Mathematical modelling and optimization of Respiratory Syncytial Virus

Charlène Krick and Rafael J. Villanueva Micó

Motivation

In this project, we study Respiratory Syncytial Virus (RSV) in older adults.

Recently, the impact of RSV on adults has gained attention, with up to 18% of pneumonia-related hospitalizations in patients over 65 attributed to RSV.

In Spain, RSV is responsible for 15,000 to 20,000 primary care visits each year.

RSV triggers seasonal epidemics that overlap with other widespread viral infections such as influenza, resulting in numerous hospitalizations and placing significant strain on healthcare services. It spreads easily, with frequent occurrences of nosocomial infections.

Research on RSV and other viruses, along with the development of epidemic control strategies, is essential for both health and economic reasons.

Mathematical models are invaluable for analyzing infectious disease epidemiology, predicting societal impacts, and understanding the influence of external factors. Constructing a reliable RSV model is crucial for forecasting healthcare demands in upcoming seasons.

Given that RSV is a major cause of illness and mortality among older adults, vaccines have been developed and shown to be safe in studies. With a vaccine potentially becoming available soon, planning vaccination strategies has become urgent.

Consequently, we are considering a vaccination strategy for Valencia targeting individuals aged 50 and older.

1 Data

1.1 Demographic data

The Spanish region of Valencia is located in the eastern Mediterranean Spain with an area of 23 255 km^2 and a population of 5 069 163 inhabitants (2023), composed of three provinces, Castellon (north), Alicante (south) and Valencia (middle).

We consider here two age groups: G_1 (people under 50 years old) and G_2 (over 50 years old).

1.2 Hospitalisation data

From hospital database of the Spanish region of Valencia, adults over 50 years of age with positive RSV PCR done in the emergency department of selected hospitals. They are by weeks from the 1st week of 2016 to the 52th week of 2019.
After processing the data, we obtained the weekly number of hospitalizations of adults in age group G_2 caused by RSV related infection as showed in Figure 1.

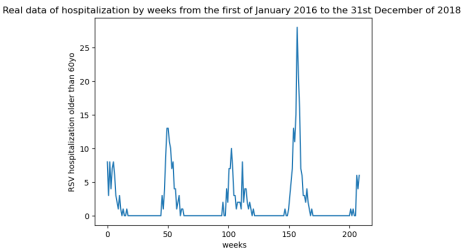


Figure 1: Weekly hospitalizations for respriratory syncitial virus infection from January 2016 to December 2019 of adults aged ¿50 years old in the region of Valencia

1.3 Comment

It should be noted that data refers to RSV infections detected in hospitals, not to all infected people. This means that only a proportion of infected are spotted in hospitals and this proportion, designated k, should be estimated in the model data fitting by the mean of a linear regression.

Moreover, the data is not really representative and viable because of the small amount of hospitalized people and then the variations are big for a small amount of data.

Consequently, to calibrate our model, we are going to use the yearly average of infectious data (more representative data than the hospital one). Indeed, according to Study of the burden of respiratory syncytial virus infection in community-dwelling older adults in Europe, infectious people represents 4,2% - 7,2% of people over 50 years old. Thus, we have to find a model where these data are confirmed.

2 Model

2.1 Mathematical modelling

We introduced an age-structured model with two age groups: the first one (i=1) corresponds to the people aged 0-50 years old and the second one (i=2) corresponds to the adults older than 50 years old.
This is justified because the disease is more acute in adults older than 50 years old, as confirmed by hospitalization data. For each age group, we have four subdivisions according to the state of the individuals with respect to the disease:

- Susceptibles $S_i(t), i = 1, 2$: the number of those at risk of contracting the disease
- Latents $L_i(t), i = 1, 2$: the number of those contracting the disease but are not contagious yet.

- Infectives $I_i(t), i = 1, 2$: the number of those infected and capable of transmitting the disease.

- Recovered $R_i(t), i = 1, 2$: the proportion of those recovered from the disease that are temporarily immune to re-infection.

