active disease and the ones that did not is shown. The number of lesions in the minipigs that developed the active disease is higher than in the ones that did not (2,4 on average, an increase of 35%). We can also observe that the mean diameter is also higher but just a 8%, 1 mm. The deviation on the minipigs that developed an active TB is higher compared with the ones that did not, and the increase is higher than this increase in mean diameter, what means that the minipigs that developed an active TB had a

	Number of lesions	Mean diameter (mm)	Diameters deviation (mm)	Lesions volume (cm^3)	Occupied volume (cm^3)
Active TB	8,2	13,7	6,2	15,9	5,2
Latent TB	5,8	12,6	3,2	7,0	2,8
Mean	6,8	13,1	4,6	10,9	3,8

TABLE V: Differences between minipigs that developed an active TB and the ones that did not.

wider distribution.
The main difference between the minipigs that developed an active TB and the ones that did not is the number of coalescence process. In fact, in table VI we can observe that the minipigs without observed coalescence remained in an a latent TB and for large number of coalescence the minipig developed an active TB. The coalescence is the main cause of an active TB when the inflammatory response is not enough.
In order to observe the effect of the β variable model we performed 1000 runs (5000 simulations) with the variable β model. The set of parameters used to perform these simulations was: $\rho = 0.05$, $r_{max} = 6.5\ mm$, $\beta_{min} = 0.04\ mm^{-1}$, $\beta_{max} = 0.12\ mm^{-1}$, $f = 0.005\ mm^{-2}$ and $z_0 = -25\ mm$. We found that in 46% of the cases the minipig developed an active TB. Despite all the initial infections started on the lower lobe ($z_{ini} < -25\ mm$) most of the sick lesions (74%) were seen on the upper lobe ($z_{lesions} > -25\ mm$). This is in accord with usual clinical observation. The maximum number of sick lesions observed was 6, but mainly (in more than 50% of the cases) we only observed 1 sick lesion. It is observed that the growth of lesions is caused by the coalescence process. Then, when the breathing amplitude is small the coalescence is more probable. This experiment is the evidence that we needed to conclude that the breathing amplitude is a very important factor to cause an active TB. Despite the initial lesions are on the lower lobe, the probability of causing an active TB on the lower lobe is small and, if the infection arrives at the upper lobe (where the breathing amplitude is smaller), the probability of developing an active TB is high.

	Number of coalescence						
	≤ 2	3	4	5	6	7-11	≥ 12
% of ill	0	34	57	76	88	97	100
% of infected	100	66	43	24	12	3	0

TABLE VI: Number of coalescence effect on the illness rate.

IV. CONCLUSIONS

- The configuration of the initial infection (number of lesions and location) can not be induced from CT data. It strongly determines final location of lesions.

- The exploration of the $r_{max} - \beta$ parameter space revealed three zones with regards to the trigger of the disease. A small r_{max} ($< 5\ mm$) ensures the maintenance of a latent infection, while a high r_{max} ($> 10\ mm$) causes the development of the active disease. In the transition zone ($5\ mm < r_{max} < 10\ mm$) the key for triggering an active disease falls on β.

- The sensitivity analyses showed that, on the latent zone, lesions' individual growth parameters are relevant, while breathing amplitude only determines the spatial dispersion. There is a lack of coalescence. On the transition zone, a small breathing amplitude clearly increases coalescence, which becomes a determining mechanism for lesions to growth.

- According to the model, the mechanism that would cause a latent infection to divert towards an active disease are a high inflammatory response ($r_{max} \uparrow$) or a moderate inflammatory response together with a small breathing amplitude ($\beta \downarrow$). This is in accord with usual clinical observations.

[1] World Health Organization. Global tuberculosis report. Technical report, World Health Organization, 2015.
[2] Bermudez LE, Danelishvili L, Early J. Mycobacteria and macrophage apoptosis: complex struggle for survival. *Microbe* 1:372-375, 2006.
[3] Cardona P-J. Revisiting the natural history of tuberculosis. *Arch Immunol Ther Exp* 58:7-14, 2010.
[4] Català M. Modelling and simulation of tuberculosis lesions dynamics in a minipig bronchial tree. Bachelor's thesis, Universitat Politècnica de Catalunya, 2015.
[5] Vegué M. Model tridimensional de l'arbre bronquial humà per a l'estudi de la disseminació de *Mycobacterium tuberculosis*. Master's thesis, Universitat de Barcelona, 2012.
[6] Prats C, Vilaplana C, *et al.* Local inflammation, dissemination and coalescence of lesions are key for the progression towards active tuberculosis: the bubble model. *Front Microbiol* 7:33, 2016.
[7] Guo B, Xu XG, Shi C. Real time 4D IMRT treatment planning based on a dynamic virtual patient model: proof of concept. *Med Phys* 38:2639-2650, 2011.