A system of differential equations assessing the impact of vaccination on lung function decline in individuals with Cystic Fibrosis

Department of Mathematics
Utrecht University

Bachelorthesis by:

Marieke Smit-Muntinghe

6430449

Abstract

Cystic Fibrosis is an autosomal recessive disease that causes a progressive loss of lung function. Currently, there exists no effective vaccine against the bacteria primarily responsible for lung function decline. In this study, we develop a system of differential equations to evaluate changes in lung function over time in individuals diagnosed with CF. These changes are measured using the $FEV_1\%$ value, comparing vaccinated and unvaccinated scenarios. Our goal is to assess whether vaccination can slow disease progression. To achieve this, we will use a Python script to explore solutions and analyze outcomes. The results indicate that vaccination can benefit individuals across all age groups, as it leads to less lung function deterioration. However, 50 years from now, a vaccine with 100% efficacy would only lead to a 2-year increase in life expectancy.

17-1-2025

Supervised by:
Dr. M.C.J. Bootsma

Contents

1 Introduction

Cystic Fibrosis (CF) is an autosomal recessive genetic disorder caused by mutations in the CFTR (cystic fibrosis transmembrane conductance regulator) gene. This gene encodes for the eponymous protein, which plays a critical role in regulating the movement of salt and water across the cells that produce mucus, sweat, and digestive fluids. Due to these mutations, the normal mucus production of the CFTR protein is disrupted, leading to the production of thick and sticky mucus in various organs of the body, particularly in the lungs and the digestive system [1]. Consequently, the mucus can obstruct the lung airways, pancreas ducts, the reproductive system and the gastrointestinal tract, trapping bacteria and creating an environment conducive to persistent infections and chronic inflammation. Over time, this can result in progressive lung damage, reduced lung function, and respiratory failure, due to (persistent) lung infections. It will also cause complications in other body parts [2].

The median age of people with CF is now 22 years, with a median age of 21 days at diagnosis. While there is no cure for CF, important advancements in treatment over the years have enabled patients to live longer and healthier lives. Treatments include medications that help to thin and clear the thick mucus from the airways, enzymes to support the absorption of fats and nutrients, and antibiotics to fight infections. Additionally, new oral medications known as protein modulators directly target fixing the defect in the CFTR protein. These protein modulators have greatly improved life expectancy and offer enormous benefits for individuals living with CF. [3]. Based on the latest available data, the median predicted survival age for people born between 2019 and 2023 with CF is 65 years [10].

There are several bacteria giving rise to infections causing lung damage. The major pathogen in the CF lung is the highly prevalent gram-negative bacterium called *Pseudomonas aeruginosa* (*P. aeruginosa*). About 50% of individuals with CF are infected with *P. aeruginosa* [4]. It causes lung infections in people with an abnormal functioning immune system like CF patients, and once acquired, a chronic infection will almost inevitably develop, resulting in an accelerated decline of lung function [5].

The loss of lung function is commonly measured by the Forced Expiratory Volume in 1 second (FEV_1), an indicator of airway obstruction that has played a crucial role in clinical care and research. It can be corrected to a percentage of normal values, by dividing the patients FEV_1 by the average FEV_1 of people with similar age, height, gender, race and ethnicity. This provides the $FEV_1\%$ value. Declines in $FEV_1\%$ are mainly due to the interaction between chronical bacterial infections, mentioned above, and the induced inflammatory immune response [9]. For CF patients, a target $FEV_1\%$ measured of 85% or higher is considered optimal, as this indicates (almost) normal lung health. However, each individuals target $FEV_1\%$ will vary based on their own lung function results and progression [7]. Carriage of the bacteria *P. aeruginosa* contributes to a more rapid decrease in $FEV_1\%$ and a higher mortality. Additionally, lower $FEV_1\%$ facilitates the acquisition of these bacteria, reduces their rate of loss, which increases mortality as well. Therefore, curing or even preventing infection of *P. aeruginosa* through vaccination could improve survivorship and $FEV_1\%$ [6].This brings us to the following research question:

"What is the impact of vaccination on lung function decline and disease progression in individuals with CF?"

To provide an answer to this question, we will start by explaining how we modeled a system of differential equations to study the progression of lung function in CF patients. Each key

component of the model will be examined in detail to provide a clearer understanding of how they contribute to the system. We will define how we can use this model to investigate the influence of vaccinating against bacteria such as *P. aeruginosa*. Following this, we will focus on solving the problem and implementing the solution in a Python code. Finally, we will present the results and interpret them to provide an answer to the research question.

2 Methods

Understanding the dynamics behind the decline in lung function in people diagnosed with CF can help develop interventions and improve patient outcomes. To address this, we modeled a system of differential equations that describes the transition of individuals of a certain age with CF through stages of decreasing $FEV_1\%$, accounting for natural progression of the disease, mortality and aging. We will use parameters derived from the 2022 Annual Data Report of the UK Cystic Fibrosis Registry, where they annually register data from over 11,000 people with CF across the UK [7]. Additionally, we aim to evaluate the influence of vaccines on disease progression by evaluating their impact on the transition rates r, which represent the rate at which individuals move from one stage of $FEV_1\%$ to the next.

2.1 Outline of the system of differential equations

We will begin by outlining the structure of the system of differential equations we will be working with. To enhance clarity, we will also provide a visual representation of the model.

In this model, individuals from different age groups are categorized based on their $FEV_1\%$ value. The size of each lung function group changes due to mortality and aging, as well as individuals entering from a previous group or moving to a lower lung function group both caused by decreasing $FEV_1\%$. The system of differential equations incorporates the following key components:

1. Lung function (V_i)
    - $V_i(t)$ denotes the population of individuals in the i-th lung function group whose $FEV_1\%$ value falls within a specific range.
    - The initial value is determined by the data from the 2022 Annual Data Report of the UK Cystic Fibrosis Registry and represents the number of people of a certain age per lung function group at $t = 0$.
2. Transition rate (r_i)
    - Individuals can enter the group V_i from the previous group V_{i-1} with a rate of r_{i-1}. This is associated to progression of the disease.
    - Individuals can also leave the group V_i with a rate of r_i as their lung function deteriorates further, entering the next group V_{i+1}.
3. Mortality rate (μ_i)
    - Within each group a proportion of individuals die with rate μ_i, resulting in a reduction in group size.
4. Rate of aging (u_i)
    - Within each age group, individuals transition to the next age group as they age with a certain rate.

This can be pictured as follows:

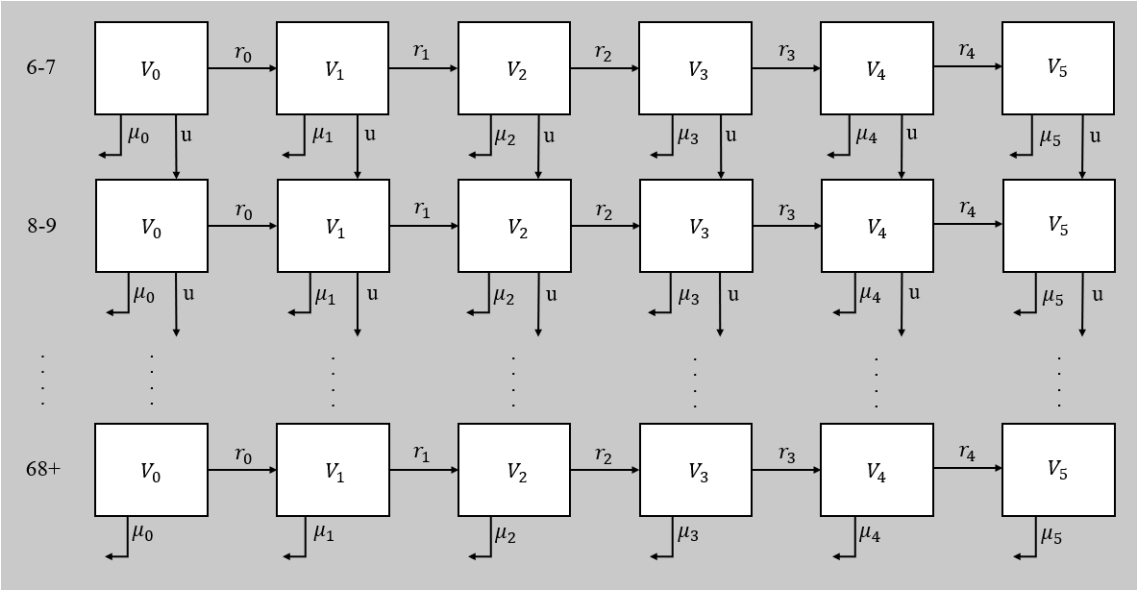


Figure 1: Visualization of the system of differential equations, where V_i represents the lung function groups, r_i the transition rates, μ_i the mortality rates and u the rate of aging. While only three age groups are illustrated, the model includes all age groups.

The population dynamics for each age group and lung function category are described by the following system of differential equations:

$$\begin{cases} V_i^{\text{age}}(0) = n_i^{\text{age}} \\ \frac{dV_0^{6-7}}{dt} = 213 - \left(r_0^{6-7} + \mu_0^{6-7} + u^{6-7}\right) V_0^{6-7} \\ \frac{dV_i^{\text{age}}}{dt} = -\left(r_i^{\text{age}} + \mu_i^{\text{age}} + u^{\text{age}}\right) V_i^{\text{age}} + r_{i-1}^{\text{age}} V_{i-1}^{\text{age}} + u^{\text{age-1}} V_{i-1}^{\text{age-1}}. \end{cases} \tag{1}$$

For the hightest $\text{FEV}_1\%$ group V_0^{6-7} the equation shows the outflow due to progression r_0, mortality μ_0 and aging u_0. The constant term accounts for new individuals aged 6–7 entering the dataset, assuming they have healthy lung function upon entry. We arrived at this value because it represents half of the individuals initially in this age group, which we will explain in further detail later. The data supporting this can be found in Figure 8 in the Appendix.
The dynamics of the subsequent groups report an inflow from the previous group and an outflow due to progression, mortality and aging. The first age group defined in the 2022 Annual Data Report of the UK Cystic Fibrosis Registry consists of individuals aged 6–7. As our study relies on data from this registry, our model starts with the 6–7 age group. Therefore, transitioning into this age group from an earlier one is not possible, causing the positive u term to disappear. Similarly, since it is not possible to exit the 68+ age group, here the negative u term disappears.

We will now systematically examine each key component in detail, to provide a better understanding of how they contribute to the system. This will help us to clearly define the differential equation system employed in this study.