In this model, it is assumed that there are some simplifying hypothesis :

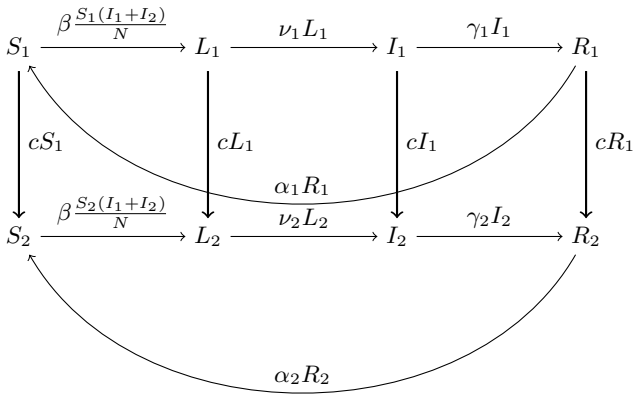
- Primary and secondary infections have the same rate of recovery and infectivity but it is not the case in real clinical situations.

- The mixing between the age group is homogeneous.

- There is no maternal antibody protection.

We note that $G_i(t), i = 1, 2$ is the number of people of the group i where the global population N is defined by $N = G_1 + G_2$.

We obtain this model for the two groups:



Where :

- μ is the yearly birth number.

- c is the transition rate from G_1 to G_2

- $d_i, i = 1, 2$ is the death rate for each group.

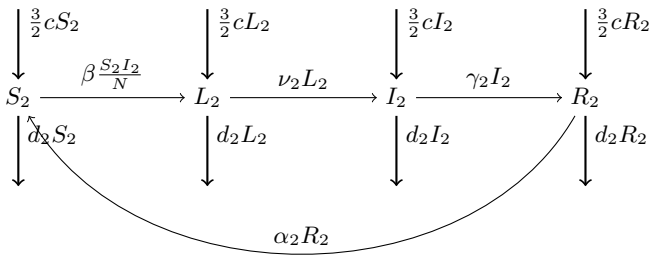
- $\beta = b_0(1 + b_1 cos(\frac{2\pi}{52}t + \phi))$ is the disease transmission rate, expressed as such in order to represent the yearly seasonality of the disease.Here, it is expressed weekly, that is why the time is divided by 52.

- ν_i where $\frac{1}{\nu_i}, i = 1, 2$ is the average time an individual of age group G_i becomes contagious after being in a latent state.

- γ_i, where $\frac{1}{\gamma_i}, i = 1, 2$ is the average time an individual of age group G_i recover from the illness.

- α_i where $\frac{1}{\alpha_i}, i = 1, 2$ is the average time an individual of G_i remains immune against re-infection.

As we focus on the second group only, we will obtain this model :



Indeed, the group under 50 years old is approximately 1,5 more populated than over 50 years old, thus we use the approximation : $S_1 = \frac{3}{2}S_2$, $L_1 = \frac{3}{2}L_2$,$I_1 = \frac{3}{2}I_2$ and $R_1 = \frac{3}{2}R_2$. Assuming that b_0 in β is much higher than in the first model of the two groups, we can simplify and consider only I_2 to go from S_2 to L_2.

The age-structured model is then defined by the following system of differential equations:

$$S_2'(t) = \frac{3}{2}cS_2 - \beta\frac{S_2I_2}{N} - d_2S_2 + \alpha_2R_2 \tag{1}$$

$$L_2'(t) = \frac{3}{2}cL_2 + \beta\frac{S_2I_2}{N} - \nu_2L_2 - d_2L_2 \tag{2}$$

$$I_2'(t) = \frac{3}{2}cI_2 + \nu_2L_2 - d_2I_2 - \gamma_2I_2 \tag{3}$$

$$R_2'(t) = \frac{3}{2}cR_2 - d_2R_2 + \gamma_2I_2 - \alpha_2R_2 \tag{4}$$

2.2 Numerical values

Numerical values of most of the coefficients described above can be assigned. Using demographic data of the Spanish region of Valencia for the year **2023**, where hospitalization data are available, we obtain :

- The mean population : $N = 5\ 069\ 163$.

