When is it appropriate to model transmission of tuberculosis using a dose response model?

Duayne Strydom * **Ralf R. Küsel** * **Ian K. Craig** *

* *Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa*

Abstract: The risk of transmission of tuberculosis (TB) in confined spaces is analyzed using the Gammaitoni and Nucci model and its variant, a dose response model. The dose response model with its additional parameters has the benefit of taking the immune status and susceptibility of an individual into account, as a separate term from the generation parameter. It is shown via a sensitivity analysis and a model algebraic identifiability analysis that there is little benefit in using the more complex dose response model unless the number of infectious TB particles in the air can be measured. This is because the uncertainties associated with the additional dose response model parameters are otherwise lumped into the parameter representing the generation of infectious TB particles.

Keywords: Tuberculosis, risk of transmission model, quanta, risk in transmission control.

1. INTRODUCTION

Tuberculosis (TB) is a bacterial disease caused by *Mycobacterium tuberculosis* (*Mtb*) (WHO, 1999). The disease is most commonly an infection of the lungs, known as pulmonary TB, and is spread when droplet nuclei, containing *Mycobacterium tuberculosis* bacilli, are expelled from persons with active pulmonary TB, and subsequently inhaled by uninfected persons. It is estimated that one third of the worlds population is latently infected with TB and that approximately 5-10% of these people will become actively infected (WHO, 1999).

A quantum is defined as the number of droplet nuclei that would infect 63.2% of exposed individuals to that number of droplet nuclei. It is a method to quantify the infectiousness of airborne agents. The quanta unit can be seen to be a measure of both the infectivity as well as the quantity of the infectious agent in the room (Beggs et al., 2003). A dose-response (DR) model based on the Wells-Riley model makes use of infectious dose data to determine the probability of infection, where the dose is the number of pathogen required to infect a certain amount of the population. A 50% infectious dose is therefore the dose of pathogen required to infect half of the exposed population (Sze To and Chao, 2010; Nicas, 1996).

Because of the immunological response of each individual and the characteristics of the pathogen, it is impossible to directly measure quanta to determine the infectivity of the pathogen. This means that quanta is a theoretical unit of measure. Although the quanta is a theoretical unit of measure, it allows for mathematical comparisons between different scenarios of the risk of transmission, and the impact of environment control mechanisms.

In trying to refine the modelling of the risk of transmission, care should be taken that the factors used for this still indicate the uncertainty transparently (Issarow et al., 2016; Nardell, 2016). This paper investigates whether the use of infectious dose as a unit of measure would reduce the uncertainty surrounding the infectious particle generation parameter. The paper also investigates what measurements would be required for the uncertainty to be reduced.

2. TRANSMISSION MODELS

Disease models can be used to good effect to improve understanding (Craig et al., 2004), and also to study disease transmissibility (du Toit and Craig, 2015). For TB, the Wells-Riley model (WR) is the more commonly used model for the risk of transmission, but comparisons between the Gammaitoni and Nucci (GN) model showed that the models are fundamentally the same (Beggs et al., 2003). However, the GN model allows the use of non-steady state initial conditions of the infectious particles and the state-space format is ideal for identifiability analysis (Gammaitoni and Nucci, 1997; Xia and Moog, 2003). State-space models also allow for controller design for nonlinear and multiple-input, multiple output systems or the design of an observer that can estimate unknown or immeasurable states (Nise, 2011). In this paper, two models are compared. The difference between the two models is that in the second model the infectious particle source is changed to incorporate a measurable variable. The models that will be used are based on the GN model. The model consists of two states, which are the number of susceptibles $S(t)$ and the infectious particles $C(t)$ (Beggs et al., 2003).

$$\frac{dS}{dt} = -\frac{p}{V}CS \tag{1}$$

$$\frac{dC}{dt} = -NC + \gamma \tag{2}$$

In this model, p is the pulmonary ventilation rate of the susceptible population, V is the volume of the confined space in which the susceptible population is, and N is the air flow rate of the extraction of air from the room (measured in air changes per hour). The generation of infectious particles is given by γ. The room ventilation rate can be expressed as,

$$Q = NV \tag{3}$$

The generation of infectious particles, γ, determines the measurement of the infectious particles, C. If

$$\gamma = I\phi, \tag{4}$$

then the infectious particles will be measured in quanta. But if

$$\gamma = IG\beta, \tag{5}$$

where G is the number of airborne TB bacilli released per infector per unit of time and β is the fraction of infectious particles deposited in the alveolar region (Nicas, 1996), then the infectious particles are measured in infectious doses. The first model, the GN model, uses (4) to describe the generation of infectious particles, and the second model, the dose-response (DR) model, uses (5) for this purpose. Otherwise the two models are the same.