2.1.1 Lung function groups

We will make use of data from The 2022 UK CF Registry Annual Data Report to divide all the participants into distinct lung function groups. Firstly, we define six groups based

on different intervals of $\text{FEV}_1\%$ values, categorized as follows:

- V_0: $\text{FEV}_1\% > 100\%$
- V_1: $\text{FEV}_1\%$ between $90 - 100\%$
- V_2: $\text{FEV}_1\%$ between $80 - 90\%$
- V_3: $\text{FEV}_1\%$ between $70 - 80\%$
- V_4: $\text{FEV}_1\%$ between $60 - 70\%$
- V_5: $\text{FEV}_1\% < 60\%$

The data report presents the number of people in each age group, along with the average $\text{FEV}_1\%$ and its standard deviation (see Appendix, Figure 8). Using the properties of a normal distribution, we can determine which group individuals fall into based on the mean $\text{FEV}_1\%$ values and the standard deviation. To compute the number of individuals in every lung function groups, we will use the z-score and the z-table. The z-score is a measure that indicates how many standard deviations a specific data point (in this case, a specific value of $\text{FEV}_1\%$) is from the mean of the distribution [13].

We define:

$$Z = \frac{X - \mu}{\sigma}.$$

Proposition 2.1 If $X \sim N(\mu, \sigma^2)$, then $Z \sim N(0, 1)$.

Proof. The cumulative distribution function of Z is

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq z\right) \\ &= \mathbb{P}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma} \leq z\right) \\ &= \mathbb{P}(X \leq z\sigma + \mu) \\ &= F_X(z\sigma + \mu). \end{aligned}$$

Thus,

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_X(z\sigma + \mu) \\ &= \sigma f_X(z\sigma + \mu). \end{aligned}$$

We assumed that $X \sim N(\mu, \sigma^2)$, in other words f_X is a normal density function with parameters μ and σ. Therefore $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. So we find that, after substitution,

$$\begin{aligned} f_Z(z) &= \sigma \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(z\sigma+\mu)-\mu}{\sigma}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \end{aligned}$$

From this we can conclude that Z follows a normal distribution with parameters $\mu = 0$ and $\sigma = 1$, thus $Z \sim N(0, 1)$ [13]. □

We can use this result to compute $F_X(x)$:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right). \end{aligned}$$

Since $Z \sim N(0,1)$, it follows that $\mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, which we can derive from the z-table. The cumulative z-table provides the area under the curve left of the z-score: $\frac{x-\mu}{\sigma}$. This corresponds to the probability that X is less or equal to x.

We will now focus more deeply on the age group 6-7 and show the calculation for this group in detail. The other age groups will be analyzed by a similar approach. From the data report we get that there are 425 individuals in the age group 6-7, with a mean $\text{FEV}_1\%$ of 97.1% and a standard deviation of 15.8%. For every V_i we will determine the number of individuals from this age group who fall within the specified $\text{FEV}_1\%$ range.

- V_0: $\text{FEV}_1\% > 100\%$
  - We start with calculating the probability of having a $\text{FEV}_1\% > 100\%$:

$$\begin{aligned} \mathbb{P}(X > 100) &= 1 - \mathbb{P}(X \leq 100) \\ &= 1 - \Phi\left(\frac{100-\mu}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{100-97.1}{15.8}\right) \\ &= 1 - \Phi(0.18) \\ &\approx 0.429 \end{aligned}$$

    So the probability is 42.9%.
  - This corresponds to $425 \cdot 0.429 \approx 182$ people.
- V_1: $\text{FEV}_1\%$ between $90 - 100\%$
  - We start with calculating the probability of having a $\text{FEV}_1\%$ between $90-100\%$:

$$\begin{aligned} \mathbb{P}(90 < X \leq 100) &= \mathbb{P}(X \leq 100) - \mathbb{P}(X \leq 90) \\ &= \Phi\left(\frac{100-97.1}{15.8}\right) - \Phi\left(\frac{90-97.1}{15.8}\right) \\ &\approx 0.571 - 0.326 \\ &\approx 0.245 \end{aligned}$$

    So the probability is 24.5%.
  - This corresponds to $425 \cdot 0.241 \approx 104$ people.
- V_2: $\text{FEV}_1\%$ between $80 - 90\%$
  - Calculated in the same way as before, we obtain a quantity of ≈ 79 people.
- V_3: $\text{FEV}_1\%$ between $70 - 80\%$
  - Calculated in the same way as before, we obtain a quantity of ≈ 41 people.

- V_4: $\text{FEV}_1\%$ between $60 - 70\%$
  - Calculated in the same way as before, we obtain a quantity of ≈ 14 people.
- V_5: $\text{FEV}_1\% < 60\%$
  - Before we are able to calculate the initial value for this group, we must prove the following proposition:

    Proposition 2.2 Let X be a continuous random variable. Then for every $x \in \mathbb{R}$:

    $$\mathbb{P}(X < x) = \mathbb{P}(X \leq x).$$

    Proof. First of all we need to proof that the cumulative distribution function F_X is continuous. F_X is defined as $F_X(x) = \mathbb{P}(X \leq x)$ and for a continuous random variable this can be expressed as the integral of its probability density function $f_X(u)$:

    $$F_X(x) = \int_{-\infty}^{x} f_X(u)du.$$

    The Fundamental Theorem of Calculus states that if $f_X(u)$ is a continuous function on the interval, then the integral of $f_X(u)$, $F_X(x)$, is differentiable and its derivative is given by $\frac{d}{dx}F_X(x) = f_X(x)$. Since $f_X(x)$ is continuous, as X is a continuous random variable, $F_X(x)$ is differentiable on the whole interval and thus continuous.

    It follows that for every $x, y \in \mathbb{R}$

    $$\mathbb{P}(X \leq x) = F_X(x) = \lim_{y \uparrow x} F_X(y) = \mathbb{P}(X < x)$$

    where we used the property of the cumulative distribution function $\lim_{x \uparrow y} F_X(x) = \mathbb{P}(X < y)$ [15]. □
  - Using this result, we can express $\mathbb{P}(X < 60) = \mathbb{P}(X \leq 60)$. Following a similar calculation as before, we obtain a quantity of ≈ 4 people.

The other age groups will be analyzed by a comparable approach. The results can be found in the table in Figure 2. Logically, as people age, their lung function tends to decline, and the amount of measurements become less.

Age-groups	Lung function groups ($FEV_1\%$)					
	V_0	V_1	V_2	V_3	V_4	V_5
6-7	182	104	79	41	14	4
8-9	194	118	84	39	11	2
10-11	192	133	112	62	23	7
12-15	324	236	205	126	51	18
16-19	207	156	157	117	64	36
20-23	143	109	129	131	107	137
24-27	119	101	132	146	127	196
28-31	91	83	110	128	117	207
32-35	63	64	97	114	116	216
36-39	42	45	67	81	85	176
40-43	40	39	57	70	73	162
44-47	19	20	31	42	44	108
48-51	16	18	27	35	37	81
52-55	13	13	19	25	25	61
56-59	10	10	15	17	18	39
60-63	5	6	9	12	13	29
64-67	5	4	6	8	8	22
68 +	8	6	9	10	9	20
Total	1673	1265	1345	1204	942	1521

Figure 2: Initial group size values.

Now that we have determined the number of individuals within each lung function group, we can establish the baseline for our model. We set $t = 0$ as the time when the report is published (2022). In this context, the initial values for each group correspond to $V_i(0)$, where $V_i(0)$ represents the number of individuals in group i at the starting point of the analysis. This establishes a basis for analyzing the progression of lung function changes over time within each group. Starting from the initial state, we can model the transitions between groups while accounting for mortality and aging rates.

2.1.2 Mortality rates

The mortality rate is established by two factors: the general mortality risk for healthy individuals in each age group, and the hazard ratio, which quantifies the relative risk of death associated with different levels of $FEV_1\%$ impairment compared to healthy individuals. We multiply them to determine the mortality rates for our model. By doing so, we are essentially combining the general mortality risk for an age group with the additional risk due to poor lung function. This is reasonable because individuals with poorer lung function experience a higher likelihood of death, so the total mortality risk should reflect both the standard risk and the relative risk from impaired health. We define

$$\mu_i^{\text{age}} = \mu^{\text{age}} \cdot \mu_i \tag{2}$$

where μ^{age} represents the general mortality risk and μ_i the mortality risk associated with lung function.

The standard mortality risk for each age group can be obtained from data collected by the CBS (Statistics Netherlands). We use the combined data for both men and women from 2023. To determine the mortality risk for, for example, the 6-7 age group, we use the average of the mortality risks for individuals aged 6 and 7 [14].

A study publiced in The Lancet Global Health in May 2019 described the association between $FEV_1\%$ and mortality. $FEV_1\%$ was categorized as no impairment ($FEV_1\%$ 112·9 (95% CI 106–122)), mild impairment ($FEV_1\%$ 91·5 (95 % CI 86–96)), moderate impairment ($FEV_1\%$ 70·7 (95 % CI 64–75·4)), and severe impairment ($FEV_1\%$ 46·9 (95% CI 35·4–53)). Follow-up was done every 3 years to collect information on mortality among 126.359 adults with acceptable spirometry data available. A pattern of increasing mortality risk with decreasing $FEV_1\%$ was observed across most geographical regions and also in healthy individuals. The findings of the study show that individuals with mild $FEV_1\%$ impairment have a 27% higher risk of mortality compared to those without impairment. This is indicated by the hazard ratio of 1.27. Secondly, individuals with moderate $FEV_1\%$ impairment have a 74% higher risk of mortality compared to the no-impairment group (HR of 1.74) and lastly individuals with severe $FEV_1\%$ impairment have a 154% higher risk of mortality compared to those without impairment (HR of 2.54) [12].

To calculate the mortality rates for our model, we use the findings from the study described above. Plotting the relative risks against their corresponding $FEV_1\%$ values shows an exponential rise in relative risks as $FEV_1\%$ decreases, see Appendix, Figure 9. This graph enables us to estimate the relative risks for different $FEV_1\%$ values, and consequently for our lung function groups. For instance, for V_1, we identified the relative risks for $FEV_1\%$ values of 90 and 100, then averaged them to define μ_i for this group. We took the average because the graph is nearly linear between those $FEV_1\%$ values. A similar method was applied to the other groups. After determining μ_i for all V_i we calculate μ_i^{age} using equation (4). The results are shown in the table in Figure 3. By determining the mortality rates, we can proceed to the next step in clarifying the system of differential equations: defining the rates of aging.

