

Dynamical Models of Tuberculosis among Canadian Immigrants

Jeremy Chiu^{1*}, William Ruth^{2†}, Albert Wong^{1†}, Timothy Lee¹,
Kezia Wijaya¹

^{1*}Mathematics and Statistics, Langara College, 100 E. West 49th
Avenue, Vancouver, V5Y 2Z6, B.C., Canada.

²Statistics, Université de Montréal, 2900, boul. Édouard-Montpetit,
Montréal, H3T 1J4, PQ, Canada.

*Corresponding author(s). E-mail(s): jeremychiu@langara.ca;

Contributing authors: william_ruth@sfu.ca; alwong@langara.ca;
tle74@mylangara.ca; kwijaya01@mylangara.ca;

[†]These authors contributed equally to this work.

Abstract

This research developed a dynamical system model for the Canadian immigrants for the ten year period between 2011 and 2020.

Keywords: Tuberculosis, Dynamical Models, Canadian Immigrants

1 Introduction

We will write this section last

- Literature review – biology of TB, Canadian foreign-born population and screening and immigration (with TB), mathematics
- Model. Theirs (Guo & Wu) without any adaptations
 - Data
- Show problems with their model:
 - calculate incidence, it's very off [Fig2, top graph]
 - $R_0 = 0$ has problems.

- Optimization using fmincon, their parameters
 - R_0 same problem
 - we include L_0 as an optimization (estimated) parameter, but it's extremely dependent on our "initial choice" of L_0 . The optimized L_0 stays nearly identical to our initial choice
- Other graphs:
 - Sensitivity analysis for ν and ω (and many other parameters that don't really matter)
 - After sensitivity analysis, we inspect the combinations of parameters that yield low error
 - plot error vs (certain) parameter
 - post-sensitivity analysis, we have distributions of optimal q_1 , q_2 , E_0 , L_0 , R_0
 - how to pick L_0
 - (?) updating L_0 at every step causes problems in that optimizer no longer changes E_0
- other model updates:
 - arrows pointing from R to T

2 Literature Review

Background and TB in general

- WHO report on TB (number of cases) [?]
- Canadian numbers and recent studies
- Canada TB numbers [?] and [?]
- Canadian context [?] and [?] and [?]
- Canadian communities [?] and [?] [?]
- Canadian Immigrant population [?]
- TB Modelling SIIR Model for South Sulawesi [?] VSEIT Model for Algeria [?]
- SIS-SEIS Model for Indonesia [?] SEII model [?] SELT Model with early and late latency [?]
- Impact of immigrants [?]
- Reinfection Contribution of reinfection to ARI and incidence of TB [?] SEIRE Model [?]

2.1 Papers and work that heavily influence our work

- Guo paper on the basic model with early and late latency and without re-infection [?]
- TB Surveillance reports, which include data we use [?]

2.2 In 2008 to 2020, what countries did Canadian immigrants come from

See Wikipedia.

This shows that a significant proportion of Canada’s recent immigrants are from TB high-incidence countries, e.g. India [?], China [?], and Philippines[?].

2.3 [TIM] Screening Procedure in Canada (and other low-incidence countries)

In Canada, the primary detection method for active Tuberculosis disease is a series of diagnostic tests consisting of a chest radiography, followed by an MRI or similar scans, as well as a test for microbiological TB presence via sputum samples (Public Health Agency of Canada [PHAC],2014). The chest radiography involves scanning the lungs for abnormalities that have been observed in high proportions of previous confirmed cases of tuberculosis disease (Canadian Thoracic Society [CTS], 2019). For example, 56.2 percent of previously confirmed positive TB cases were observed to have varying degrees of cavitation in the upper lungs (CTS, 2019). Other abnormalities include, but are not limited to, calcified pulmonary modules, intrathoracic adenopathy, and fluid accumulation in the pleural cavity (CTS, 2019). These abnormalities are statistically *associated* with active pulmonary TB, but are not exclusively TB symptoms. (CTS, 2019). To err on the side of caution, all documented abnormalities should be looked for during screening, as to minimize false negative diagnoses. As a result, the chest radiography has a fairly high sensitivity estimated to be over 0.95 when considering all associated lung irregularities, but has a specificity as low as 0.75 (CTS, 2019). To improve specificity, individuals tested positive for the chest radiograph are required to undergo sputum testing to confirm the diagnosis (PHAC, 2014). MRI and other scans are also performed in conjunction to detect for extra-pulmonary tuberculosis and further increase specificity (CTS, 2019). Extra-pulmonary TB is assumed to have no mechanism for for transmission, and so is not transmissible from an infected host [?].