The DR model has the benefit of taking the immune status and susceptibility of an individual into account, via the β parameter in (5). This term is separate from the generation of infectious dose parameter whereas in the GN model, the generation and susceptibility is one term. It is of great importance that a realistic level of exposure in susceptible individuals is determined when using DR models. As with the GN model, interspecies extrapolation is also required for the DR model (Armstrong and Haas, 2007).

3. SIMULATION DATA

Simulation data was obtained from a published paper of an experiment conducted at the AIR (Airborne Infections Research) facility in eMalahleni, South Africa (Mphaphlele et al., 2015). In the paper, sentinel guinea pigs were used as the measure of the risk of transmission. The facility has six-bed inpatient wards that are connected by airtight ventilation systems to two identical guinea pig chambers (Mphaphlele et al., 2015).

For the models in (1) to (2), it was assumed that the susceptible guinea pigs and the infected individuals shared the same airspace, and that the guinea pigs do not add to the number of infectors. Constant air changes, N, of 6 per hour were assumed for the control simulations and 24 air changes per hour (ACH) were assumed for the intervention simulations (to verify the models) respectively and the number of infectors were cycled each day between two and zero infectors. The results from the published experiment are shown in table 1 (Mphaphlele et al., 2015).

It is assumed that only the number of infected can be measured through the diagnosis of TB disease. For simplification purposes, it will be assumed that the number of susceptibles is equal to the difference between the initial number of susceptibles and the number of infected. That is:

Table 1. Results showing the number of infected guinea pigs from AIR Facility experiment (Mphaphlele et al., 2015).

Period	**Control**	**Intervention**
Month 1	17 *guinea pigs*	4 *guinea pigs*
Month 2	31 *guinea pigs*	12 *guinea pigs*
Month 3	1 *guinea pigs*	0 *guinea pigs*
Total	49 *guinea pigs*	16 *guinea pigs*

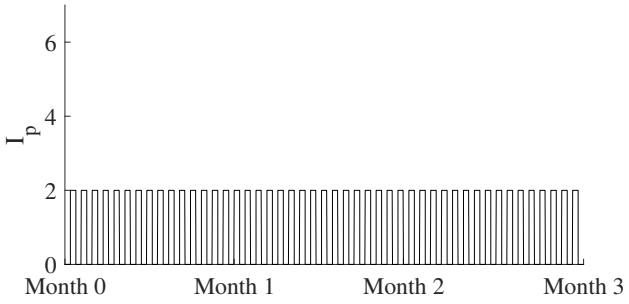

Fig. 1. Number of patients recorded in the ward.

$$S = S_0 - I \tag{6}$$

where the number of infected guinea pigs can be measured and the initial susceptible population is known. In this way, it will be assumed that the number of remaining susceptibles can be measured. For the following analysis, in which the two models are compared, it will be assumed that the number of infectious particles that is used by the DR model's infectious dose generation parameter, cannot be measured. The measured variable is therefore,

$$y(t) = S(t) \tag{7}$$

3.1 Simulations

Quanta and dose generation rates calculated using parameter estimation was applied to the GN and DR equations described in the previous section (Sze To and Chao, 2010). The data that was used is summarized in table 2.

Fig. 1 shows the assumed average number of patients in the ward. The number of patients were taken as zero when the clean (outside) air was directed into the animal rooms. This is done in order to simulate the switching of the airflow between the two animal rooms. Each patient was assumed to produce the same amount of quanta or infectious dose and each of the animals were assumed to have the same susceptibility. Fig. 2 shows the predicted number of infected guinea pigs compared to the measured number of guinea pigs for the control and intervention experiments.

The number of susceptible animals present in the ward remained constant at 90 guinea pigs. The ward has a paddle fan, with a calculated average of 60 ACH. Compared to the number of air changes extracted by the ventilation, the room air is assumed to be well mixed.