Age-groups	Mortality rates					
	μ_0	μ_1	μ_2	μ_3	μ_4	μ_5
6-7	0,00007	0,00008	0,00010	0,00012	0,00015	0,00022
8-9	0,00008	0,00009	0,00011	0,00014	0,00018	0,00025
10-11	0,00009	0,00010	0,00013	0,00016	0,00020	0,00029
12-15	0,00010	0,00011	0,00015	0,00018	0,00023	0,00033
16-19	0,00022	0,00025	0,00032	0,00041	0,00051	0,00073
20-23	0,00027	0,00031	0,00040	0,00051	0,00064	0,00091
24-27	0,00034	0,00039	0,00050	0,00063	0,00080	0,00114
28-31	0,00039	0,00045	0,00057	0,00072	0,00091	0,00130
32-35	0,00049	0,00057	0,00073	0,00091	0,00115	0,00164
36-39	0,00064	0,00074	0,00095	0,00120	0,00151	0,00216
40-43	0,00090	0,00105	0,00134	0,00169	0,00213	0,00304
44-47	0,00125	0,00145	0,00186	0,00234	0,00294	0,00420
48-51	0,00181	0,00210	0,00270	0,00339	0,00427	0,00610
52-55	0,00276	0,00320	0,00411	0,00517	0,00651	0,00929
56-59	0,00417	0,00484	0,00622	0,00782	0,00985	0,01405
60-63	0,00609	0,00706	0,00906	0,01141	0,01436	0,02050
64-67	0,00938	0,01087	0,01396	0,01758	0,02212	0,03158
68 +	0,11292	0,13093	0,16820	0,21169	0,26644	0,38035

Figure 3: Mortality rates, where μ_i denotes the mortality risk of individuals in V_i.

2.1.3 Rates of aging

Since individuals are divided into age groups, as they grow older, they can transition to the next age group. The rate at which they transition is determined by the number of years within each group. We assume that individuals of every age are equally represented within a specific age group. For instance, in the 6-7 age group, 50% of individuals are 6 years old and 50% are 7 years old. As a result, after one year, half of the individuals in this group moves to the next age group: 8-9. This can be captured by the following equation:

$$u^{\text{age}} = \frac{1}{\text{The number of ages represented in the group}}. \tag{3}$$

In addition to the mortality and aging rates, there is one remaining factor that influences the growth of the lung function groups. This brings us to the last step in clarifying the model: specifying the transition rates.

2.1.4 Transition rates

The 2022 UK CF Registry annual data report reveals there is a gradual decline in $\text{FEV}_1\%$ in CF patients aged six years and older who have not had a lung transplant. We use the best $\text{FEV}_1\%$ predicted value each year as lower values, often measured during acute exacerbations, vary significantly based on the severity and stage of each exacerbation. Therefore, the best $\text{FEV}_1\%$ predicted value is the most consistently reproducible value and most reliable. Additionally, patients who have had a lung transplant are excluded, because their 'new' lungs may have lung health similar to a person without CF, and therefore are not representable. Because treatments have been improved over the last years, the decline in $\text{FEV}_1\%$ has become less severe. However, the report's findings reveal that this condition persists, leading to a decline

in lung function as the disease progresses [7]. Without new treatments, this process is expected to continue.

In the absence of a vaccine, we predict a steady-state scenario. In this model, an additional 213 individuals join the cohort each year. Over time, they transition to different groups due to lung function decline, aging, or a combination of both. Some individuals will pass away, exiting the cohort. Despite these dynamics, the overall distribution of individuals across specific age and lung function groups is expected to stay relatively stable. In other words, the initial sizes calculated for each group in the previous section are anticipated to remain constant over the years. This can be expressed in the context of our system of differential equations as:

$$\frac{dV_i^{\text{age}}}{dt} = -\left(r_i^{\text{age}} + \mu_i^{\text{age}} + u^{\text{age}}\right)V_i^{\text{age}} + r_{i-1}^{\text{age}}V_{i-1}^{\text{age}} + u^{\text{age-1}}V_{i-1}^{\text{age-1}} = 0$$

for all $i = 0, ..., 5$ and all ages.

Since we already calculated the initial values for V_i^{age}, the mortality rates μ_i^{age} and the rates of aging u^{age}, we now need to determine the transition rates that satisfy this expression. In order to find this, we start by redefining the system. We can write the system as a non-homogeneous system of first order differential equations:

$$\frac{dV_i}{dt} = \sum_{j=0}^{n} a_{ij}V_j, + \, b_i \quad i = 1, 2, ..., n$$

with $n = 108$ and the constant coefficients $a_{ij} \in \mathbb{R}$ and $b = (213, 0, 0, ..., 0)^T$. Or, in vector notation;

$$\frac{dV}{dt} = AV + b$$

where $A \in M_n(\mathbb{R})$ is a constant n x n-matrix with elements a_{ij} and $V, b \in \mathbb{R}^n$.

In our model, the matrix A contains four different types of elements: $-\left(r_i^{\text{age}} + \mu_i^{\text{age}} + u^{\text{age}}\right)$, r_i^{age}, u^{age} or 0. Therefore, the only unknowns in the matrix are the transition rates. For instance, in the first row of the matrix A, only r_0 is unknown. Since we know that the result of this row multiplied by V plus b, i.e. the differential equation of V_0^{6-7}, equals zero, we can determine r_0. Moving to the second row of A, the only transition rates are r_0 and r_1. Having already determined r_0, we can now solve for r_1. This process is repeated until all 108 transition rates are determined. However, we did apply certain restrictions to the transition rates. First, every transition rate corresponding to V_5 is set to zero, as individuals cannot transition out of the last lung function group due to disease progression. Additionally, we observed that the code occasionally produced negative values for r, which is not realistic since transitioning to a better lung function group is not possible. To address this, we implemented a condition ensuring that r is the maximum of the calculated value from the script and zero. The details of this procedure are outlined in the Python script provided in Appendix A.3.1.

Having outlined how we defined and calculated the transition rates, we now proceed to the last phase. Since the primary goal of this study is to assess the impact of vaccines on disease progression by examining their influence on transition rates, the following step involves defining the effect of vaccination.

2.1.5 Prevention trough vaccination

Declines in $FEV_1\%$ are mainly due to the interaction between chronical bacterial infections and the induced inflammatory immune response. The primary pathogen in the CF lung is the gram-negative bacterium *P. aeruginosa*. Therefore, preventing infections with *P. aeruginosa* through vaccination can positively influence disease progression, potentially increasing life expectancy. Numerous studies have explored the efficacy of various *P. aeruginosa* vaccines. However, despite over 50 years of research, clinical vaccine development for *P. aeruginosa* has largely been unsuccessful [20]. Nonetheless, with continued advancements in vaccine development—such as the recent PLGA vaccine using alginate antigen—a successful vaccine is likely to emerge in the future. In this study, we assume the existence of ten different vaccines, each reducing the transition rate by steps ranging from 10% to 100%, and evaluate their impact on lung function over time.

Thus far, we have clarified all aspects of the system of differential equations. We started by defining the lung function groups and their initial values, then determined the transition rates and the effect of vaccination, we calculated the mortality rates for each group, and finally defined the rates of aging. With each key element of the system described to enhance our understanding, the next step is to focus on solving the problem.

2.2 First-order linear differential equation system

We explained that our system can be written as a non-homogeneous system of first-order differential equations:

$$\frac{dV}{dt} = AV + b$$

where $A \in M_n(\mathbb{R})$ is a constant n x n-matrix with elements a_{ij} and $V, b \in \mathbb{R}^n$.
The first step in solving a non-homogeneous differential equation is to find the solution to its corresponding homogeneous equation. In Section A.1.1 of the Appendix, we provide the proof showing that the solution to the homogeneous initial value problem is given by $V(t) = e^{tA}V_{i_0}$, with the initial condition $V_i(0) = V_{i_0}$. A crucial element in computing e^{tA} is the fundamental matrix.

2.2.1 Fundamental matrix

The fundamental matrix of a system of n homogeneous first order differential equations is defined as the n x n matrix $\Psi(t)$ whose columns are linear independent solutions of the system. A detailed definition, along with the conditions for a matrix to qualify as a fundamental matrix, can be found in Section A.1.2 of the Appendix. We finished the proof with the observation that e^{tA} is a fundamental matrix of a first order system of differential equations. This result will be used in the next theorem.

Theorem 2.1 *Let $\Psi(t)$ be a fundamental matrix of the system of differential equations. Then the flow, i.e. the exponential matrix e^{tA} is given by*

$$e^{tA} = \Psi(t)[\Psi(0)]^{-1}.$$

In other words, the product of the fundamental matrix of the given system with its inverse at $t = 0$ must equal e^{tA}.

Proof. The solutions of the homogeneous system form a linear subspace within the linear space of all functions $\mathbb{R} \to \mathbb{R}^n$, closed under pointwise addition and scalar multiplication. In other words, every solution of the system of differential equations is given by a linear combination of the columns of the fundamental matrix:

$$V(t) = \sum_{i=1}^{n} a_i V^{(i)}(t) = \Psi(t)a$$

with $a = (a_1, a_2, ..., a_n) \in \mathbb{R}^n$.

For example, let $V^{(1)}(t)$, $V^{(2)}(t)$ be linear independent solutions of the system, then $\alpha V^{(1)}(t) + V^{(2)}(t)$, $\alpha \in \mathbb{R}$ is also a solution

$$\begin{aligned}\frac{d}{dt}(\alpha V^{(1)}(t) + V^{(2)}(t)) &= \frac{d}{dt}\alpha V^{(1)}(t) + \frac{d}{dt}V^{(2)}(t) \\ &= \alpha A V^{(1)}(t) + A V^{(2)}(t) \\ &= \alpha A \left(V^{(1)}(t) + A V^{(2)}(t) \right).\end{aligned}$$

Now, consider $\Psi(t)$ and $\Phi(t)$ two fundamental matrices of the system of differential equations. Then, since every solution of the system is given by a linear combination of the columns of the fundamental matrix, each column of $\Phi(t)$ can be written as a linear combination of the columns of $\Psi(t)$. I.e. there exists vectors c_j, $j = 1, ..., n$ such that

$$\varphi^{(j)}(t) = c_j^1 V^{(1)}(t) + c_j^2 V^{(2)}(t) + ... + c_j^n V^{(n)}(t)$$

Consequently, let C be a matrix with columns $c_1, c_2, .., c_n$ where $c_j = (c_j^1, c_j^2, .., c_j^n)^T$, then $\Phi(t) = \Psi(t)C$.

