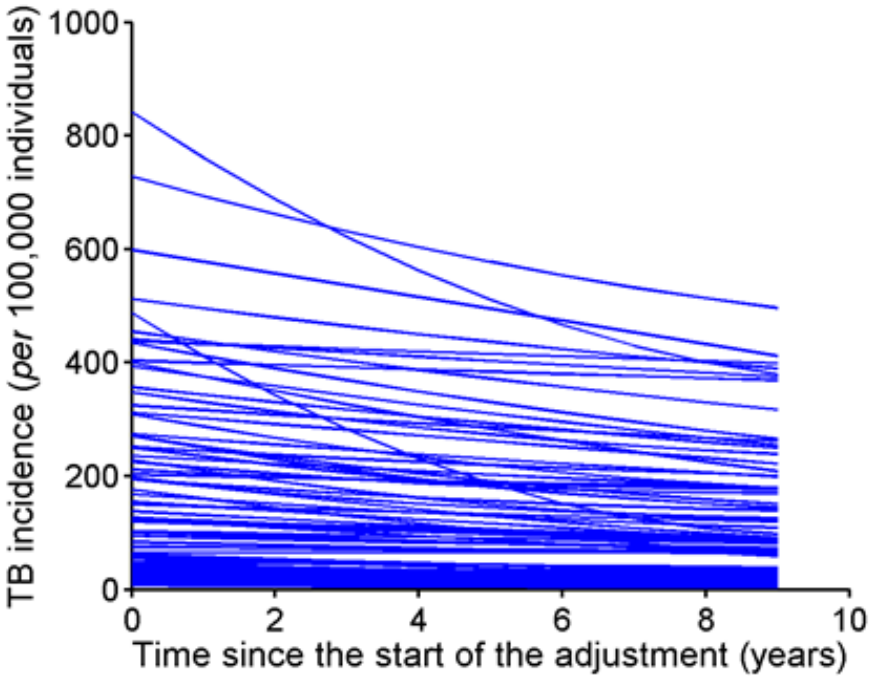
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Agricultura de Barcelona

Analysis and assessment of tuberculosis epidemiology dynamics with a compartment based mathematical model



Treball final de grau Enginyeria de Sistemes Biològics

Edgar Sánchez Prados

Juny, 2016

Direcció: Dra. Àngels Orcau (Agència de Salut Pública de Barcelona) i Dr. Daniel López (Grup de Recerca en Biologia Computacional i Sistemes Complexos, Universitat Politècnica de Catalunya)

Resum

La tuberculosis (TB) és una malaltia infecciosa de transmissió aèria causada per *Mycobacterium tuberculosis*, que ha co-evolucionat amb la humanitat. La seva estratègia per a l'èxit consisteix en romandre desapercebut i actuant lentament. L'Informe Global de la Tuberculosi 2015 de l'Organització Mundial de la Salut (OMS) estima que, el 2014, 9,6 milions de persones van desenvolupar TB i 1,5 milions van morir com a conseqüència de la malaltia.

D'acord amb l'informe de l'Agència de Salut Pública de Barcelona (ASPB), el 2014 es van detectar 300 casos de tuberculosi en residents a Barcelona, el que equival a una taxa d'incidència de 18,6 casos per cada 100.000 habitants. La incidència més alta de TB a Barcelona es troba al barri de Ciutat Vella.

Hem observat que els comportaments de la major part de països s'ajusten a funcions parabòliques, mentre que d'altres funcions, com l'exponencial, presenten ajustos pitjors. Per això, hem desenvolupat un programa per ajustar automàticament una funció parabòlica a les dades epidemiològiques extretes de la base de dades de l'OMS. Els ajustos obtinguts proporcionen una primera visió general de la situació epidemiològica de cada país i permeten afirmar que el comportament observat més freqüent és de tipus decreixent.

Les regularitats observades, en països amb situacions molt diferents, s'han de poder explicar pel comportament de la malaltia. Per fer-ho, hem desenvolupat un model matemàtic amb l'objectiu de descriure la dinàmica observada en l'epidemiologia de la tuberculosi. Encara que simple, el model és robust i permet l'estudi de la dinàmica de la malaltia i fer prediccions a mig i llarg termini de l'evolució de la malaltia en diferents circumstàncies. El model matemàtic es va ajustar adequadament per a diversos països i també per a la ciutat de Barcelona.