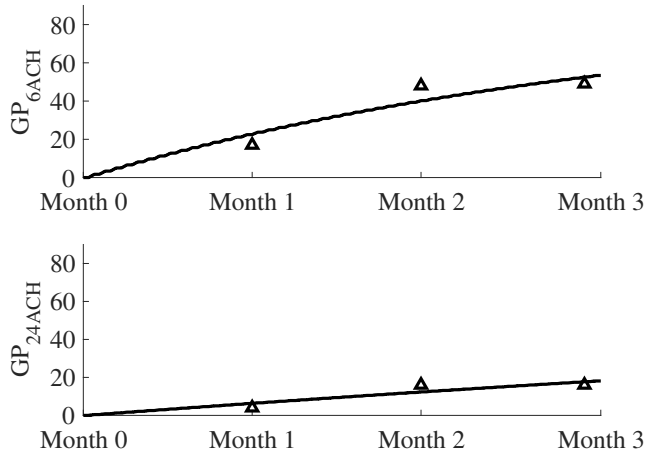

Fig. 2. Predicted number of infected guinea pigs for the GN and DR models. The solid lines show the predicted number of infected guinea pigs and the triangles the measured number of guinea pigs for the 6 and 24 ACH experiments.

Table 2. Equation parameters for GN and dose-response models.

Parameters	Values
S	90 *animals* (Mphaphlele et al., 2015)
ϕ	28.1560 $quanta \cdot h^{-1}$
G	46.9267 $dose \cdot h^{-1}$
β	0.6 (Sze To and Chao, 2010)
p	0.23 $m^3 \cdot h^{-1}$

The amount of quanta in the room is dependent on the air ventilation rate as well as the number of patients in the ward. The resulting predicted number of newly infected animals is shown in Fig. 2 for ventilation rates of 6 and 24 ACH.

3.2 Parameter Estimation

The infectious dose and quanta were backwards calculated from historic data (Mphaphlele et al., 2015). The quanta generation rate and infectious-dose generation rates were estimated using a Nelder-Mead search function to minimize a cost function, J. The cost function is used to minimise the difference between the simulation and measured number of infected animals, and penalties are also incurred if any of the model parameters fall outside the ranges specified in table 3. The cost function is shown in (8) in which a least squares formulation is used. A Heaviside function, $\mathcal{H}$, is used to implement the constraints on ϕ.

$$J = \sum_{i=1,2,3}^{k} (I(t_k) - I_k)^2 + \mathcal{H}(\phi_{min} - \phi)(\phi_{min} - \phi)^2 + \mathcal{H}(\phi - \phi_{max})(\phi - \phi_{max})^2) \quad (8)$$

where $I(t_k)$ is the simulated number of infected animals, I_k is the measured number of infected animals, and ϕ_{min} and ϕ_{max} are the minimum and maximum quanta generation rates. ϕ_{min} and ϕ_{max} are specified in table 3. The cost function is minimized when the quanta generation rate is 28.1560 $quanta \cdot h^{-1}$, with a final minimization cost value of $J = 103.1162$.

Table 3. Minimisation function ranges for the ϕ parameter.

Parameters	Values
ϕ_{min}	1.25 (Beggs et al., 2003)
ϕ_{max}	30840 (Beggs et al., 2003)

The infectious dose generation rate was calculated to match that of the quanta generation rate. The two models predicted the number of infected animals fairly closely, with a predicted value of 55 infected animals at the end of the experiment for both models, compared to actual measurement of 49 (see Fig. 2).

4. SENSITIVITY ANALYSIS

In order to determine the sensitivity to deviation of the model parameters, the simulation parameters were deviated by 10% from the values in table 2. The results are shown in Fig. 3 to 8, and the sensitivities for the GN and DR models are plotted separately.

Fig. 3 shows a deviation of a single patient instead of a 10% deviation, seeing that there cannot be a 10% increase in the number of patients. Fig. 4 shows the effect of a 10% deviation of the quanta and infectious dose generation rates. Fig. 5 shows the effect of a 10% deviation of the fraction of infectious particles deposited in the alveolar region. The sensitivity to a deviation of the pulmonary ventilation rate is shown in Fig. 6. The effect of a 10% deviation of the ward extraction ventilation rate is shown in Fig. 7. Fig. 8 shows the effect of a larger deviation of 30% in the ranges of the quanta and infectious dose generation rates.