- The mean population of both sub-populations: $G_1 = 3\ 233\ 613$ and $G_2 = 1\ 835\ 550$.

- The total of births per year is 45 800, thus the number of weekly births :$\mu = 880.77$.

- The mean weekly death rate for each age group: $d_1 = 1.08 * 10^{-5}$ and $d_2 = 0.00148$.

Moreover, the evolution of the population is described by the following system :

$$G_1' = \mu - d_1G_1 - cG_1 \tag{5}$$

$$G_2' = cG_1 - d_2G_2 = \frac{3}{2}cG_2 - d_2G_2 \tag{6}$$

$$N = G_1 + G_2 \tag{7}$$

Under the assumption of constant population for each age group ($G_1' = G_2' = 0$), some parameters can be determined.

$$c = \frac{\mu}{G_1} - d_1 \tag{8}$$

Thus, we obtain the transition rate from group 1 to group 2: $c = 2.6 * 10^{-4} weeks^{-1}$. However, we can also estimate this parameter by $c = \frac{1}{60}$ (over 60 years, one year is transitioning to group 2).Then, weekly, it is $c = \frac{\frac{1}{60}}{52}$ and we obtain

$c^* = 3.2 * 10^{-4} weeks^{-1}$. It is the same magnitude, consequently, it is coherent with our model. Moreover, if the hypothesis of constant population for each age group is valid, d_2 would satisfy the equation (9), ie:

$$d_2^* = c\frac{G_2}{G_1} = 0.00046 \tag{9}$$

Finally, we choose a new d_2^* satisfying the equations, which is different from the real value of d_2 because the population is not really constant in the four year interval chosen. Nevertheless, the difference is sufficiently small to be ignored in our study.

The time to recover from RSV is about 3 to 10 days, here, as there is a latency period, the average time to recover from RSV will be estimated to five days.
The time to be contagious is 4 to 5 days. For the same reasons, the average time to be contagious will be estimated to four days.
Finally, the average time to lose immunity is about 200 days.
The following parameters are defined by (in $weeks^{-1}$):

$$\gamma_2 \in [\frac{7}{10}, \frac{7}{3}] \tag{10}$$

$$\alpha_2 \in [\frac{7}{230}, \frac{7}{180}] \tag{11}$$

$$\nu_2 \in [\frac{7}{5}, \frac{7}{4}] \tag{12}$$

Hence, as can be seen, only parameters b_0,b_1 and ϕ remain unknown, i.e parameters dependent upon the seasonality, but also the coefficients of transition between different states (ν_2, α_2 and γ_2) can be found more precisely.

3 Calibration

3.1 Linear regression

To find the scale between hospitalized people $H(t)$ and infectious people $I(t)$, we need to minimize the error:

$$min_k ||H(t) - kI(t)|| \tag{13}$$

Here, we use the linear regression implemented in python from $sklearn$. For random values $b_0 = 4.98, b_1 = 0.86, \phi = 3.42$, and $\gamma_2 = \frac{7}{5}$ (5 days to become infectious),$\alpha_2 = \frac{7}{200}$ (200 days to lose immunity) ,$\nu_2 = \frac{7}{4}$(4 days to recover) , we find : r_2 score $= 0.68$ and $k = 0.00047332$.

The r_2 score can be better with other values of the undefined parameters. Moreover, because the data is not viable, the coefficient of determination will never be close to 1.

3.2 Error function

We need, firstly, to calibrate with the infectious data that we have.
The inside/outside error is defined by :

$$IO(p, [a, b]) = \begin{cases} 0, & \text{if } p \in [a, b] \\ min(|p - a|, p - b|), & \text{otherwise} \end{cases}$$

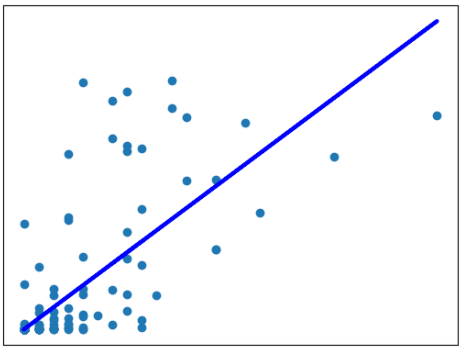


Figure 2: Linear regression between the hospitalizations and our model of infectious for values $b_0 = 4.98, b_1 = 0.86$ and $\phi = 3.42$