We proved e^{tA} is a fundamental matrix of the system, thus $e^{tA} = \Psi(t)C$. Setting $t = 0$ gives $I = \Psi(0)C$, which implies $C = [\Psi(0)]^{-1}$. Therefore, $e^{tA} = \Psi(t)[\Psi(0)]^{-1}$ [18, 19]. □

The last step remaining in the computation of e^{tA} is to determine the equations for $V^{(1)}(t), V^{(2)}(t), ..., V^{(n)}(t)$, the columns of $\Psi(t)$. The procedure of computing $V^{(i)}(t)$ is reported in Section A.1.2 of the Appendix. With this, we have reached the stage where we can compute e^{At}. This allows us to proceed with solving the non-homogeneous differential equation.

2.2.2 Non-homogeneous solution

Theorem 2.2 *Let $A \in M_n(\mathbb{R})$ be a matrix and $b : \mathbb{R} \to \mathbb{R}^n$ a continuous vector function, then for all $v \in \mathbb{R}^n$ the solution of the initial value problem*

$$\frac{dV}{dt} = AV + b(t), \quad V_i(0) = V_{i_0}$$

is given by

$$V(t) = e^{tA}V_{i_0} + \int_0^t e^{(t-\varphi)A} b(\varphi) d\varphi$$

Proof. We have already established that e^{At} serves as a solution to the homogeneous system. This result will be used to formulate an appropriate substitution

$$V(t) = e^{tA} c(t)$$

where $c(t)$ is a vector function to be determined, and $c(0) = V_{i_0}$. Taking the derivative of this substutution we get

$$\frac{dV}{dt} = Ae^{tA}c(t) + e^{tA}\frac{dc}{dt}$$

Starting with the equation $\frac{dV}{dt} = AV + b(t)$, we can conclude $e^{tA}\frac{dc}{dt} = b(t)$. It follows that $\frac{dc}{dt} = \left(e^{tA}\right)^{-1} b(t)$. And since,

$$e^{-tA}e^{tA} = e^{-tA+tA} = e^0 = I$$

we conclude that the inverse of e^{tA} is e^{-tA} and thus we get $\frac{dc}{dt} = e^{-tA}b(t)$.

To determine $c(t)$, we solve the differential equation, using the initial condition $c(0) = V_{i_0}$. According to The Fundamental Theorem of Calculus, the solution is given by

$$c(t) = V_{i_0} + \int_0^t e^{-\varphi A}b(\varphi)d\varphi.$$

We are left with substituting this result into the equation we defined for $V(t)$

$$\begin{aligned} V(t) &= e^{tA}c(t) \\ &= e^{tA}\left(V_{i_0} + \int_0^t e^{-\varphi A}b(\varphi)d\varphi\right) \\ &= e^{tA}V_{i_0} + \int_0^t e^{(t-\varphi)A}b(\varphi)d\varphi \end{aligned}$$

which is what we aimed to prove [18, 19]. □

3 Results

We have provided a detailed description of the model's key components to enhance an understanding of their contributions to the system. We used data from The UK Cystic Fibrosis Registry to determine the initial values $V_i(0)$ by computing the sizes of the lung function groups at $t = 0$. Furthermore, we calculated the mortality rates and determined the aging process for each group. We also defined the transition rates and explored how vaccination could affect them. Lastly, we presented a step-by-step proof of how to derive the solution for a non-homogeneous system of differential equations, such as the one in our study. In this section, we will use the provided information to create a Python code that calculates the solution to our problem. Our goal is to determine the solution for $V(t)$ both in the absence of vaccination and with vaccinations of varying efficacy. Through this, we aim to visualize the impact of vaccination on the changes in group size of the lung function groups over time.

3.1 Programming

We wrote a script in Jupyter notebook that simulates and visualizes the time evolution of $V_i(t)$ with and without vaccination. We will give a brief explanation, since the entire script can be found in Section A.3 of the Appendix.

The script imports three essential libraries: **numpy** for numerical operations, **scipy** for advanced mathematical functions (specifically matrix exponentiation in this case),

and **matplotlib.pyplot** for plotting the results. We started by computing the matrix A, using the function **a** to ensure it satisfies our model. Next, we developed two functions: **create_V** and **calc_results**. **Create_V** computes the time-dependent solution $V(t)$ of the system. It modifies the input matrix A by appending the vector b and preparing it for matrix exponentiation. The function returns a lambda function that calculates $V(t)$ while incorporating the initial values V_0. The second function **calc_results** calculates the evolution of $V(t)$ over a specified time range by evaluating the function at multiple time points between t_start and t_stop.
The script generates 11 models for every age group - one without vaccination and 10 with vaccines of varying efficacy - using **create_V** for both the no-vaccination and vaccination cases. It then computes the results for each case over a time range of 0 to 50 years, divided into 50 time steps. Finally, the number of individuals in each lung function group within a specific age group is plotted over time for all scenarios, allowing a comparison between them.

3.2 Visualization of solutions

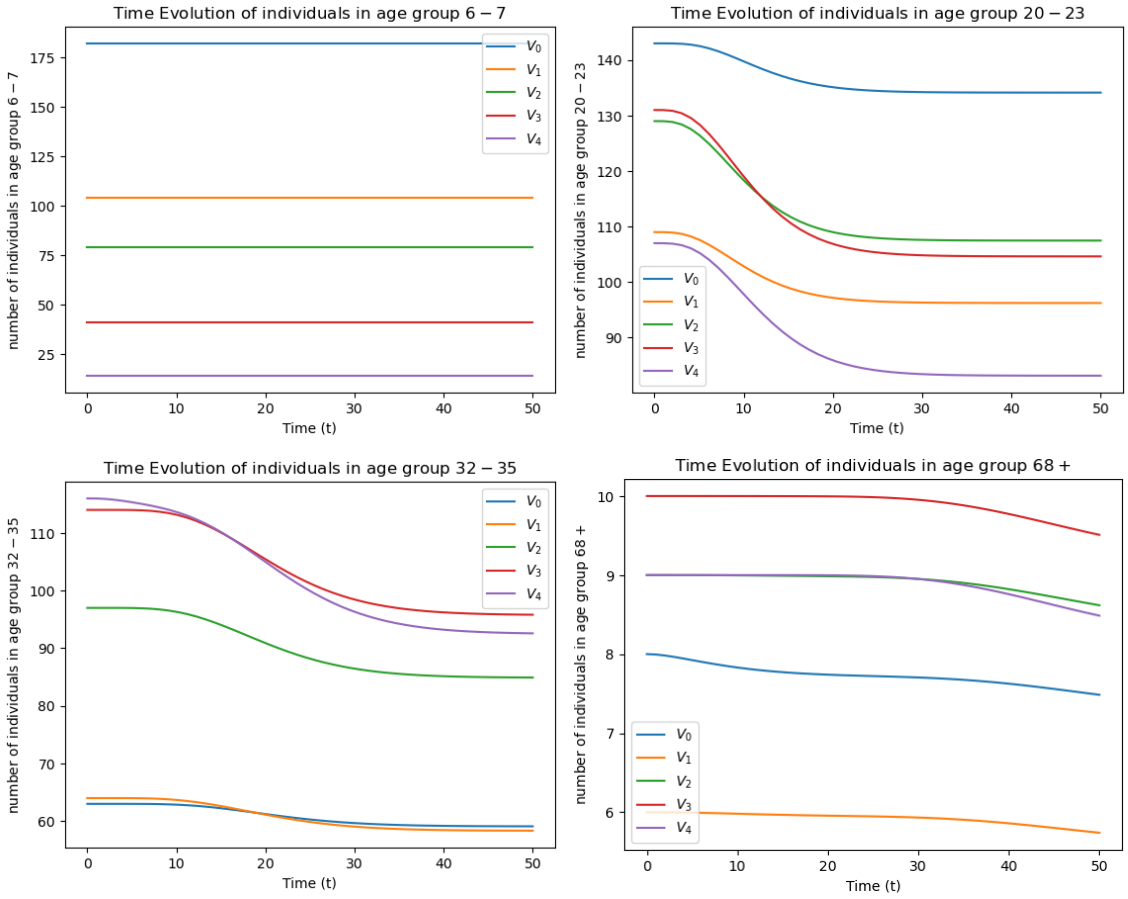
Executing the code generates 108 solutions $V(t)$, corresponding to six lung function groups, $V_0 - V_5$, across all 18 age groups. This means that for the unvaccinated scenario alone, we already have 108 solutions. To compare this outcome with the vaccinated scenarios, we run the code again using an adjusted A matrix incorporating the modified transition rates, which produces another set of 108 solutions per vaccine. Given the large number of solutions, it is impractical to analyze each one individually, necessitating a selection process.
Our goal is to focus on the groups we consider most relevant in providing a meaningful answer to the research question. We executed the code for all age groups and various vaccines, noticing a consistent pattern within groups receiving the same vaccine. The selection of the vaccines we will evaluate is based on the idea that examining the best-case scenario and the worst-case scenario, and one in between would provide a comprehensive overview of the impact of vaccination. Thus, the three vaccination scenarios we will analyze are: one with no vaccine, one with a vaccine of 50% efficacy that reduces the transition rate by half, and one with 100% efficacy that eliminates the transition rate entirely. For each scenario, we will illustrate the impact of vaccination across four different age groups. These groups were selected based on specific assumptions. First, we chose the youngest group, as vaccinating early in life could potentially have a significant impact on these patients. Second, we included the 20-23 age group, as this represents the median age of individuals currently living with CF [10], allowing us to assess whether vaccination remains relevant for the majority of patients. Third, we selected adults aged 32-35, as this group is well-represented in the data registry and, with advances in life expectancy, may reflect the median age of people with CF in the near future. Finally, we included the eldest group to evaluate whether vaccination can still provide benefits for these patients.

3.2.1 Absence of vaccination

In Section 2.1.4, we explained how the transition rates -the rates at which individuals move from the previous lung function group or leave to the next - were determined. We assumed that, in the absence of a vaccine, there will not be significant changes in the number of individuals in each lung function group. New individuals will enter the cohort, and those already in the cohort will age and experience disease

progression at a consistent rate. Additionally, individuals will pass away each year, with the number of deaths expected to balance out the number of new entrants into the cohort. Therefore, without vaccination, a steady-state scenario will occur. This means that, over time, the distribution of individuals across specific age and lung function groups is expected to remain relatively stable. In other words, the initial sizes calculated for each group in the previous section are projected to remain constant over the years. This is illustrated in the graphs, where all the lines are flat, showing almost zero slope.