The tuberculin skin test (TST) and interferon gamma release assay (IGRA) are used to detect latent TB infection, but yield a positive result for both LTBI and active TB (CTS, 2019). TST is an *in vivo* examination in which tuberculin, proteins which contain *M. Tuberculosis* antigen, is injected into the skin of the testing patient. The localized region of skin is observed for reactions after 48-72 hours, which is indicative of prior infection (PHAC, 2014). The degree of swelling of the skin region is examined, with a ≥ 5 mm rise in the skin being considered the threshold for a positive latent TB diagnosis, while any changes under 5mm is generally considered negative (CTS, 2019). On the other hand, IGRA test is an *in vitro* test which involves injecting blood samples with *M. Tuberculosis* antigen (PHAC, 2014). Immune responses from adapted T cells will emit a protein known as Interferon-gamma (Tau & Rothman, 1999), which is a litmus for infection if detected in the blood sample above a certain threshold (≥ 0.35 IU/mL) (CTS, 2019). It should be noted that the thresholds for both tests were chosen to balance specificity and sensitivity, but can be raised or lowered to maximize one over the other (CTS, 2019). For example, if a rise of skin over 10 mm were considered a positive diagnosis in the skin test rather than 5mm, then there would be less false positive diagnoses, but more false negatives as a result (increasing specificity for a decrease in sensitivity) (CTS, 2019). The thresholds used in TST and IGRA are loosely standardized, and may vary by medical practitioner (CTS, 2019). Furthermore, immunizations have been shown to cause false negatives (PHAC, 2014),

and other diseases (including active TB disease) can cause false positives in LBTI tests (CTS, 2019). As a result, both TST and IGRA tests have relatively low sensitivity as well as specificity (CTS, 2019).

According to *Canadian Tuberculosis Standards 7th Edition: 2014*, all permanent migrants are required to be undergo chest radiography to be screened for active tuberculosis disease. Any positive tests must undergo further diagnosis and necessary treatment before allowed entry (PHAC, 2014). For visitors /temporary migrants, individuals who are on extended stays of over 6 months, come from high incidence countries with over 30/100000 TB incidence rates, or have healthcare-related occupations are also screened for active TB pre-entry (PHAC, 2014). LBTI screening is not performed en masse for migrant groups pre-entry, and is instead performed on *select* migrants who are at higher risk to have been exposed to TB infection (PHAC, 2014). The incidence rate in the country of origin, proximity to active cases, and occupation are factors used to assess risk and selection for LBTI screening (PHAC, 2014). Those who are tested positive for LBTI are still allowed entry, but are subject to medical surveillance (PHAC, 2014). As of the 8th edition of the Canadian Tuberculosis Standards, LBTI tests are still discouraged for mass screening migrants due to its cost:benefit ratio, and reserved for high risk travellers (CTS, 2019).

-On average, Canada receives 280 000 permanent resident arrivals and 380 000 temporary resident arrivals per year, with 70 percent of entries being from high incidence countries (PHAC, 2014). In 2011, of the 500992 pre-entry medical tests, 0.09 percent resulted in positive for active TB disease (PHAC, 2014)

-only 69 percent of arrivals were tested for LBTI before *or* after entry. This lack of mass screening for LBTI, as well as the lack of accuracy in the LBTI tests described earlier, are likely the main reasons why data regarding migrants w/ LBTI is not well-documented, and is the main crux of this paper.