El model matemàtic ha permès definir un paràmetre (q) que es pot utilitzar per tal d'avaluar la qualitat dels programes de control de la tuberculosi.

Resumen

La tuberculosis (TB) es una enfermedad infecciosa de transmisión aérea causada por *Mycobacterium tuberculosis*, que ha co-evolucionado con la humanidad. Su estrategia para el éxito consiste en permanecer desapercibido y actuando lentamente. El Informe Global de la Tuberculosis 2015 de la Organización Mundial de la Salud (OMS) estima que, en 2014, 9,6 millones de personas desarrollaron TB y 1,5 millones murieron como consecuencia de la enfermedad.

De acuerdo con el informe de la Agencia de Salud Pública de Barcelona (ASPB), en 2014 se detectaron 300 casos de tuberculosis en residentes en Barcelona, lo que equivale a una tasa de incidencia de 18,6 casos por cada 100.000 habitantes. La incidencia más alta de TB en Barcelona se encuentra en el barrio de Ciutat Vella.

Hemos observado que los comportamientos de la mayoría de países se ajustan a funciones parabólicas, mientras que otras funciones, como la exponencial, presentan peores ajustes. Por este motivo, hemos desarrollado un programa para ajustar automáticamente una función parabólica a los datos epidemiológicos extraídos de la base de datos de la OMS. Los ajustes obtenidos proporcionan una primera visión general de la situación epidemiológica de cada país y permiten afirmar que el comportamiento observado más frecuente es de tipo decreciente.

Las regularidades observadas, en países con situaciones muy diferentes, se han de poder explicar por el comportamiento de la enfermedad. Para ello, hemos desarrollado un modelo matemático con el objetivo de describir la dinámica observada en la epidemiología de la tuberculosis. Aunque simple, el modelo es robusto y permite el estudio de la dinámica de la enfermedad y hacer predicciones a medio y largo plazo de la evolución de la enfermedad en diferentes circunstancias. El modelo matemático se ajustó adecuadamente para varios países y también para la ciudad de Barcelona.

El modelo matemático ha permitido definir un parámetro (q) que puede ser utilizado para evaluar la calidad de los programes de control de la tuberculosis.

Abstract

Tuberculosis (TB) is an airborne infectious disease caused by *Mycobacterium tuberculosis* that has co-evolved with humanity. Its strategy for success consists of remaining almost hidden and acting slowly. The Global Tuberculosis Report 2015 from the World Health Organization (WHO) estimates that, in 2014, 9.6 million people developed TB and 1.5 million died as a consequence of the disease.

According to the report of the Public Health Agency of Barcelona (ASPB), in 2014 were detected 300 cases of tuberculosis in Barcelona residents, equivalent to an incidence rate of 18.6 cases *per* 100,000 inhabitants. The highest TB incidence in Barcelona is found in the Ciutat Vella neighbourhood.

We have observed that the behaviour of most of the countries can be adjusted to parabolic functions, while other functions, such as exponentials, show worst adjustments. For this reason, we have developed a program to automatically adjust a parabolic function to epidemiological data retrieved from the WHO's database. The adjustments obtained provided a first overview of the epidemiological situation of each country and stated that the most frequent behaviour observed was of a decreasing type.

The regularities observed, in countries with very different situations, should be explained by the behaviour of the disease. For this purpose, a mathematical model was developed with the aim to describe the dynamics observed in TB epidemiology. Although simple, the model is robust and allows the study the dynamics of the disease and to make medium and long-term predictions of the evolution of TB disease under different circumstances. The mathematical model was properly adjusted for several countries and also to the city of Barcelona.

The mathematical model has allowed to define a parameter (q) that can be used to evaluate the quality of TB control programs.