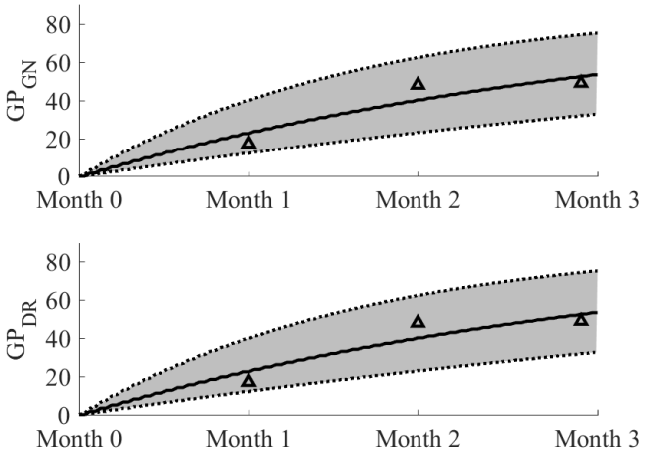

Fig. 3. Sensitivity to deviation of the number of infectors for the dose-response and the GN models.

The analysis shows that the two models have deviations of 10% in the predicted number of infected guinea pigs for the γ, G, β, p and N parameters (see Fig. 3 to 7). The number of infectors could not be deviated by a fraction and was deviated by a single patient. A large deviation

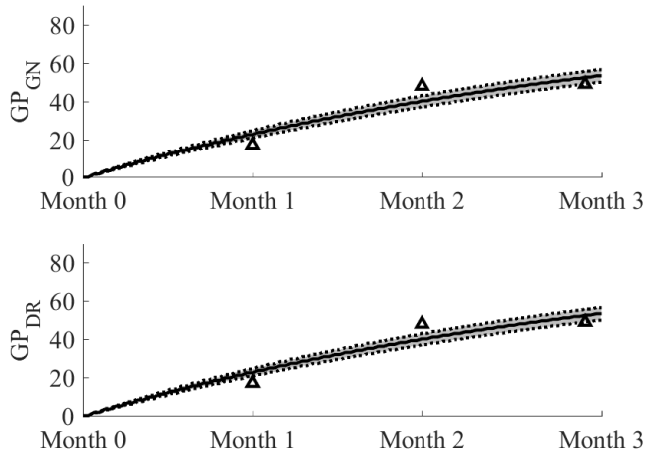

Fig. 4. Sensitivity to deviation of the quanta and infectious dose generation rates for the GN and dose-response models.

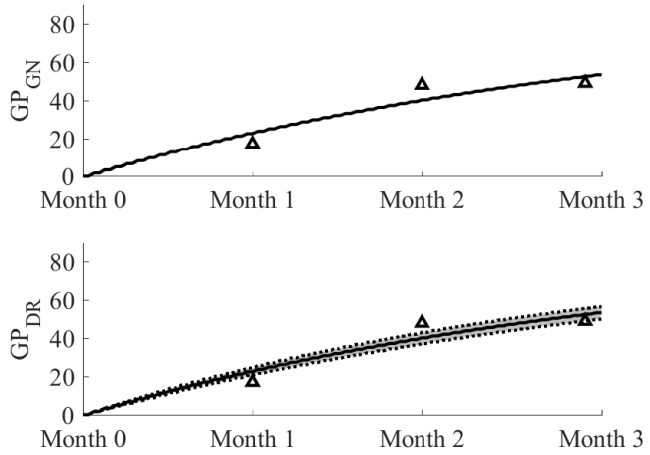

Fig. 5. Sensitivity to deviation of the beta parameter for the dose-response model.

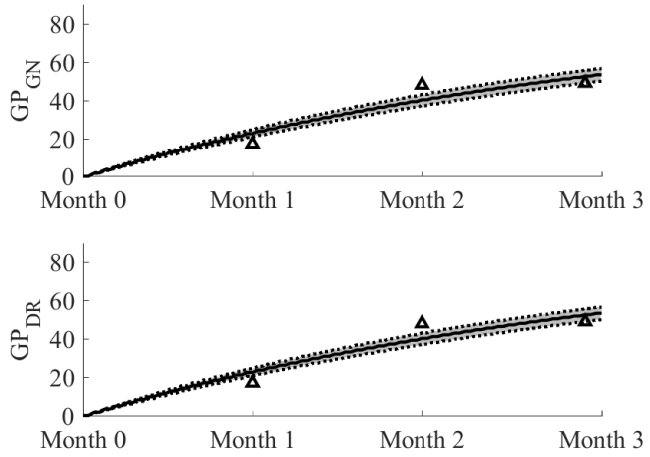

Fig. 6. Sensitivity to deviation of the pulmonary ventilation rate for the dose-response and GN models.