Then, we calculate the yearly average of infectious after the warm-up. We call it AI. So,

$$error_1 = IO(AI, [G_2 * a, G_2 * b]) \tag{14}$$

We know that $a = 4,2\%$ and $b = 7,2\%$,

$$error_1 = IO(AI, [77093.1, 132159.6]) \tag{15}$$

To calculate AI, we have to count unique infectious over 4 years to find the average. But we have only access, in our model, to the number of weekly infections even it is the same case.
Indeed, a person is infectious for $\frac{1}{\gamma_2}$ weeks. If we suppose, it is equally spread over a week, we can conclude that $\frac{7}{\gamma_2} - 1$ cases out of 7 cases (days) is on two weeks. Hence, $\frac{7}{\gamma_2} - 1$ cases out of 7 are counted twice.

If N_i is the number of infections over 4 years, based from our model and N_u is the number of separate cases of infections over 4 years, $N_i = (N_u + \frac{\frac{7}{\gamma_2} - 1}{7} N_u)$, thus, $N_u = (\frac{6}{7} + \frac{1}{\gamma_2}) * N_i$

$$AI = \frac{\frac{1}{\frac{6}{7} + \frac{1}{\gamma_2}} * N_i}{4} \tag{16}$$

This error is about 10^4 when it is not 0.
Thus, We have to find β which minimize $error_1$. (N_i is a functiun of β).
Error 2 is calculated from the hospitalizations data that we have over the four last years. Hence, the error to minimize is :

$$||kI(t) - H(t)|| \tag{17}$$

Where:

H(t) the weekly hospitalization from January 2016 to December 2019 from the database of the Valencian hospitals.

I(t) the weekly infectious from our model. (I(t)=f(β(t))).

k the scale which depends of the parameters of β.
This error is however about 10^1.

4 Parameters

To find the best parameters to minimize the error:

$$error = error_1 + error_2$$

Firstly, we use Nelder Mead method, already implemented in python library $scipy.optimize$, to minimize our problem.
However, our function error is not convex and, therefore, has several minima. To solve this problem and find the global minimum, we have to start from different points and find the minimum of the minima. Indeed, we simulate initial random values with bounds :

$$b_0 \in [0, 50] \tag{18}$$

We simulated several values and those bounds seemed more plausible.

$$b_1 \in [0, 1] \tag{19}$$

$$\phi \in [0, 7](2\pi) \tag{20}$$

And γ_2,ν_2 and α_2's bounds are described in (10), (11) and (12).

Then, we looked for the minimum with twenty random starting points to find a minimal error.

We found those parameters :

$$b_0 = 7.86232479 \tag{21}$$

$$b_1 = 0.69422302 \tag{22}$$

$$\phi = 3.3167752 \tag{23}$$

$$\gamma_2 = 2.33333333 \tag{24}$$

$$\alpha_2 = 0.03043478 \tag{25}$$

$$\nu_2 = 1.68162664 \tag{26}$$

By considering these parameters, the error is about 29.2 : $error_1 = 0$ and $error_2 \approx 29.2$. The average yearly rate of infectious over 50 years old is about $6.58\% \in [4.2\%, 7.2\%]$.

The scale obtained is : $k = 0.00065488$.

We compare the data of hospitalization (blue) that we got with our model (orange) and we obtain :

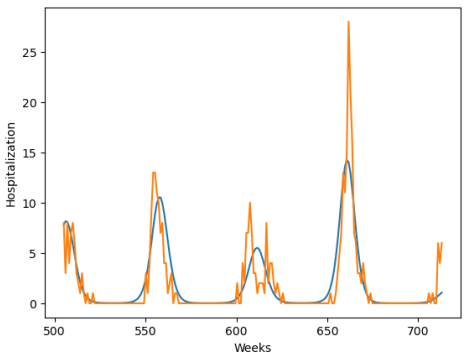


Figure 3: Hospitalization model and data over 4 years

We can observe that, with these parameters, the model is really close to our data.