Figure 4: Absence of a vaccine



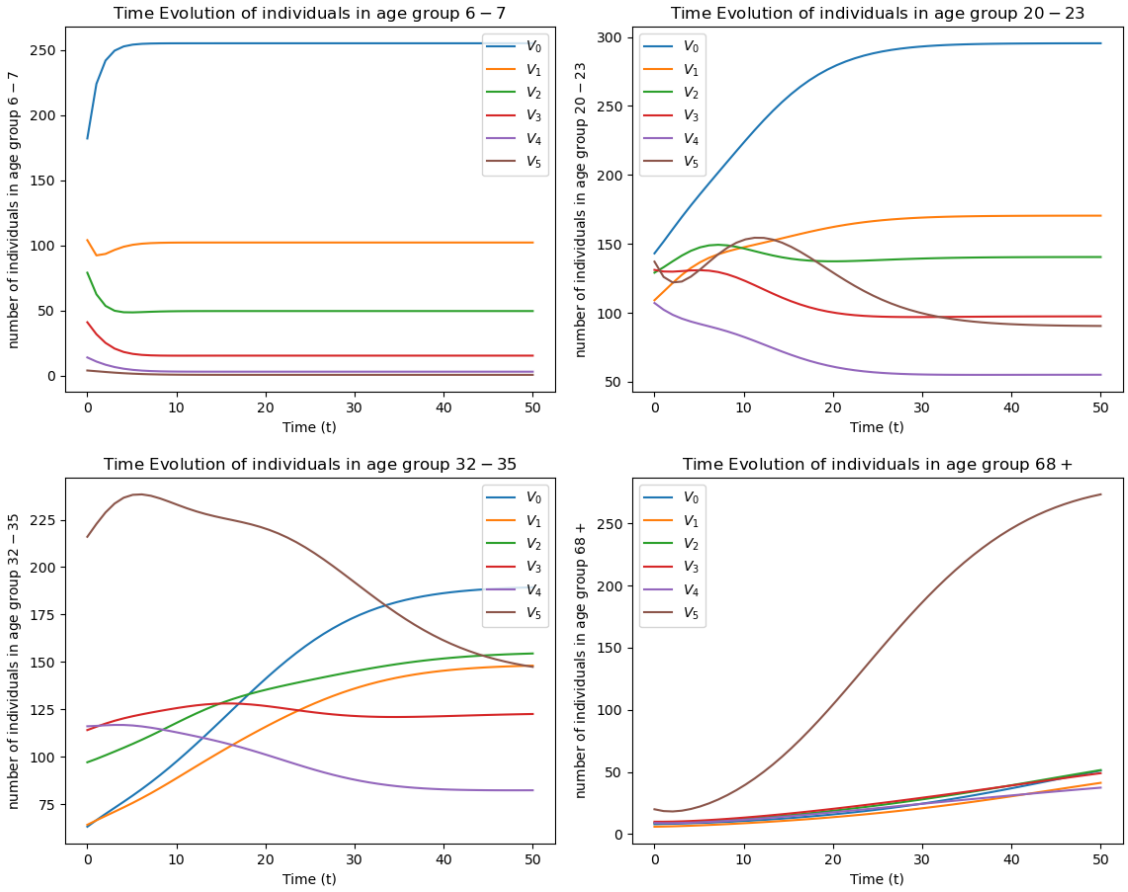
Note that V_5 is not included in the graph. As mentioned in section 2.1.4 every transition rate corresponding to V_5 is set to zero, as it is impossible to move to a poorer lung function group. However, a limitation of this approach is that the size of V_5 changes over time, which contradicts our goal of achieving a steady-state. Therefore, it is not meaningful in the context of the no-vaccination scenario. Another observation is that for the age groups 20–23 and 32-35, there is a slight decline in group size. This is due to the transition rates calculated from the code being negative and, as outlined in Section 2.1.4, subsequently being set to zero.

3.2.2 Vaccine with 50% efficacy

A vaccine with 50% efficacy reduces the transition rate by a factor of 0.5, leading to a larger cohort over time as individuals transition more slowly between groups. As illustrated in the figure below, we observe that in the 6-7 age group the vaccine has a positive effect. Specifically, the size of the best lung function group V_0 increases,

V_1 stays stable and the other groups decrease in size. This indicates an overall improvement in lung function as a result of vaccination. For the 20–23 age group, we observe a similar pattern: a relatively large increase in the size of V_0, a smaller but still notable increase in V_1 and a very slight increase in V_2. Both V_4 and V_5 decrease, with V_5 showing a greater decrease than V_4. These observations also highlight a positive effect of vaccination, with a greater absolute impact compared to the previous age group.

Figure 5: Vaccine with 50% efficacy



In the third graph, representing the 32–35 age group, we observe a similar outcome to the previous graphs: an increase in the size of the healthier groups and a decrease in the poorer groups as a result of vaccination. The slight increase in the size of V_5 can be attributed to the fact that vaccination causes more people to remain in the overall cohort, leading to an increase in the total number of individuals. Over time, however, individuals tend to distribute more across the other lung function groups.

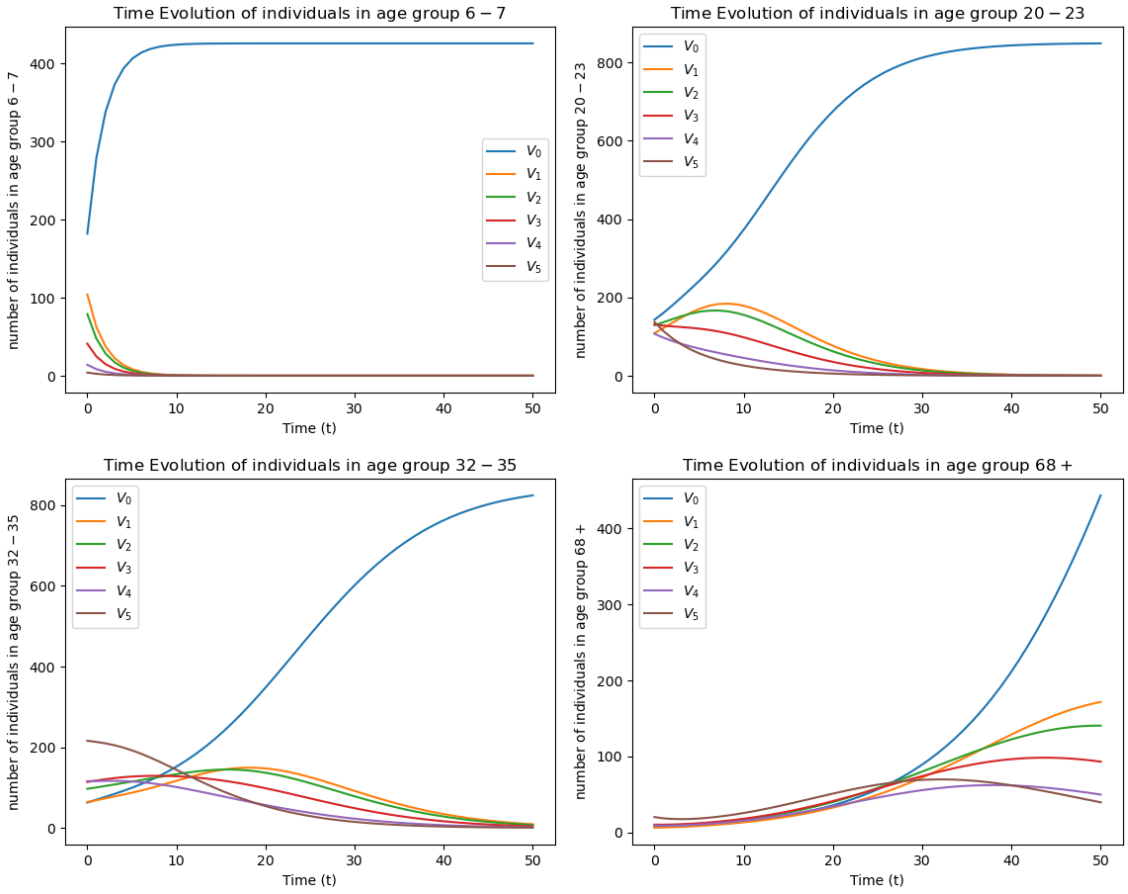
The eldest group presents a slightly different pattern. As mentioned, with vaccination individuals are less likely to move out of their group, although there still is an annual inflow. Consequently, the final stage that individuals can reach—the oldest age group with the poorest lung function—will see an increase in size. For the other lung function groups, we do not observe a significant effect of vaccination, as their sizes remain quite stable. However, there is a slight increase after 50 years, although the absolute effect is minimal. This can be attributed to individuals staying longer in the 68+ group due to a slower mortality rate.

3.2.3 Vaccine with 100% efficacy

As previously stated, a 100% effective vaccine completely eliminates the transition rate, resulting in $r_i^{\text{age}} = 0$ for all i and all age groups. This implies that individuals who are initially in a specific age and lung function group at $t = 0$ can only exit that group through aging. As a result, individuals can progress to higher age groups but cannot transition to poorer lung function groups. Furthermore, all new individuals enter the cohort in the first age group with the healthiest lung function, V_0^{6-7}, meaning their only possible movement is to V_0 of higher age groups.

This is illustrated as follows: in all age groups, except the youngest one, the lung function groups V_1 to V_5 experience growth until it reaches a peak. This growth is driven by the transition of individuals from preceding age groups. Therefore, there is no growth in the 6-7 age group, as no one can enter it. This pattern is consistent across all age groups, although the timing of the peak varies. In higher age groups, the peaks occur later because there is a longer period of time during which individuals can enter these groups, as people transition from multiple younger groups. After reaching the peak, the sizes gradually decrease until they ultimately reach zero after a sufficient number of years.
The change in sizes for V_0 can be explained by the constant inflow of 213 individuals each year into the group V_0^{6-7}. So, unlike the other lung function groups, this group will never die out, as there is always an influx from above. In the absence of a vaccine, newly entered individuals would not only move to higher age groups but also to poorer lung function groups, resulting in a more evenly distributed population across groups. However, with the introduction of the vaccine, we observe significant growth in V_0 across all age groups. As expected, the growth begins earlier in the younger age groups and later in the older ones.

Figure 6: Vaccine with 100% efficacy



In conclusion, the relative effect of the 100% effective vaccine is consistent across all age groups. However, the absolute effect varies. Since the highest age group contains significantly fewer individuals compared to the others, the absolute benefit of vaccination is smaller in this group.