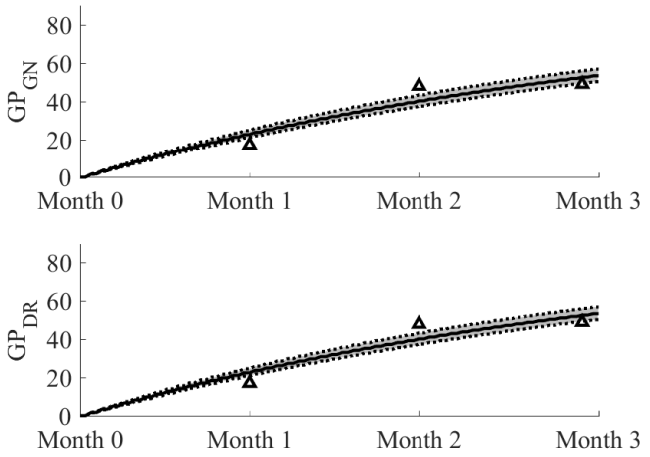

Fig. 7. Sensitivity to deviation of the ward ventilation rate for the dose-response and GN models.

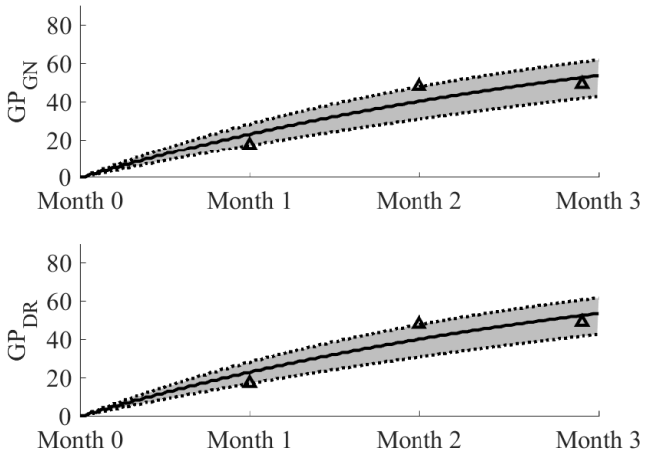

Fig. 8. Sensitivity to deviation of the range of the quanta generation rate for the GN and dose-response models.

of more than 40% was observed when the infectors were increased by one.

The quanta and infectious dose generation parameters ranges were also changed to deviate by 30% from calculated values, in order to determine whether the relationship is linear, resulting in a 15% increase in predicted number of infected. The non-linear relationship of the deviation in parameters and the predicted number of infected adds to the model uncertainty.

5. MODEL ALGEBRAIC IDENTIFIABILITY

The sensitivity analysis showed little variation between the two models concerning the uncertainty of the models. To investigate the difference in the identifiability of the model that using quanta or infectious doses makes, an algebraic identifiability method was used. Algebraic identifiability is a means to determine whether various parameters of an equation can be distinguished from output measurements given known inputs (Xia and Moog, 2003). The algebraic

identifiability gives a theoretical indication of what variables are independently identifiable in the model.

Since only the susceptibles state is measurable, the GN model must be brought into a form where the infectious particles state is removed. The number of susceptibles are measured using guinea pigs as sentinel animals.

$$\begin{aligned} \frac{dC}{dt} &= -NC + \gamma \\ e^{Nt}\left(\frac{dC}{dt} + NC\right) &= e^{Nt}\gamma \\ Ce^{Nt} &= \frac{\gamma}{N}e^{Nt} + K \\ C &= \frac{\gamma}{N} + Ke^{-Nt} \end{aligned}$$

The infectious particles at time zero are taken as $C(0) = C_0$ and is substituted into equations (1) and (2).

$$\frac{dS}{dt} = -\frac{p}{V}\left(\frac{\gamma}{N} + \left(C_0 - \frac{q}{N}\right)e^{-Nt}\right)S \tag{9}$$

For simplicity the most common case was considered, where the number of infectious particles are zero for the initial time frame considered.

$$\frac{dS}{dt} = -\frac{p\gamma}{VN}\left(1 - e^{-Nt}\right)S \tag{10}$$

To solve this algebraically, let

$$\theta = -\frac{p\gamma}{VN}\left(1 - e^{-Nt}\right) \tag{11}$$

and expressing the equation in terms of the measured variable gives,

$$\dot{y} = \theta y \tag{12}$$

The rank is calculated in order to determine whether the parameters are identifiable.

$$\begin{aligned} [y]\,[\theta] &= [\dot{y}] \\ rank\,[y] &= 1 \end{aligned}$$

Since the rank is the same length as the unknown parameter θ, this parameter is algebraically identifiable. From (12) it is evident that at least two measurements are necessary to identify the parameters.