Then, we tried to find optimal parameters with Particle swarm optimization algorithm but we did not find any better value of error than previously found with the Nelder Mead algorithm.

The model for infectious people over 4 years from 1st January of 2016 to 31 December of 2019 is :

We can predict the following years with the parameters found. Over 30 years after the end of the data, we obtain :

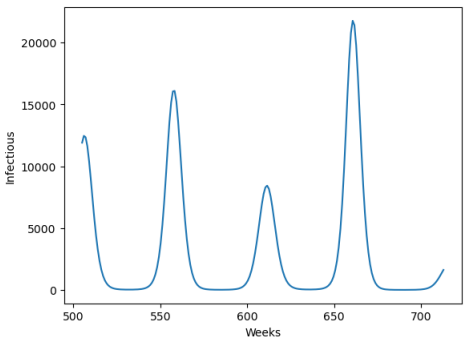


Figure 4: Infectious model over 4 years

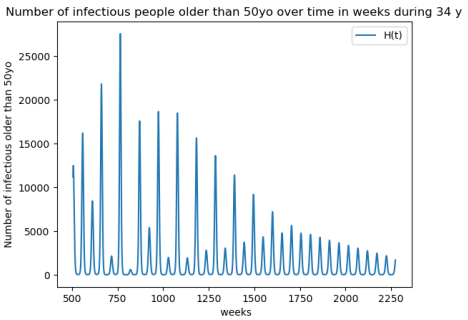


Figure 5: Infectious model over 34 years

As we can see, the model has two different amplitudes of periodicity: one decreasing and the other one increasing. Then, when the two different periodicities reach the same amplitude, the model decrease until 0.
Nevertheless, it can not decrease. In fact, the yearly average amount of infectious cases is about 4% on 10 years predicted, then about 2% on 30 years which is not in the normal interval of [4.2%,7.2%].
This result is not the one expected probably because of the poor quality of the data set.

5 Introduction of the vaccine

We are going to implement an agent-based model (ABM) to describe the vaccination. Vaccination is a way to get people out of the Susceptible - Latent - Infectives - Recovered - Susceptible contagion loop.
The idea is to keep track of the vaccinated nodes independently and avoid problems of integrating vaccination into the differential equations model due to the complexity of the vaccination campaign and the protection offered by the vaccine.
Here, we see how to combine the model with a system of differential equations with an agent-based model (ABM), which nature is mainly discrete, as we are going to see.
The vaccination campaign consists of vaccinating older people over 60 years old (59% of people over 50 years old according to our demographic data) following a pattern (vector vec) that depends on:

1. the vaccination coverage (percentage of people vaccinated).

2. the protection given by the vaccine (decreasing over time).

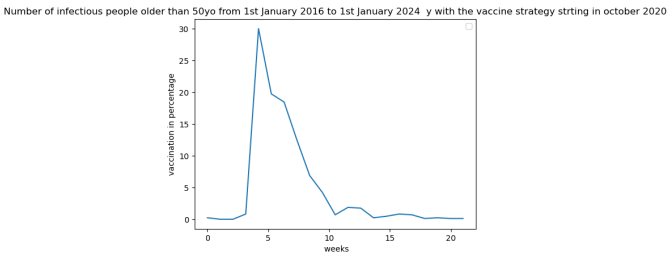


Figure 6: Fixed vaccination coverage, vaccination in percentage per week

In older people may be around 50%, but, we should simulate different coverages.
We assume that the vaccination of people over 60 years of age for RSV will follow the same vaccination pattern that people over 60 years of age followed to get vaccinated as in the 2016-2017 flu campaign in the Valencian Community.
For a fixed vaccination coverage, vaccination in percentage per week was :

$$vac = \{0.23, 0.00, 0.00, 0.82, 30.02, 19.74, 18.46, 12.50, 6.89, 4.21, 0.70, 1.87, 1.75, 0.23, 0.47, 0.82, 0.70, 0.12, 0.23, 0.12, 0.12\} \quad (27)$$