2.4 TB Reactivation - relapse and reinfection [Kezia, add to this]

Patients who develop active Tuberculosis more than once in their lifetime can be classified into one of two categories – exogenous reinfection, or endogenous relapse / reactivation. Cases of relapse are usually linked with drug resistance, commonly found within the first 2 years after treatment ends [?] [?]. Two Tuberculosis treatment trials conducted in the United States and Canada found that 96% of recurrent Tuberculosis cases were from reactivation of the initial infecting strain[?][?], see also [?]. Although some sources claim reinfection is more common than expected [?], the proportion of recurrent TB cases due to relapse or reinfection is highly associated with the respective country. Cases of TB reinfection are often linked with the coexistence of HIV [?].

[?] “We observed a saturation effect in the number of cases per household and estimated that protective immunity conferred up to 35% reduction in the risk of disease.”

[?] “the protection afforded to hosts against a new infection from an established one, ... in a population are hard to estimate and cannot be observed directly.”

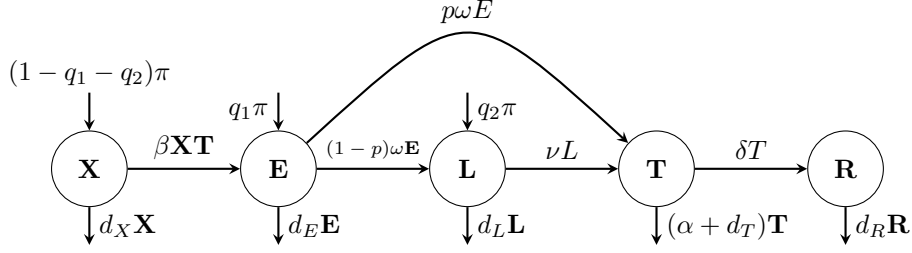


Fig. 1 Flow diagram for our model of TB transmission.

3 The Compartmental Model

[?] present a sophisticated compartmental model for the spread of tuberculosis (TB) within the foreign-born population¹ of Canada. We adopt the structure of their model, but update their parameter values based on more current data. In this section, we first present the model of [?], then give our updated parameter values. We were unable to locate reasonable values for some parameters. We discuss our approach to these unknown parameters in Section 3.2.

The model used by [?] to model TB transmission divides people into five compartments based on their disease status: Susceptible (**X**), Early Latent (**E**), Late Latent (**L**), Active (**T**) and Recovered (**R**). In brief, individuals begin their life susceptible. A person may become infected upon contact with an active TB case. Immediately after infection, an individual moves to the early latent stage (the *latent period* is defined as time from infection to infectious

[awk] All the assumptions give rise to the following system of differential equations.

$$\begin{cases} X' &= (1 - q_1 - q_2)\pi - \beta X T - d_X X, \\ E' &= q_1 \pi + \beta X T - (d_E + \omega) E, \\ L' &= q_2 \pi + (1 - p)\omega E - (d_L + \nu) L, \\ T' &= p\omega E + \nu L - (d_T + \alpha + \delta) T \\ R' &= d_T - d_R R \end{cases} \quad (1)$$

Figure 1 gives a diagram of our compartments, along with arrows and rates for all possible transitions. Table 1 gives more detail on each of the parameters in our model, and Table 2 gives the initial conditions used for our model. Note that some parameters and initial conditions are pulled from the literature, while others are estimated from the data.

See [?] for a detailed stability analysis of this model.

¹Depending on the source, there may or may not be a difference between the terms “foreign-born” and “immigrant”. In practice, the operational definition must be checked for each source, regardless of which term is used. In this work, we use the terms interchangeably, and define them both to mean “something” (we need to be careful and precise about this definition. We may actually not be able to get away with a single definition, in which case we can mention the inconsistency of what is being measured as a limitation of our study in the Discussion section).

Parameter	Description	Value
π	Annual Immigration Rate	reported [?]
q_1	Proportion of Early Latent Cases Among Immigrants	Estimated
q_2	Proportion of Late Latent Cases Among Immigrants	Estimated
β	TB Transmission Rate	
d_C	Death Rate for Compartment C	
ω	Rate of Progressing Past Early Latent Stage	
p	Fraction of Cases Which Skip Late Latent Stage	
ν	Rate of Progressing Past Late Latent Stage	
δ	Recovery Rate from Active TB	

Table 1 Parameters of our model, as well as their specified values and corresponding references. Parameters for which we were unable to find suitable references are labelled as “estimated”.