Using this method of algebraic identifiability, it can be seen that only a single parameter of the model is identifiable from measured data. However, what this also reveals, is that any backwards calculation of any of the parameters would contain any of the deviations of any of the other parameters from the actual value.

For example, if the average pulmonary ventilation rate of the guinea pigs was 10% less for this situation compared to the value taken from the literature, and the source of infectious particles is backwards calculated, then the source of infectious particles would be estimated to be 10% greater than it actually is.

This implies that it is impossible to distinguish the difference of one parameter's uncertainty from another parameter's uncertainty.

Assuming that a method becomes available to measure the number of infectious particles, and following the same process, the measured variables become,

$$y_1(t) = S(t) \tag{13}$$
$$y_2(t) = C(t) \tag{14}$$

The model in equations (1) and (2) can be expressed in terms of these measured variables,

$$\begin{aligned} \dot{y_1} &= -\frac{p}{V}y_1 y_2 \\ \dot{y_2} &= -Ny_2 + \gamma \end{aligned}$$

To solve this algebraically, let

$$\theta_1 = -\frac{p}{V} \tag{15}$$
$$\theta_2 = -N \tag{16}$$
$$\theta_3 = \gamma \tag{17}$$

which gives,

$$\dot{y_1} = \theta_1 y_1 y_2 \tag{18}$$
$$\dot{y_2} = \theta_2 y_2 + \theta_3 \tag{19}$$

Equation 18 was used to determine the identifiability of θ_1.

$$\begin{aligned} [y_1 \quad y_2]\,[\theta_1] &= [\dot{y_1}] \\ rank\,[y_1 \quad y_2] &= 1 \end{aligned}$$

this is the same as the length of the vector containing the unknown θ_1, meaning that this unknown can be estimated from these measurements. As with the GN model identifiability, two measurements of y_1 and one measurement of y_2 is required.

Taking equation 19, the identifiability of θ_2 and θ_3 can be determined,

$$\begin{aligned} \begin{bmatrix} y_2 & 1 \\ \dot{y_2} & 0 \end{bmatrix}\begin{bmatrix} \theta_2 \\ \theta_3 \end{bmatrix} &= \begin{bmatrix} \dot{y_2} \\ \ddot{y_2} \end{bmatrix} \\ rank\begin{bmatrix} y_2 & 1 \\ \dot{y_2} & 0 \end{bmatrix} &= 2 \end{aligned}$$

this is the same as the length of the vector containing the unknowns θ_2 and θ_3, meaning that these unknowns can also be estimated from these measurements. However, in order to estimate these parameters, at least two measurements of y_1 and three measurements of y_2 are needed. This shows that three parameters are identifiable for the case where both the number of susceptibles and number of particles can be measured.

6. DISCUSSION AND CONCLUSION

Numerous confined-space models have been developed from basic SEIR (Susceptible, Exposed, Infected and Removed) models, most notably the WR and GN models

that describe the transmission of airborne diseases in confined spaces. The susceptibility of each individual is also assumed to be the same in each of the models, meaning each individual has the same immune response. A quanta generation rate of 28.1560 was backwards calculated using a least squares cost function and a Nelder-Mead minimization function. The simulations predicted 53 infected animals compared to the 49 measured.

Simulating each model's sensitivity to deviation in parameters showed no benefit in using the dose-response model when backwards calculating the infectious dose. The sensitivity to deviation is the same for both quanta and infectious dose cases, and in the case of the dose-response model, an increased uncertainty arises from the β parameter. In order to improve the accuracy of the dose-response model, measurements of the number of infectious TB particles are needed. The proposed sensor to measure the number of infectious TB particles would need to be accurate enough to measure the number of TB bacilli present in the air (or air sample). Using sensors for which the measured organisms are indistinguishable from TB, such as a particle counter, would result in the infectious dose parameter having the same uncertainty as the quanta parameter. In this case, when estimating the other parameters, increased uncertainty would arise due to each of the parameters being estimated from the infectious dose and number of susceptibles measurements.