Summary

Agraïments

En primer lloc, vull agrair al meu tutor, Daniel López, la oportunitat de poder treballar en aquest projecte. També vull donar les gràcies a en Daniel i a la Clara Prats per tot el que m'han ensenyat al llarg d'aquest procés d'aprenentatge i descobriment. Gràcies pel vostre suport, ajuda i orientació. Gràcies per haver-me descobert el món de la modelització i simulació dels organismes vius, dels models matemàtics... però també gràcies per la confiança que heu dipositat en mi. Gràcies per descobrir-me el màster de bioinformàtica, per ajudar-me en tot el procés… fins i tot a triar el pla d'estudis!

Finalment, m'agradaria donar les gràcies a la meva família i amics pel seu suport.

1. Introduction

Tuberculosis (TB) is an airborne infectious disease caused by *Mycobacterium tuberculosis* that has co-evolved with humanity. Its strategy for success consists of remaining almost hidden and acting slowly. The Global Tuberculosis Report 2015 from the World Health Organization (WHO) estimates that, in 2014, 9.6 million people developed TB and 1.5 million died as a consequence of the disease (WHO, 2016).

In figure 1.1. a map with the estimation of the worldwide TB incidence is shown.

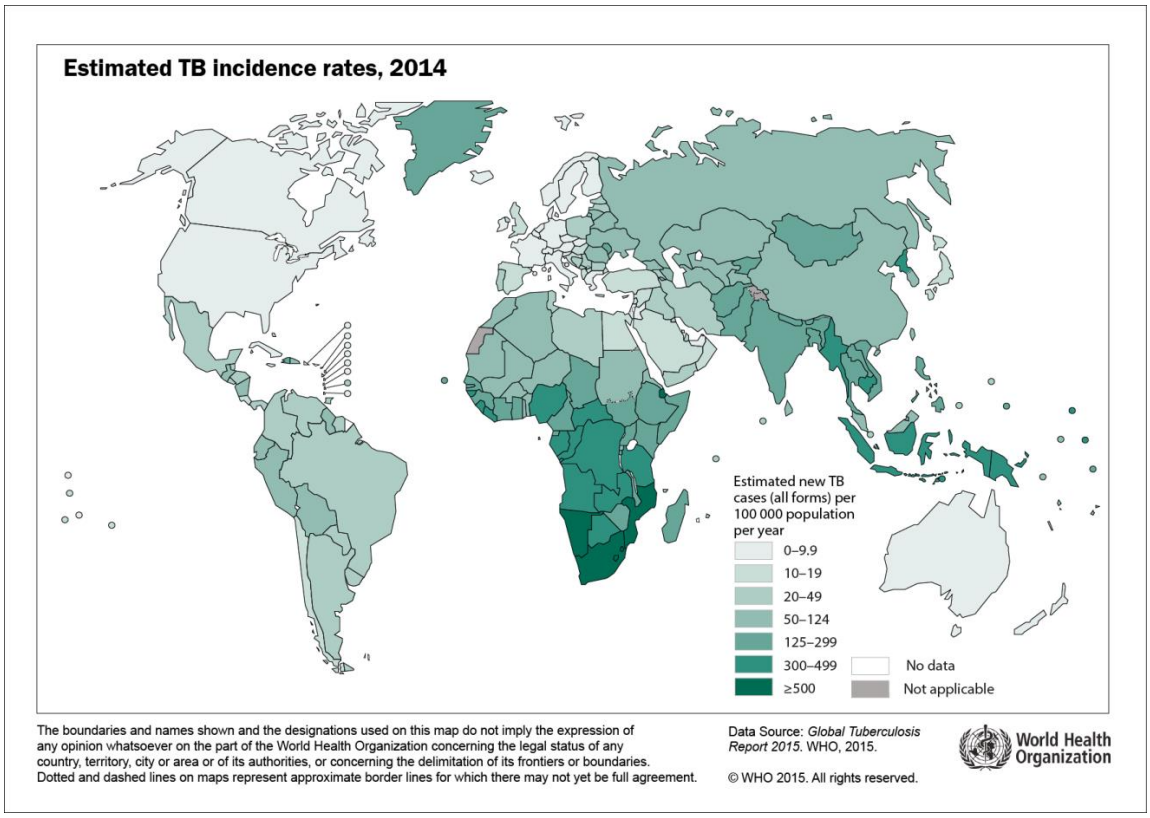


Figure 1.1. Map of the tuberculosis incidence worldwide, data corresponding to the *Global TB Report 2015* (WHO, 2016).