Compartment	Description	Value
\mathbf{X}_0	Susceptible	
\mathbf{E}_0	Early Latent	Estimated
\mathbf{L}_0	Late Latent	Estimated
\mathbf{T}_0	Active TB	Reported
\mathbf{R}_0	Recovered	

Table 2 Initial conditions for compartment sizes (i.e. size in 2010). Values are either given (with references) or are labelled as “estimated”.

3.2 Biology and Parameters for Simulation

3.2.1 Infectivity β

[?] use $\beta = 1 \times 10^{-8}$, while [?] use $\beta = 7 \times 10^{-6}$.

3.2.2 Latent Tuberculosis

Parameters that relate to the E and L compartment. q_1 , q_2 , p , ω , ν .

[?] uses $\nu = 0.00256$, which corresponds to 5 percent probability of development of disease over 20 years during the long-term LTBI stage.

Many TB infections remain latent and never develop to active TB – [?] estimate only 5-10% of latently infected individuals develop active TB.

[?]: “The mean waiting time between first and second diagnoses in households with 7 adults was 3.5 years, and the distribution was skewed toward shorter times, with a median interval between cases of 1.65 years”

Newly infected patients (≤ 2 years from infection) are $15\times$ more likely to develop active TB than people with no known risk factor [?] (note the 2022 Standards [?] do not report this number ; hence we use the 2014 Standards [?] reported number). Thus

$$p\omega = 15 \cdot \nu$$

[?] report $p = 5\%$, with a range of $2\% - 15\%$.

ReportedImmigration = [xx259110260036263101267924240763323192272707303325313601284157226309]

4 Analysis

The parameters of our model are broadly divided into two groups: those obtained from the literature, and those we estimate. Our analysis treats these groups separately. We begin by fixing all unestimated parameters at plausible values. We then optimize over the estimated parameters, calibrating our model to match observed TB incidence. Finally, we investigate sensitivity to our choices of unestimated parameter values by repeating the above process with several different values for each unestimated parameter (awk). In this section, we describe our treatment of the unestimated and unestimated parameters in more detail, and discuss what outputs we extract from our individual simulations and sensitivity analysis.

4.1 Estimation

Citations needed for methodology. I have some I can use for optimization, but it's worth looking into how Matlab likes to be cited.

We begin with the estimated parameters, q_1 , q_2 , \mathbf{E}_0 and \mathbf{L}_0 . For fixed values of the non-estimated parameters, we estimate q_1 , q_2 , \mathbf{E}_0 and \mathbf{L}_0 by minimizing the least-squares error for the observed incidence rate of TB among foreign-born individuals in Canada between the years of 2010 and 2020 (confirm this). More precisely, for particular values of the non-estimated parameters, we use the `ode23` function in `Matlab` (citation for Matlab? for ode23?) to solve System 1. This solution consists of trajectories for the sizes of the compartments in our model. We then compute the TB incidence at each time point using the expression in ?]: $pwE + vL$. Our objective function is the squared \mathcal{L}^2 difference between predicted and observed incidence at each time point where we have data (normalization?).

Optimization is performed using the `fmincon` function in `Matlab`. This function iteratively attempts to take a “direct step”, in which the KKT equations are approximated using the finite difference method, then linearized and solved. Alternatively, if the Hessian obtained from the finite difference method is not positive definite, the Conjugate Gradient method is used instead for the current iteration.

Figure ??? gives the observed incidence trajectory, as well as our estimated incidence trajectory after optimizing q_1 , q_2 , \mathbf{E}_0 and \mathbf{L}_0 , where we have used values from ?] for the non-estimated parameters. For reference, we also include the estimated incidence trajectory obtained by using values from ?] for the estimated parameters. Note that optimizing parameters here improves the fit dramatically.