The algebraic identifiability indicates that both the quanta and the dose models are identifiable. However, with only the measurement of the susceptibles available, it is only possible to solve for (11). This means that the effect of the different parameters (like p and γ) are indistinguishable from each other. This means that only one of the parameters that make up θ may be unknown. Any deviation in the actual value of the variables, that are set as fixed and derived from literature, will result in a compensating deviation of the variable that is solved for. As an example, if γ is chosen to be solved for, and all the other parameters in θ are chosen from literature, but the actual value of p, which is a value from literature, is 10% greater for this group of guinea pigs, then the solution of γ will be 10% less than the actual value of γ for this group. This implies the variable that is solved for has significant uncertainty.

If the measurement of the number of infectious particles is available, then the situation changes. This would allow the different parameters of the model to be estimated independently. This would greatly reduce the uncertainty 'lumping' onto the infectious particles generated per infector. Although the different parameters that define the make-up of γ are not differentiable from each other, the ability to determine γ independently means that the uncertainty of the model parameters is at least low enough where a theoretical unit of measure is of concern. If the measurement can be made in real time, optimal control solutions could be employed to possibly reduce the transmission of the disease.

In the opinion of the authors, the measurement of infectious TB particles would be required to justify using infectious doses instead of quanta of infection when modelling the risk of transmission of tuberculosis. The dose response model can be useful for simulation studies without such a measurement. However, there is little benefit in using the more complex dose response model when representing experimental data unless the number of infectious TB particles in the air can be measured. This is because the uncertainties associated with the additional dose response model parameters are otherwise lumped into the parameter representing the generation of infectious particles.

REFERENCES

Armstrong, T.W. and Haas, C.N. (2007). A quantitative microbial risk assessment model for legionnaires' disease: Animal model selection and dose-response modeling. *Risk Analysis*, 27, 1581–1596.

Beggs, C.B., Noakes, C.J., Sleigh, P.A., Fletcher, L.A., and Siddiqi, K. (2003). The transmission of tuberculosis in confined spaces: An analytical review of alternative epidemiological models. *International Journal of Tuberculosis and Lung Disease*, 7, 1015–1026.

Craig, I.K., Xia, X., and Venter, J.W. (2004). Introducing HIV/AIDS education into the electrical engineering curriculum at the University of Pretoria. *IEEE Transactions on Education*, 47(1), 65–73.

du Toit, E.F. and Craig, I.K. (2015). Selective pinning control of the average disease transmissibility in an HIV contact network. *Physical Review E*, 97(012810).

Gammaitoni, L. and Nucci, M.C. (1997). Using a mathematical model to evaluate the efficacy of tb control measures. *Emerging Infectious Diseases*, 3, 335–342.

Issarow, C.M., Wood, R., and Mulder, N. (2016). Seminal mycobacterium tuberculosis in vivo transmission studies: Reanalysis using probabilistic modelling. *Mycobact. Dis.*, 6(3), –.

Mphaphlele, M., Dharmadhikari, A.S., Jensen, P.A., Rudnick, S.N., Reenen, T.H.V., Pagano, M.A., Leuschner, W., Sears, T.A., Milonova, S.P., Walt, M.V.D., Stoltz, A.C., Weyer, K., and Nardell, E.A. (2015). Institutional tuberculosis transmission: Controlled trial of upper room ultraviolet air disinfection: A basis for new dosing guidelines. *American Journal of Respiratory and Critical Care Medicine*, 192, 477–484.

Nardell, E.A. (2016). Mycobacterial diseases wells revisited : Infectious particles vs . quanta of mycobacterium tuberculosis infection dont get them confused. *Mycobact. Dis.*, 6, 5–7.

Nicas, M. (1996). An analytical framework for relating dose, risk, and incidence: An application to occupational tuberculosis infection. *Risk Analysis*, 16, 527–538.

Nise, N.S. (2011). *Control Systems Engineering*. John Wiley & Sons, Hoboken, 6th edition.

Sze To, G.N. and Chao, C.Y. (2010). Review and comparison between the wellsriley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor Air*, 20, 2–16.

WHO (1999). World health organization–tuberculosis, http://www.who.int/topics/tuberculosis/en/, accessed 28 march 2016.

Xia, X. and Moog, C.H. (2003). Identifiability of nonlinear systems with application to hiv/aids models. *IEEE Transactions on Automatic Control*, 48, 330–336.