1.1. Tuberculosis main characteristics

TB has its origins 7,000 years ago in Africa (Perrin, 2015). In the XVIII and XIX centuries it produced deadly epidemics in urban centres of Europe, for that reason it is also known as White Plague. It was not until 1882 that Robert Koch, German physician, discovered the germ causing tuberculosis, Koch's bacillus. Tuberculosis is a disease closely linked to poverty. It is transmitted by droplets and affects mostly the lungs. Some of its symptoms include destruction of tissues, cough and fever. Most human infections remain in a latent stage, with approximately 10% probability to develop active disease during their lives. This risk is strongly encouraged when the individual has immunodeficiency disorders, as in the case of AIDS, malnutrition or diabetes, as well as tobacco consumers. People who are infected with HIV^+ are 20 to 30 times more likely to develop active TB.

TB has clearly a higher incidence in men than in women. For most countries, TB notification is twice higher in men than in women (Neyrolles & Quintana-Murci, 2009). In 2014 there were estimated 5.4 million cases among men and 3.2 million among women. TB is among the top 5 causes of death for women aged 15 to 44.

TB is the second leading cause of death from infectious disease after AIDS. It is estimated that approximately one third of the world's population has been infected at some point with an annual incidence of around 9 million new infections. In addition, it is a disease associated with a sociological component of rejection and discrimination, often due to lack of knowledge and ignorance towards the disease. Without proper treatment, the death rate is high. About 70% of people with sputum smear-positive pulmonary TB die within 10 years. Even so, an important remark is that with an appropriate health system, including surveillance and treatment, TB is treatable and preventable.

There are two differentiated stages in TB infection. We can distinguish between latent TB infection (LTBI) and active TB. Individuals with a latent infection are the largest reservoir of the disease (Michel De La Rosa et al., 2007), as the bacillus remains inside the body without showing any symptoms that can be detected. People with LTBI do not have any symptom. At this stage people do not have any disease, they are infected but not diseased. Since the infection remains unnoticed it is said that it is a silent or hidden disease. After a certain time, the infection may become active and the individual often becomes infectious.

Common symptoms of active pulmonary TB are cough with sputum and blood at times, chest pains, weakness, weight loss, fever and night sweats. When the patient coughs, it spreads a spray that contains droplets of between one and five microns in diameter, capable of transporting bacilli through the air into the pulmonary alveoli o the contacts, where the conditions are optimal for their development.

Tb can be diagnosed through microscopic examination of sputum. In most infectious cases, the test is positive. But the ultimate test is the cultivation of a clinical sample, which may be positive in two or four weeks. In all cases the sensitivity to drugs must be studied because drug resistance cases of the disease are difficult to cure.

Recently, rapid tests to diagnose the disease have been marketed. This test allows to diagnose the disease and at the same time, to determine the sensibility to one of the most important drugs to treat the disease, rifampicin. The test *Xpert MTB/RIF* represented a major change for the diagnosis of disease and detection of resistant strains. WHO has promoted its use, especially in countries of high disease burden.

TB is treatable and curable. The treatment of active drug-susceptible TB implies a standard 6-month course of 4 antimicrobial drugs with information, supervision and support to the patient by a health worker in order to guarantee treatment adherence and avoid spreading the disease.

Unfortunately, although the number of tuberculosis cases has been decreasing since the development of the first antibiotic against it, streptomycin, the disease can kill over 50% of its victims if individuals that suffer from active infection do not follow medical treatment. In recent years a certain number of drug-resistant TB strains have appeared. Inappropriate or incorrect use of anti-TB drugs or the use of poor quality medicines, combined with a long-term use of standard anti-TB drugs have caused some disease strains to develop resistance to anti-TB drugs. First-line (or standard) anti-TB drugs have been used for decades allowing the resistance to the medicines to become wide spread. The WHO reports that disease strains that are resistant to a single anti-TB drug have been documented in every country surveyed.