4.2 Sensitivity

Those parameters not estimated (i.e. all except q_1 , q_2 , \mathbf{E}_0 and \mathbf{L}_0) are fixed at levels determined from the medical and epidemiological literature. In order to explore the impact of these parameters on the overall model, we select a range of values for each

and repeat the analysis described in Section 4.1 at every combination of parameter values. This gives us a measure of how sensitive our findings are to each parameter over the range of values we explore. For more on sensitivity analysis, see [citation needed].

[awk] Each non-estimated parameter has either three or four levels (see Tables 1 and 2), and there are a total of ??? combinations. For each such parameter, we divide all analyses based on the levels of that parameter. We then plot all incidence trajectories obtained from a single level of that parameter in the same plot, along with the average trajectory. We then plot the average trajectory for each level of the parameter, as well as the global average trajectory. This combination of plots gives a visual representation of the relative variability of estimated incidence across levels of the parameter. Comparing plots across parameters allows us to assess the relative sensitivity of our model to each non-estimated parameter.

In addition to the above trajectory plots, we also investigate the distribution of parameter estimates across all configurations of the non-estimated parameters. Figure 2 gives a histogram for our final estimates of each of q_1 , q_2 , E_0 , L_0 and R_0 . These histograms display the range of values for each parameter. Recall that \mathbf{L}_0 is minimally affected by optimization, so the three bars in its histogram correspond to the three initial values for this parameter. In order to see how well fitted models match the available data, we also give a histogram of the optimal objective function value achieved under each configuration of the non-estimated parameters. See Figure ??.

While it is instructive to investigate the breadth of our parameter estimates, we are mostly interested in the best such estimates. To this end, we restrict attention to the 3? 5? parameter estimates that most closely match our dataset (i.e. with smallest optimal objective function value). These “best” estimates are shown in Figure 2 by vertical lines. The difference in objective function between the first and (third? fifth? best estimates is negligible (confirm this).

5 Simulation and Results

5.1 Defining the loss function

Use Matlab’s `ode23` routine, we solve the system of differential equations from [?], which returns population vs time. From the population, we compute the incidence estimated TB incidence $\hat{f}(t)$

$$\text{Estimated TB Incidence} = \frac{100,000}{X(t) + E(t) + L(t) + T(t) + R(t)} \cdot (pwE(t) + vL(t)),$$

which can be compared to the reported TB incidence in from official Canadian reports $f(t)$ [?]. Our loss function is defined to be the difference between our computed (estimated) incidence and the reported incidence,