3.3 Impact on life expectancy

To provide a precise answer to our research question, we need to quantify the effect of all the scenarios described in the previous section. We will accomplish this by calculating the vaccines' impact on life expectancy. To determine the average age at which individuals die, we will use the following equation:

$$\overline{\text{age}} = \frac{\sum_{j=0}^{17} \sum_{i=0}^{5} \mu_i^j V_i^j a_i}{\sum_{j=0}^{17} \sum_{i=0}^{5} \mu_i^j V_i^j}$$

The numerator denotes the calculation of the sum across all lung function groups within each age group, where each term is the product of the size of a lung function group, the corresponding mortality rate, and the mean age of individuals within this group. The denominator is the the sum over all lung function groups within each age group, where each term is the product of the size of a lung function group and the corresponding mortality rate. By dividing these values, we obtain the predicted average life expectancy. The results will be displayed in a plot illustrating life expectancy as a function of vaccine efficacy. For each vaccine efficacy, we calculated

the predicted life expectancy of individuals diagnosed with CF, 50 years from $t = 0$ (2072).

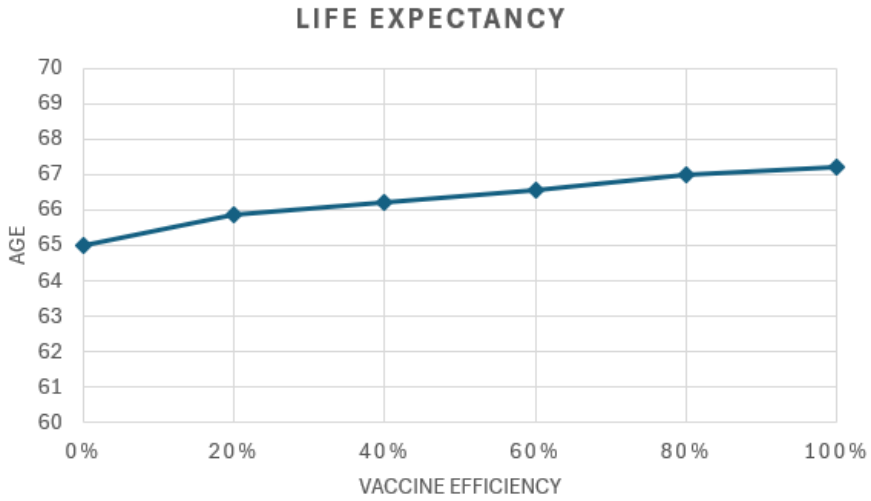


Figure 7: Life expectancy as a function of vaccine efficacy

The graphs show that an increase in vaccine efficacy results in a rise in life expectancy. However, the increase in life expectancy is not significant, as a vaccine with 100% efficacy only improves life expectancy by two years.
It should be noted that, since we calculated the predicted life expectancy in 2072 for individuals with CF, starting vaccination with 100% efficacy now means we are still dealing with individuals in poorer lung function groups. In 50 years, individuals who are currently in groups such as V_4 will remain in V_4, but in a 50-year older age group. However, after 100 years, only individuals in the V_0 group will remain, and we will observe a life expectancy equivalent to that of individuals without CF.

4 Discussion

Since CF was first described in 1938, remarkable progress has been made in extending patients' life expectancy. By 1978, the estimated median survival in the United States had increased to 11 years [21]. Today, based on the most recent data, the median predicted survival age for individuals born between 2019 and 2023 with CF has risen to 65 years. These remarkable advancements in CF treatment highlight the substantial progress made over the years [10]. However, our model assumes that transition rates remain constant over time. Specifically, we set $t = 0$ as the time the registry was introduced, used the group sizes from 2022, and assumed these sizes would remain constant over time in the absence of vaccination. In reality, group sizes do change over time due to advancements in treatment, as evidenced by the increasing survival age. Therefore, the assumption of a steady state at $t = 0$ for calculating transition rates do not accurately reflect real-world conditions. Additionally, individuals aged 50 or older in the data registry utilized for this study were born during a period when survival rates were much lower. They did not have access to the advanced treatments available compared to those aged 30 in the registry, and certainly not to the therapies the youngest age groups will benefit from in the future. As a result, the older age groups have very small sample sizes, making it difficult to

obtain significant results due to the minimal absolute differences observed.

In calculating the transition rates, we had to make two assumptions. First, we assumed that transition rates could not be negative; therefore, any negative values were set to zero. However, as seen in the results for the no-vaccination scenario, this assumption led to a contradiction with our hypothesis, namely that the size of the groups would remain constant over time. It could be meaningful to use the least squares method to determine which non-negative values of r provide group sizes that are closest to the observed ones. This is a statistical method used to find the best fit for a set of data points by minimizing the sum of the deviation of points from the plotted curve. In our case, we could use both earlier data and data up to the present to calculate the transition rates using the same method we applied previously.By observing how these rates change over time, we can apply the least squares method to identify a pattern and obtain the best approximation of both the current and future rates, adjusting them as needed.
The second assumption we made was that all transition rates associated with V_5 were set to zero, as it is impossible to transition to a poorer lung function group. This limitation makes it challenging to draw meaningful conclusions about this group in our study. A better approach would be to include more lung function groups. While this would not completely resolve the issue—since there would still be a group that cannot be thoroughly analyzed—it would no be remarkable if this group corresponded, for example, to an FEV_1% of 20%. This is because most individuals in this condition would not survive. Therefore, for the medical purpose of this study, this adjustment would not significantly impact the results.

4.1 Conclusion

The results suggest that vaccination can be beneficial for individuals of all ages. In 50 years from now, a 100% effective vaccine can increases life expectancy by 2 years. Even with a vaccine efficacy of 50%, we observe a positive outcome, as more people remain in the healthier lung function groups. However, one could question whether the increase in life expectancy is substantial enough to justify starting vaccination.

4.2 Outlook and recommendations

In this model we used the $\text{FEV}_1\%$ value as a predictor of life expectancy in response to vaccination. However, real-world factors such as genetic variability, coexisting conditions, or environmental influences, might influence this as well. Therefore, we recommend further research to incorporate these variables into the model to capture more accurately the complexity of real-world scenarios.

Another interesting point for further research can be to dive into the cost-effectiveness of the vaccine. Not only the clinical efficacy of vaccination should be considered, but also the economic consequences of producing the vaccines. The costs of vaccination depend on factors such as the price of the vaccine and the logistical challenges associated with its widespread distribution. These costs can be compared to the additional years of life that vaccination may provide. It would be valuable to assess the financial benefits of extending life expectancy by determining the monetary value of an extra year of life. A potential research question could be: What is the economic value of an additional year of life, and does it justify the cost of the vaccine?

A Appendix

A.1 Solution of first-order linear differential equation system

A.1.1 Homogeneous solution

Theorem A.1 *Let $A \in M_n(\mathbb{R})$ be a matrix. For every initial condition $V_i(0) = V_{i_0}$ the solution to the initial value problem*

$$\frac{dV}{dt} = AV, \quad V_i(0) = V_{i_0},$$

is given by

$$V(t) = e^{tA} V_{i_0}.$$

Proof. We first need to show that the power series $e^{tA} := \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k$ converges with respect to the norm in the Banach space. This convergence guarantees that the matrix exponential is well-defined for all values of t. If the series failed to converge, the infinite summation would not produce a finite matrix, making the matrix exponential invalid.

Lemma A.1 *The power series $e^{tA} := \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k$ in $M_n(\mathbb{R})$ converges for all $t \in \mathbb{R}$ with respect to the operator norm, which is defined as*

$$||A|| := sup_{y \neq 0} \frac{||Ay||}{||y||}$$

where we used the Euclidean norm on $\mathbb{R}^n$.

Proof. From the definition of the operator norm we conclude $||Ay|| \leq ||A||||y||$ for all $y \in \mathbb{R}^n$. Additionally, we use the fact that if a series converges absolutely, it is convergent as well. For all $t \in \mathbb{R}$, the series

$$\begin{aligned} \sum_{k=0}^{\infty} \left|\left| \frac{t^k}{k!} A^k \right|\right| &\leq \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \left|\left| A^k \right|\right| \\ &\leq \sum_{k=0}^{\infty} \frac{|t|^k}{k!} ||A||^k \\ &= e^{|t|||A||} < \infty \end{aligned}$$

converges, so the series for e^{tA} is absolutely convergent and, therefore, convergent [17]. □

We will now use this result to prove the theorem. We define $e^{tA} = F(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k$ and differentiate

$$\begin{aligned} \frac{dF(t)}{dt} &= \sum_{k=0}^{\infty} \frac{k t^{k-1}}{k!} A^k \\ &= A \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} A^{k-1} \\ &= A \sum_{m=0}^{\infty} \frac{t^m}{m!} A^m \\ &= AF(t) \end{aligned}$$

In other words

$$\frac{d}{dt}e^{At} = Ae^{At}$$

is a well-defined solution for the differential equation $\frac{dv}{dt}Av$ [16]. Lastly, we insert the initial value. Since $F(0) = I$, the identity matrix, we finish the proof with the conclusion that the solution of the initial value problem is $V(t) = F(t)V_{i_0} = e^{tA}V_{i_0}$. [18] □

Alternative proof. Note that it is possible to divide both sides of the equation by V

$$\frac{\frac{dV}{dt}}{V} = \frac{AV}{V}$$

And observe

$$\frac{\frac{dV}{dt}}{V} = \frac{d}{dt}ln|V(t)|$$

As a result, the equation $\frac{dV}{dt} = AV$ can be written as

$$\frac{d}{dt}ln|V(t)| = A$$

Now we can integrate both sides, where we use The Fundamental Theorem of Calculus to obtain

$$ln|V(t)| = \int A\ dt\ + c_1$$

where c_1 is the constant of the integration. After taking the exponential on both sides we derive

$$|V(t)| = e^{\int A\ dt\ + c_1} = c_2 e^{\int A\ dt}$$

since $A \in M_n(\mathbb{R})$, $e^{\int A\ dt}$ is positive and it will follow that c_2 is also positive, thus

$$V(t) = c_2 e^{\int A\ dt}$$

Finally, we note $\int A\ dt = At$ and insert the initial value $c_2 = V_{i_0}$. Using the same explanation as before, namely that $e^{0A} = I$, the identity matrix, we finish the proof with the conclusion that the solution of the initial value problem is $V(t) = V_{i_0}e^{tA}$ [19].