Multidrug-resistant TB (MDR-TB) is a form of TB caused by bacteria that do not respond to, at least, isoniazid and rifampicin, which are the most powerful first-line anti-TB drugs. It is treatable and curable with the use of second-line drugs but the options of this second-line treatment are limited

and recommended medicines may not be always available. It requires an extensive chemotherapy which is costlier and can produce severe adverse drug reactions in patients. Unfortunately, a more severe drug resistance can develop leading to extensively drug-resistant TB (XDR-TB) which is a form of multidrug-resistant TB that responds to even fewer available medicines.

MDR-TB and XDR-TB are becoming an actual problem in the fight against tuberculosis.

1.2. Tuberculosis epidemiology

1.2.1. Basics of epidemiology

Epidemiology (Porta, 2014) is the study of the occurrence and distribution of health-related events, states, and processes in defined populations, including the study of the determinants influencing such processes, and the application of this knowledge to control relevant health problems. It can also be described as the study and analysis of the patterns, causes and effects of health and disease conditions. Epidemiology is the cornerstone of Public Health and shapes policy decisions and evidence-based practice by the identification of risk factors for disease and targets for preventive healthcare. It is not only a branch of medicine treating of epidemics, but it may also study disease in populations of animals and plants.

Studying the epidemiology of a phenomenon implies surveillance, observation, screening, hypothesis testing, analytic research, experiments, and prediction. Distribution refers to analysis by time, place or space, and population.

Epidemiology is concerned with the frequency and pattern of health events in a population. Frequency refers to the number of health events and the relationship of that number to the size of the population, whereas pattern refers to the occurrence of health-related events by time, place and person. Characterizing health events by time, place and person are activities of descriptive epidemiology. Another use is the search for determinants, which are the causes and other factors that influence the occurrence of disease and other health-related events.

In epidemiology there is not just one level of work. Epidemiology can be used to study a phenomenon in very distinct levels, which can range from local levels, such as a single neighbourhood, scaling up to countries, regions or the whole world.

Epidemiologists employ a range of study designs from the observational to experimental that are generally categorized as descriptive, analytic and experimental. Analytic studies aim to further examine known associations or hypothesized relationships. In general, epidemiological studies aim at revealing unbiased relationships between exposure to a determined factor, for instance, a biological agent, to mortality or morbidity.

In order to ease the reader, in this chapter we will briefly describe some of the basic vocabulary and ideas about epidemiology that may be used later on the text.

- Mortality: Measure of the frequency of occurrence of death in a defined population during a specified interval of time.

- Morbidity: The measure of any departure, subjective or objective, from a state of physiological or psychological well-being. According to the WHO Expert Committee on Health Statistics, the morbidity of a period of time can be measured in terms of three units: persons who were ill, the illness that these persons experienced and the duration of these illnesses. Sometimes it is also used as synonymous of the incidence rate.

- Incidence: Measure of the probability of occurrence of a given medical condition in a population within a specified period of time. It is usually expressed as the proportion of the number of new cases of a disease divided by the population at risk in a given period, usually a year. In TB epidemiology it includes the new and relapse cases.

- Prevalence: A measure of the occurrence of a given medical condition. It measures the total number of individuals who have the condition divided by the population at risk in a given period, usually a year. In TB epidemiology it measures the total number of cases of TB at a given point in time.

- WHO Regions: The World Health Organization Member States are grouped into 6 WHO Regions that include African Region, Region of the Americas, South-East Asia Region, European Region, Easter Mediterranean Region and Western Pacific Region. Figure 1.2. shows a world map coloured with all six regions.

- Bacteriologically confirmed case of TB: A patient from whom biological specimen is positive by smear microscopy, culture or WHO-approved rapid diagnostic test.

- New case of TB: A patient who has never been treated for TB or has taken anti-TB drugs for less than one month.

- Retreatment case of TB: A patient who has been treated for one month or more with anti-TB drugs in the past. Retreatment cases are further classified into four categories depending on the outcome of their most recent course of treatment: relapse patients, treatment after failure patients, treatment after loss to follow-up patients and other previously treated patients.