We are on day t, and x is a vaccinated person. We have the function of protection against infection

$$p(t) = \begin{cases} 82.43, & \text{if } t \in [0, 365] \\ max(-0.23t + 168.3, 0), & \text{otherwise} \end{cases} \quad \text{t in days} \quad (28)$$

This function is decreasing and reaches 0 for t = 732 days after vaccination. Let the protection drop vector be:
P = (p(1), p(2), ... , p(732)) / 100
which is a vector of decreasing values from $P_1 = p(1)/100 \leq 1$ to $P_{732} = p(732)/100 = 0$. Then, (*)

1. We generate a random number between 0 and 1.
2. If r $\geq P_1$, then the number of days this node x will be protected against infection is i = 1.
3. Otherwise, we search for element i of vector P such that
$$P_{i-1} > r \geq P_i.$$
Thus, the number of days this node x will be protected against infection is i days.
4. After i days this node x loses the protection and change its state to susceptible (where it may become infected).

This process is only used during the vaccination campaign for older people.
We are in **week t**.
This week, we are going to vaccinate

$$N = \frac{COV \times vac[t]}{100} \text{ people}$$

Let Vaux be a vector with N elements.

And with this new value of S(t), we keep solving the system of differential equations to obtain the values of all the subpopulations in the system of differential equations model.

Algorithm 1 Vaccination process in vaccination campaign

$Vaux \leftarrow [] \times N$
$N \leftarrow \frac{\text{COV} \times \text{vac}[t]}{100}$
for $k \leftarrow 1$ to N **do**
  $Vaux[k] \leftarrow$ days of vaccine protection assigned, as in section (*).
$counter \leftarrow 0$
for $k \leftarrow 1$ to length(V) **do**
  $V[k] \leftarrow V[k] - 1$ ▷ every day, we lose a day of protection
  if $V[k] == 0$ **then**
    $counter \leftarrow counter + 1$ ▷ if the protection is 0, the node moves to S
    remove $V[k]$ from V ▷ remove this element from the vector
$V \leftarrow V \cup Vaux$
$S(t) \leftarrow S(t) - N + counter$ ▷ Move on to the continuous problem

6 Simulation

Here, we simulate for different coverage the infectious after a vaccination campaign.

By applying once this vaccine on october 2020 during the 21 weeks into our model starting from 1st January 2016 to 1st January 2024 with a coverage of 60 % , we obtain :

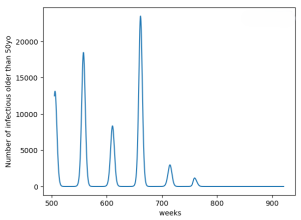


Figure 7: Infectious over 4 years after the vaccine campaign applied once (from 752th week) with a 60% coverage

We can observe that, using this coverage, the vaccination strategy proves to be efficient. During the next RSV outbreak season, the number of infectious cases consequently decreases, and over the following three seasons, RSV infections are almost nonexistent.

However, the effects of the vaccine are not immediate. In fact, during the campaign season, there is still a peak of infectious cases, around 1000 cases, before the number of cases drops significantly.

Now, we are going to look ten years ahead from the vaccination campaign from 1st January 2016 to 1st January 2030 with the vaccine strategy starting in october 2020 in Figure 8.

We can observe that from the 774th week (one week after the end of the vaccination campaign in early March 2020) to the 1014th week (five years after the start of the campaign in mid-October 2024), there is less than one infection case per day, with small fluctuations that are not visible in our graph due to their minimal size. Thus, during these 240 weeks, the vaccination campaign proves to be effective. However, infection cases begin to rise again, with a small peak of 202 cases during the 2024-2025 season (out of approximately 1400 distinct cases for that season), and subsequent seasonal peaks are higher compared to when there was no vaccination.