A.1.2 Fundamental matrix

Definition A.1 Let $t \mapsto V^{(i)}(t)$, $1 \leq i \leq n$, be linear independent solutions of the system of differential equations, specifically

$$\sum_{i=1}^{n} a_i V^{(i)}(t) = 0$$

for all $t \in \mathbb{R}$ if and only if $a_i = 0$ for all $i = 1, 2, ..., n$. Then, the fundamental matrix is the n x n matrix $\Psi(t)$ with a set of n linear independent solutions $V^{(1)}(t), V^{(2)}(t), ..., V^{(n)}(t)$ as columns:

$$\Psi(t) = \begin{pmatrix} V_1^{(1)}(t) & V_1^{(2)}(t) & ... & ... & V_1^{(n)}(t) \\ V_2^{(1)}(t) & V_2^{(2)}(t) & ... & ... & V_2^{(n)}(t) \\ ... & & & & \\ ... & & & & \\ V_n^{(1)}(t) & V_n^{(2)}(t) & ... & ... & V_n^{(n)}(t) \end{pmatrix}.$$

Lemma A.2 *A matrix $\Psi(t)$ is a fundamental matrix of a first order linear non-homogeneous system if and only if $\frac{d\Psi}{dt} = A\Psi$ and $det\Psi(0) \neq 0$.*

Proof. Let $V^{(1)}(t), V^{(2)}(t), ..., V^{(n)}(t)$ represent the columns of $\Psi(t)$. Then

$$\frac{d\Psi}{dt} = \left(\frac{d}{dt}V^{(1)}(t), \frac{d}{dt}V^{(2)}(t), ..., \frac{d}{dt}V^{(n)}(t)\right)$$

and since $V^{(1)}(t), V^{(2)}(t), ..., V^{(n)}(t)$ are solutions of the system of differential equations, we observe

$$\begin{aligned}\frac{d}{dt}V^{(1)}(t), \frac{d}{dt}V^{(2)}(t), ..., \frac{d}{dt}V^{(n)}(t) &= \left(AV^{(1)}(t), AV^{(2)}(t), ..., AV^{(n)}(t)\right) \\ &= A\left(V^{(1)}(t), V^{(2)}(t), ..., V^{(n)}(t)\right)\end{aligned}$$

So we conclude $\frac{d}{dt}\Psi(t) = A\Psi(t)$. Furthermore, we stated that the solutions $V^{(1)}(t)$, $V^{(2)}(t), ..., V^{(n)}(t)$ are linear independent. According to Definition A.1, this implies $V^{(1)}(0), V^{(2)}(0), ..., V^{(n)}(0)$ are linear dependent. These vectors, however, are linearly independent if and only if $\det\Psi(0) \neq 0$ [19]. □

As a result, we observe that e^{tA} is a fundamental matrix of a first order system of differential equations, since we proved $\frac{d}{dt}e^{At} = Ae^{At}$ and $e^{0A} = I$, so $\det(e^{0A}) = 1 \neq 0$.

The next theorem explains how to determine the columns $V^{(1)}(t), V^{(2)}(t), ..., V^{(n)}(t)$ of the fundamental matrix $\Psi(t)$.

Theorem A.2 *If all eigenvalues of $A \in M_n(\mathbb{R})$ are simple, i.e. have an algebraic multiplicity of 1, then the fundamental matrix is defined as the matrix with the columns*

$$V^{(i)}(t) = e^{\lambda_i x}\eta^{(i)},\ 1 \leq i \leq n$$

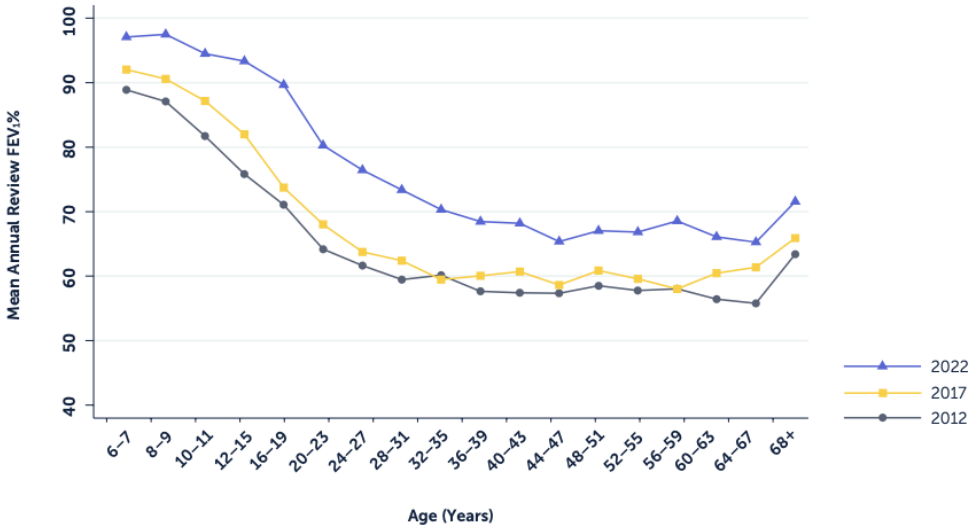
with $\lambda_1, ..., \lambda_n \in \mathbb{R}$ the eigenvalues of A and their corresponding eigenvector $\eta^{(i)} \in \mathbb{R}^2$.

Proof. According to the definition of an eigenvector, all eigenvectors corresponding to distinct eigenvalues are linearly independent. So if all eigenvalues are simple, it follows that the functions $V^{(i)}(t)$ as defined above are linearly independent for all $t \in \mathbb{R}$. On top of that, $V^{(i)}(t)$ are solutions of the differential equation for all $1 \leq i \leq n$

$$\frac{d}{dt}V^{(i)}(t) = e^{\lambda_i x}\lambda_i\eta^{(i)} = e^{\lambda_i x}A\eta^{(i)} = Ae^{\lambda_i x}\eta^{(i)} = AV^{(i)}(t)$$

where we used the property of eigenvalues and their eigenvectors; $\lambda\eta = A\eta$ [18]. □

A.2 Data



	2012		2017		2022		
Age (years)	n	FEV₁% :Mean (SD)	n	FEV₁% :Mean (SD)	n	FEV₁% :Mean (SD)	p-values (t-test*)
6-7	375	88.9 (15.9)	526	92.0 (16.0)	425	97.1 (15.8)	<0.001
8-9	399	87.1 (15.8)	511	90.6 (15.9)	449	97.5 (14.7)	<0.001
10-11	392	81.7 (16.5)	488	87.2 (15.7)	528	94.5 (15.5)	<0.001
12-15	914	75.8 (18.3)	862	82.0 (17.8)	960	93.3 (16.0)	<0.001
16-19	920	71.1 (21.8)	887	73.7 (21.5)	737	89.7 (17.9)	<0.001
20-23	915	64.2 (23.6)	945	68.0 (23.3)	757	80.3 (22.4)	<0.001
24-27	747	61.6 (23.8)	860	63.7 (23.0)	820	76.5 (22.6)	<0.001
28-31	600	59.5 (22.8)	725	62.4 (23.8)	736	73.4 (23.0)	<0.001
32-35	394	60.1 (23.3)	609	59.5 (23.7)	670	70.3 (22.5)	<0.001
36-39	256	57.6 (22.5)	436	60.0 (24.3)	496	68.5 (22.8)	<0.001
40-43	241	57.4 (22.7)	276	60.7 (24.0)	441	68.2 (23.8)	<0.001
44-47	167	57.3 (25.5)	237	58.6 (23.4)	265	65.4 (23.7)	0.001
48-51	111	58.5 (23.8)	172	60.9 (25.4)	214	67.0 (22.9)	0.012
52-55	66	57.8 (26.6)	119	59.6 (25.4)	156	66.8 (24.1)	0.016
56-59	37	58.0 (22.3)	68	58.0 (24.1)	109	68.6 (23.6)	0.005
60-63	18	56.4 (26.9)	45	60.5 (23.1)	73	66.1 (22.8)	0.196
64-67	17	55.8 (23.1)	28	61.4 (22.6)	53	65.3 (25.4)	0.497
68+	17	63.4 (25.6)	41	65.9 (28.3)	62	71.6 (24.5)	0.283
<16	2080	81.5 (18.0)	2387	87.1 (17.1)	2362	95.1 (15.7)	-
≥16	4506	62.8 (23.6)	5448	64.5 (23.9)	5589	74.6 (23.5)	-
<18	2561	79.8 (18.8)	2802	85.5 (17.9)	2735	94.6 (15.8)	-
≥18	4025	61.7 (23.7)	5033	63.5 (23.9)	5216	73.4 (23.5)	-

Figure 8: Annual review $FEV_1\%$ predicted over time in patients aged six years and older who have not had a lung transplant [7].

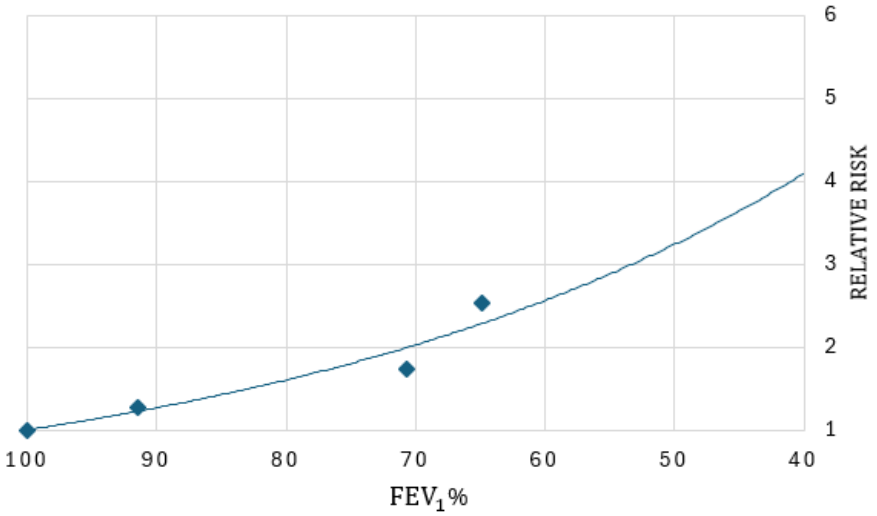


Figure 9: Relative mortality risks plotted against their corresponding $FEV_1\%$ values.