$$\text{Error} = \int_{2010}^{2020} (\hat{f}(t) - f(t))^2 dt.$$

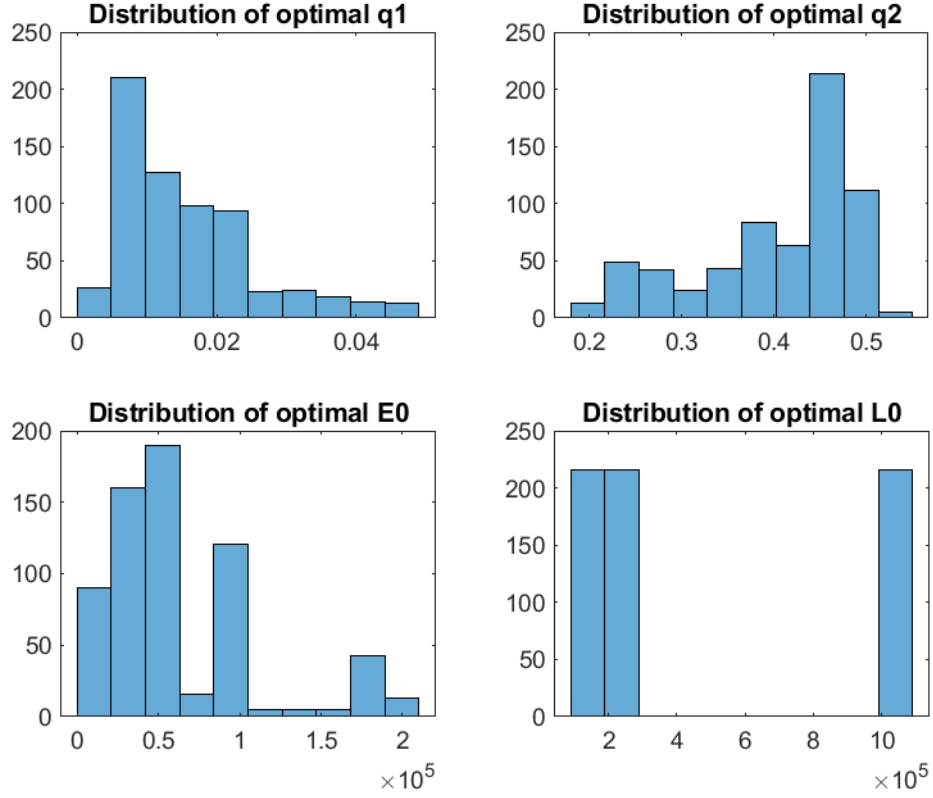


Fig. 2 Optimal values of estimated parameters under all configurations of non-estimated parameters. Vertical lines give the [3757](#) estimates with smallest objective function value.

In practice, $f(t)$ is reported annually from 2010 to 2020, and so we compute $\hat{f}(t)$ discretely. [awk]

5.2 Determining Initial Conditions

To solve the system of differential equations, initial conditions $[X_0, E_0, L_0, T_0, R_0]$ are required, which correspond to the foreign-born population distribution in 2010. T_0 is reported to be 1054 [?], and the total foreign-born population is estimated to be **WHATEVER IT IS** (found by taking the reported foreign-born population in 2011 and subtracting the immigration rate in 2010).

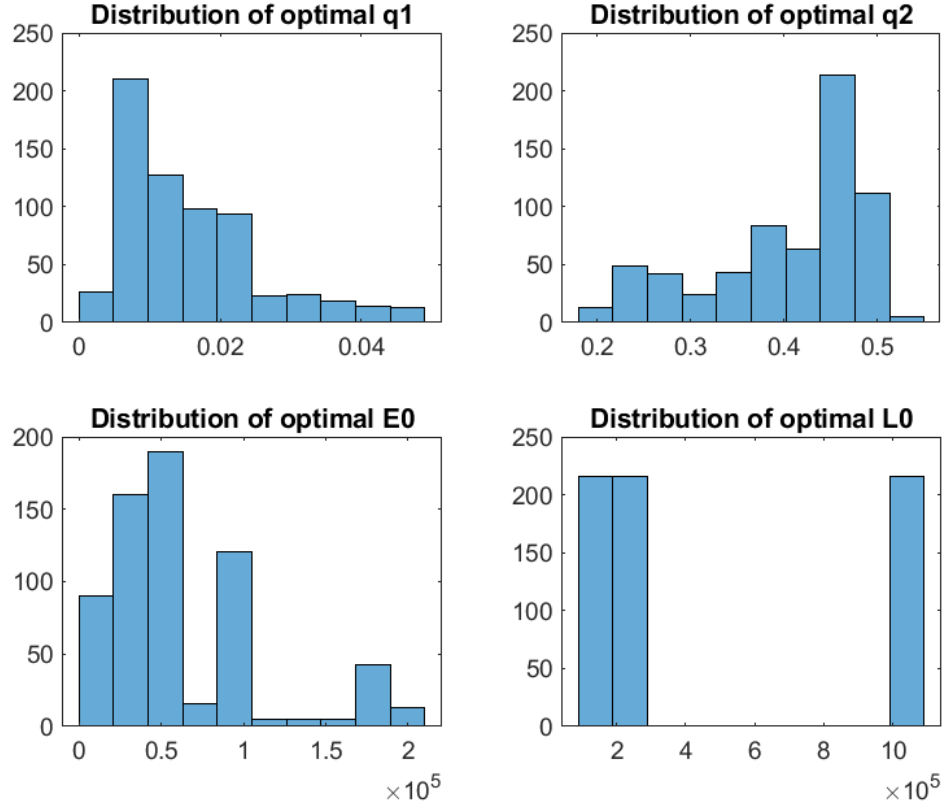


Fig. 3 After running sensitivity analysis, this is the distribution of the optimized values of q_1 , q_2 , E_0 , L_0 . Notice that L_0 has a peculiar distribution – this is because the optimized value of L_0 is always very close to the pre-optimized initial choice of L_0 .