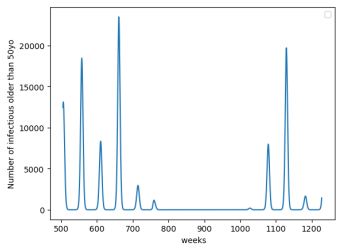


Figure 8: Infectious over 10 years after the vaccine campaign applied once (from 752th week) with a 60% coverage

We now predict 10 years ahead from one vaccination campaign for several coverages : 40 %, 30 % and 20 %.

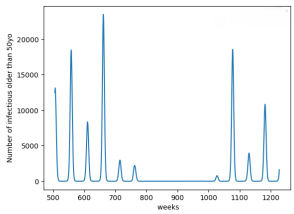


Figure 9: Coverage of 40 %

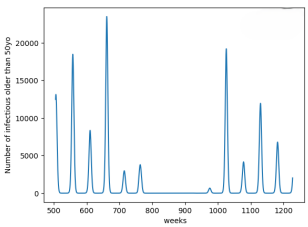


Figure 10: Coverage of 30 %

As we can expect, the smaller the coverage is, the shorter the period with less than one infectious case is.
In fact, when the coverage is about **40%**, this period lasts until the 1014th week in mid-October 2024 (**4 years**), similar to a 60% coverage but with a higher first peak of infectious cases during the campaign season (around 1800 cases at the highest week of the season) and the 2024-2025 season (around 800 cases at the highest week of the season).

Furthermore, for a **30%** coverage, this period is one year shorter than the 40% coverage, lasting **3 years**, and then exhibits the same behavior as the previous case.
However, the peak of the campaign season is higher (around 3000 cases at the highest week of the season).

Finally, for a **20%** coverage, this period lasts **2 years**, but the campaign season shows a much higher peak of infectious cases (around 7000 cases at the highest week of the season), which then drops to almost no cases. During the 2022-2023 season, there is a higher peak (around 1500 cases at the highest week of the season), and after the no-infectives period, the peaks are irregular.

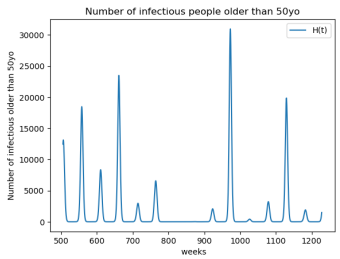


Figure 11: Coverage of 20%

To solve this issue and have a better efficiency for the vaccine strategy, we try to simulate a vaccination campaign every two years with a 20% coverage.

With this strategy we obtain 12:

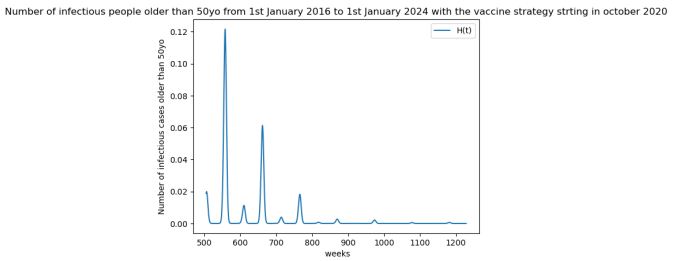


Figure 12: Vaccination strategy every two years with a 20% coverage

Over time, the intensity of the infection peaks decreases. This indicates that the periodic vaccination campaigns contribute to a long-term reduction in infections.

Toward the end of the observed period, the number of infectious cases stabilizes at a low level, indicating the cumulative effect of repeated vaccination campaigns.

Some small fluctuations remain even after several vaccination campaigns, likely due to factors such as incomplete vaccination coverage or variations in vaccine effectiveness.

The periodic vaccination campaigns with 20% coverage effectively reduce the number of infections among people over 50. Over time, the infection peaks diminish, leading to a stable and low level of infections. However, some residual fluctuations persist.

Conclusion

In this project, we aimed to predict infections among individuals over 50 years old to simulate a vaccination strategy targeting this group. We identified a potentially effective strategy.

However, our dataset is small, and the data is not entirely reliable. Therefore, we cannot be certain that our predictions are accurate. In fact, according to healthcare professionals, the number of infections does not decrease over time.

Finally, I would like to thank Rafael J. Villanueva Micó for his help, his knowledge and his support throughout this project.