A.3 Jupyter notebook script

A.3.1 Computing the transition rates

```
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt

n = 18   # Number of age groups
m = 6    # Number of disease states

mu = [0.000065, 0.000075, 0.000085, 0.0000975, 0.0002175, 0.00027, 0.0003375, 0.000385,
0.0004875, 0.00064, 0.0009025, 0.0012475, 0.00181, 0.0027575, 0.0041725, 0.006085,
0.009375, 0.1129175]
#Mortality rate for each age group
rr = [1, 1.16, 1.49, 1.87, 2.36, 3.37]
#Relative risk for each disease state
u = [0.5, 0.5, 0.5, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25,
0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0]
# Aging rate for each group

V_0 = np.array([182, 104, 79, 41, 14, 4, 194, 118, 84, 39, 11, 2, 192, 133, 112, 62, 23, 7,
324, 236, 205, 126, 51, 18, 207, 156, 157, 117, 64, 36, 143, 109, 129, 131, 107, 137, 119,
101, 132, 146, 127, 196, 91, 83, 110, 128, 117, 207, 63, 64, 97, 114, 116, 216, 42, 45, 67,
81, 85, 176, 40, 39, 57, 70, 73, 162, 19, 20, 31, 42, 44, 108, 16, 18, 27, 35, 37, 81, 13,
13, 19, 25, 25, 61, 10, 10, 15, 17, 18, 39, 5, 6, 9, 12, 13, 29, 5, 4, 6, 8, 8, 22, 8, 6,
9, 10, 9, 20])
# Initial values
b = 213

f = np.repeat(mu, m) * np.tile(rr, n) + np.repeat(u, m)
g = np.repeat(u, m)
r_true = np.zeros(n*m)
# Defining the entries of the matrix A

r_true[0] = (b / V_0[0]) - f[0]
# Calculating r_0

for i in range(1, 5):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i]
    # Calculating r_1 - r_4, since every (5 + i*6)th r is 0
    r_true[i] = np.maximum(0, r_true[i])
    # Since r cannot be negative, any negative values are set to 0
```

```
for i in range(6, 11):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])

for i in range(12, 17):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(18, 23):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(24, 29):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(30, 35):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(36, 41):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(42, 47):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(48, 53):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(54, 59):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(60, 65):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(66, 71):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(72, 77):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(78, 83):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(84, 89):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(90, 95):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(96, 101):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])
for i in range(102, 107):
    r_true[i] = ((r_true[i-1] * V_0[i-1]) / V_0[i]) - f[i] + ((g[i-6] * V_0[i-6])/V_0[i])
    r_true[i] = np.maximum(0, r_true[i])

r_true_matrix = r_true.reshape(18, 6)
# make a 18x6 matrix of r
print(r_true_matrix)
```

A.3.2 Solving the problem

```
r= 'vaccination_efficacy' * r_true_matrix
# Modified transition rate due to vaccination

def a(i, j, ii, jj):
# Computes matrix A
    if ii == i and jj == j:
        return -(mu[i] * rr[j] + u[i] + r[i][j])
        # Rate out of compartment
    elif ii == i and jj == j + 1:
```

```
            return r[i][j]
            #Disease progression
        elif ii == i + 1 and jj == j:
            return u[i]
            #Aging rate
        else:
            return 0

mat = []

for age1 in range(n):
    for dis1 in range(m):
        for age2 in range(n):
            for dis2 in range(m):
                mat.append(a(age1, dis1, age2, dis2))
#Loop over all age and disease state combinations

mat = np.array(mat).reshape(n*m,m*n)
mat = mat.transpose()  # Transpose to match the desired structure
#Reshape and transpose the matrix

print(mat)

def create_V(A, b, V_0):
    # creates solution V(t) for a system of the form dV/dt = AV +b
    A_modded = np.c_[A, b]
    # appends b as a new colomn to A
    A_modded = np.r_[A_modded, [np.zeros(A.shape[0] + 1)]]
    # appends a row of zeros to the bottem of the matrix A_modded

    return lambda t:
    (sp.linalg.expm(A_modded * t) @ np.append(V_0, 1))[:-1]
    # (sp.linalg.expm(A_modded * t) computes the matrix exponential e^{t*A_modded}
    # (np.append(V_0, 1)) appends scalar 1 to initial codition vector x_0, to handle
the constant term b in A_modded
    # [:-1] extract all elements except the last one, which corresponds to the dummy
variable b
# create_V generates lambda function - which calculates V(t) for any t, based on
the initial condition V_0, A and b - and returns it.

def calc_results(V, t_start, t_stop, num_steps):
# V is the result from create_V
    T = np.linspace(t_start, t_stop, num_steps)
    # creates an array of evenly spaced values between t_start and t_stop
    # num_step is the total number of values in array
    return T, np.array([V(t) for t in T])
     # returns both an array of time points in which the solution are evaluated
and an array of solutions of V(t) at the corresponding time

A = mat
b = np.zeros(n*m)
```

```
b[0] = 213
V_0 = [182, 104, 79, 41, 14, 4, 194, 118, 84, 39, 11, 2, 192, 133, 112,
62, 23, 7, 324, 236, 205, 126, 51, 18, 207, 156, 157, 117, 64, 36, 143,
109, 129, 131, 107, 137, 119, 101, 132, 146, 127, 196, 91, 83, 110, 128,
117, 207, 63, 64, 97, 114, 116, 216, 42, 45, 67, 81, 85, 176, 40, 39, 57,
70, 73, 162, 19, 20, 31, 42, 44, 108, 16, 18, 27, 35, 37, 81, 13, 13, 19,
25, 25, 61, 10, 10, 15, 17, 18, 39, 5, 6, 9, 12, 13, 29, 5, 4, 6, 8, 8,
22, 8, 6, 9, 10, 9, 20]
# Importing our model

V= create_V(A, b, V_0)
#solution of V(t)
T, results = calc_results(V, 0, 50, 50)
# starting time of the simulation, stop time and the number of time steps used
todiscretize the time interval

# Creating graphs that display the time evolution of V(t) for a specific age group
and vaccination efficacy.

plt.plot(T_novac[0:1000], results_novac[0:1000 , i:i+6])
plt.xlabel('Time (t)')
plt.ylabel('number of individuals in age group ...')
plt.title('Time Evolution of individuals in age group ...')
plt.legend(['V_0', 'V_1', 'V_2', 'V_3', 'V_4', 'V_5'])
plt.show()
```

References

[1] Cystic Fibrosis Foundation. CF Genetics: The Basics. Retrieved from https://www.cff.org/intro-cf/cf-genetics-basics. Accessed: 4 October 2024

[2] American Lung Association. Cystic Fibrosis Symptoms and Diagnosis. Retrieved from https://www.lung.org/lung-health-diseases/lung-disease-lookup/cystic-fibrosis/symptoms-diagnosis. Accessed: 4 October 2024

[3] American Lung Association. Cystic Fibrosis Symptoms and Diagnosis. Retrieved from https://www.lung.org/lung-health-diseases/lung-disease-lookup/cystic-fibrosis/treating-and-managing. Accessed: 6 January 2025

[4] Jane C. Davies. Pseudomonas aeruginosa in cystic fibrosis: pathogenesis and persistence. Paediatric Respiratory Reviews Volume 3, Issue 2, June 2002, Pages 128-134. Retrieved from: https://doi.org/10.1016/S1526-0550(02)00003-3. Accessed: 4 October 2024

[5] Malhotra SHayes D, Wozniak DJ (2019). Cystic Fibrosis and Pseudomonas aeruginosa: the Host-Microbe Interface. ASM Journals Clinical Microbiology Vol. 32, No. 3. Retrieved from: https://doi.org/10.1128/cmr.00138-18. Accessed: 4 October 2024

[6] Adler FR, Liou TG (2016) The Dynamics of Disease Progression in Cystic Fibrosis. PLoS ONE 11(6): e0156752. Retrieved from: https://doi.org/10.1371/journal.pone.0156752 Accessed: 4 October 2024

[7] Cystic Fibrosis Trust. UK Cystic Fibrosis Registry 2022 Annual Data Report. September 2023. Version 3. Accessed: 4 October 2024

[8] Kemp R, Pustulka I, et al (2021). Relationship between FEV1 decline and mortality in patients with bronchiolitis obliterans syndrome—a systematic literature review. Respiratory Medicine. Retrieved from: https://doi.org/10.1016/j.rmed.2021.106608. Accessed: 4 October 2024

[9] David S, Edwards CW. Forced Expiratory Volume. (Updated 2022 Aug 8). Treasure Island (FL): StatPearls. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK540970/. Accessed: 4 October 2024

[10] Cystic Fibrosis Foundation. Understanding Changes in Life Expectancy. Retrieved from https://www.cff.org/managing-cf/understanding-changes-life-expectancy. Accessed: 4 October 2024

[11] European Respiratory Society. Effect of pulmonary exacerbations on long-term lung function decline in cystic fibrosis. Retrieved from https://publications.ersnet.org/content/erj/40/1/61. Accessed: 28 November 2024

[12] Mortality and cardiovascular and respiratory morbidity in individuals with impaired FEV1 (PURE): an international, community-based cohort study Duong, MyLinh et al. The Lancet Global Health, Volume 7, Issue 5, e613 - e623.

[13] Mathematical Statistics and Data Analysis. Third Edition. John A. Rice. University of California, Berkeley

[14] Levensverwachting; geslacht, leeftijd (per jaar en periode van vijf jaren). Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37360ned/table?fromstatweb. Accessed: 6 January 2025

[15] S. Dirksen (2023). Kanrekening dictaat 2022/2023. Utrecht University.

[16] E.P. van den Ban (2022). Inleiding Analyse. Mathematisch Instituut Universiteit Utrecht.

[17] E.P. van den Ban (2022). Dictaat Functies en Reeksen. Mathematisch Instituut Universiteit Utrecht.

[18] Y.A.Kuzentsov, H.Hanßmann (2024) Basis Differentiaalvergelijkingen. Utrecht University.

[19] Martin Braun. Differential Equations and Their Applications (Fourth edition).

[20] Killough M, Rodgers AM, Ingram RJ. Pseudomonas aeruginosa: Recent Advances in Vaccine Development. Vaccines (Basel). 2022 Jul 8;10(7):1100. doi: 10.3390/vaccines10071100. PMID: 35891262; PMCID: PMC9320790.

[21] McBennett KA, Davis PB, Konstan MW. Increasing life expectancy in cystic fibrosis: Advances and challenges. Pediatr Pulmonol. 2022 Feb;57 Suppl 1(Suppl 1):S5-S12. doi: 10.1002/ppul.25733. Epub 2021 Nov 11. PMID: 34672432; PMCID: PMC9004282.