5.2.1 Including E_0 and L_0 as parameters in the solver

5.2.2 Setting E_0 and L_0 based on incidence

Recall with $t_0 = 2010$

$$\text{Estimated TB Incidence in 2010} = \frac{100,000}{X(t_0) + E(t_0) + L(t_0) + T(t_0) + R(t_0)} \cdot (pwE(t_0) + vL(t_0)),$$

The denominator is exactly the total foreign-born population **QUOTE ABOVE**, and the reported incidence in 2010 is 14.1 [?], thus we choose E_0 and L_0 so

$$\text{Estimated TB Incidence in 2010} = 14.1.$$

5.2.3 Pointwise stationary approximation

5.3 Optimization Results

See Figure 4 to compare pre- and post- optimization behaviour.

5.4 Sensitivity Analysis

We performed sensitivity analysis across several different parameters. We started by looking across parameters we did not optimize across (i.e., anything except q_1 , q_2 , E_0 , L_0). We found that the results are most sensitive to ν and ω , which is unsurprising since they explicitly appear in the loss function.

Figure 5 a distribution of the optimized parameters.

6 Conclusions

Acknowledgments. Support from Langara College - Jeremy and Albert Support from UM - William

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading ‘Declarations’:

- Funding
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors’ contributions

If any of the sections are not relevant to your manuscript, please include the heading and write ‘Not applicable’ for that section.

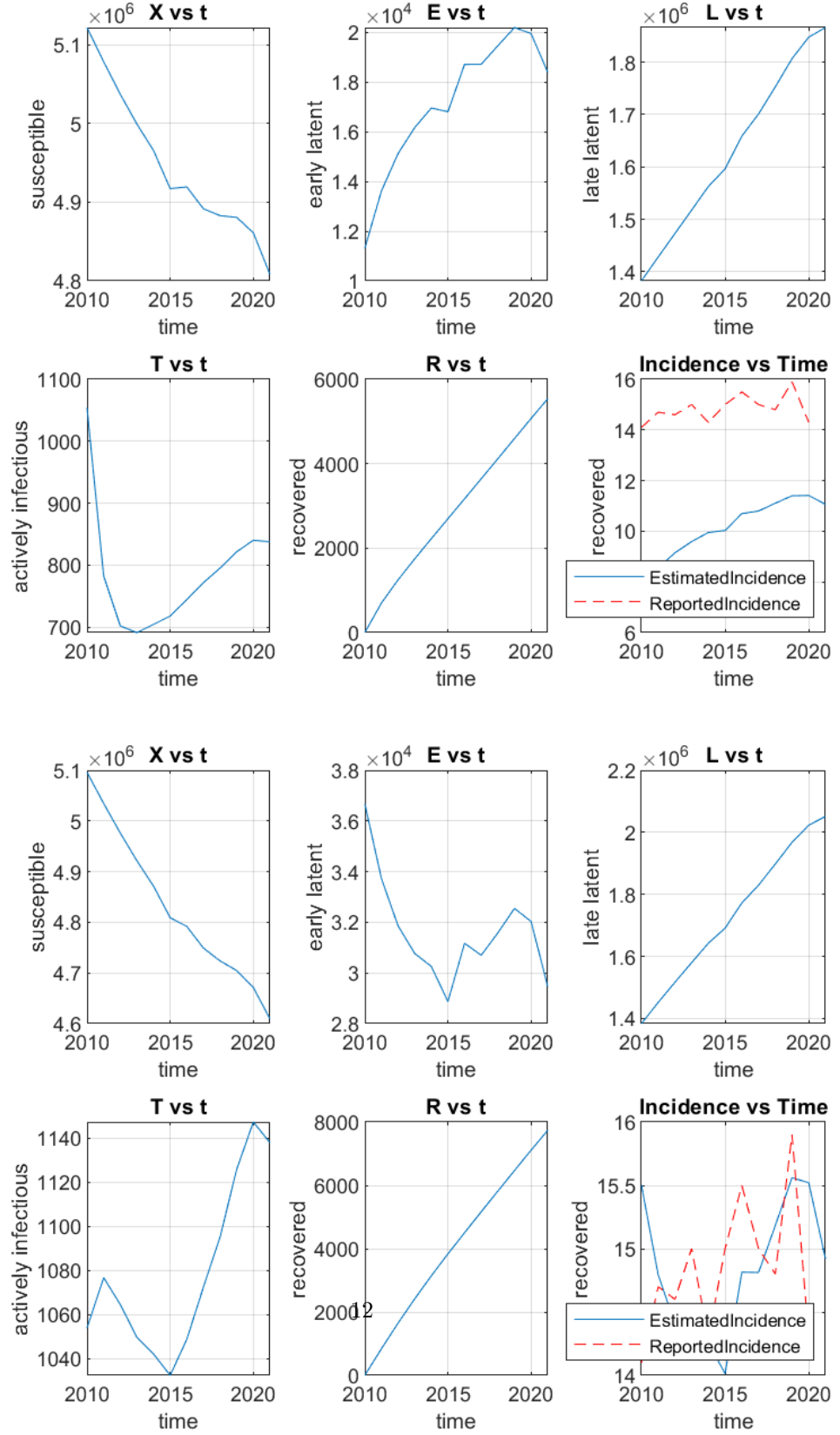


Fig. 4 Results using parameters from numerical simulations presented in [?]. In the first graph, no optimization is performed; in the second graph, parameters q_1 , q_2 , E_0 , L_0 are minimized using Matlab's `fmincon`.

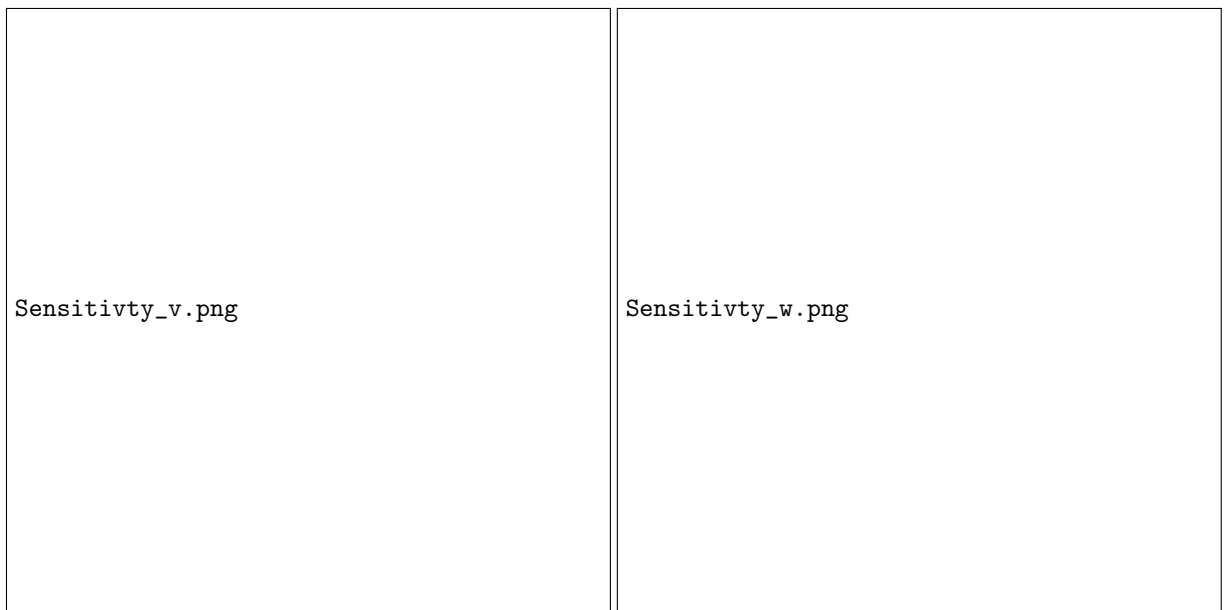


Fig. 5 Our results were most sensitive across ν and ω .