- DOTS: Directly Observed Treatment, Short-Course. It is the name given to the tuberculosis control strategy recommended by the WHO. The five elements of DOTS include political commitment with increased and sustained financing, case detection through quality-assured bacteriology, standardized treatment with supervision and patient support, an effective drug supply and management system and, finally, a monitoring and evaluation system, and impact measurement.

- CDR: Case Detection Rate. It is defined as the proportion of notified cases among the estimated number of new and relapse TB cases, thought to have occurred in a given year.

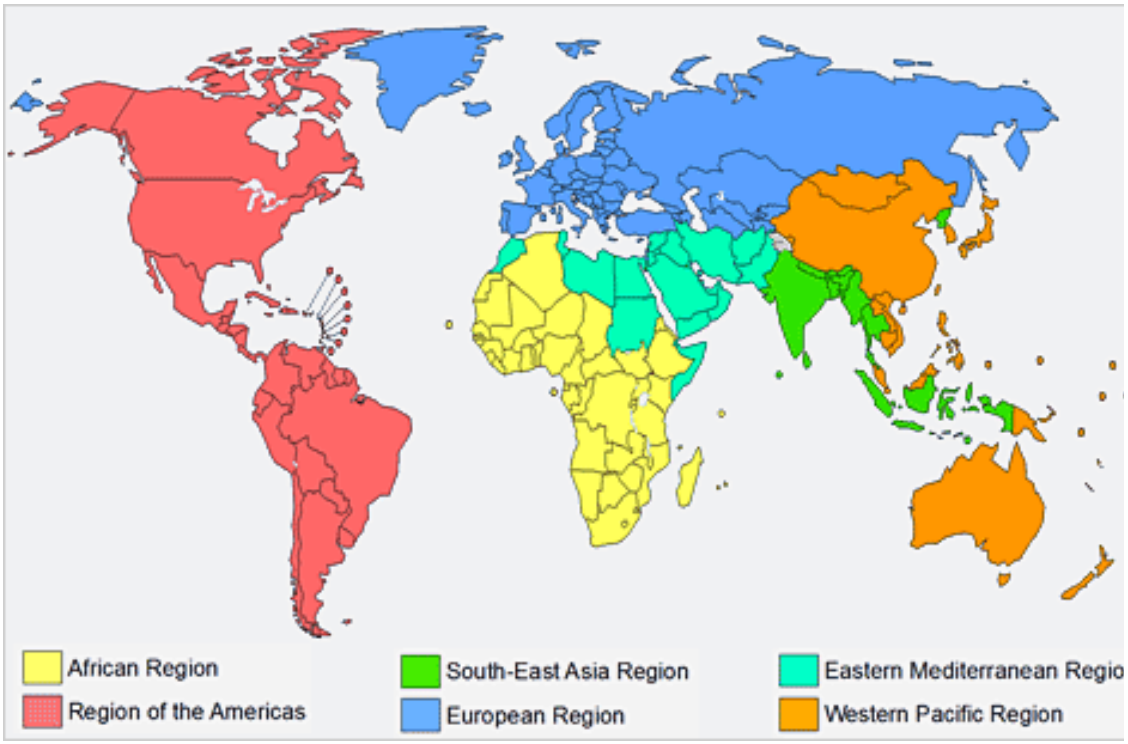


Figure 1.2. A map representing the 6 WHO Regions (WHO, 2016).

- CT: Contact Tracing. A standard procedure in the control of certain contagious diseases whereby diligent efforts are made to locate and treat persons who have had close or intimate contact with a known case. It is also known as case finding. Contact tracing aims to reduce the time required to detect and treat a case and hence reduce the ability of infectious patients to transmit the disease.

- HBCs: High TB burden countries. It is a list of the countries that, according to the WHO epidemiological data, present the highest burden of TB disease. Until 2015 it included Afghanistan, Bangladesh, Brazil, Cambodia, China, DR Congo, Ethiopia, India, Indonesia, Kenya, Mozambique, Myanmar, Nigeria, Pakistan, Philippines, Russian Federation, South Africa, Thailand, Uganda, UR Tanzania, Viet Nam and Zimbabwe. For the 2016-2020 period a new list including 30 countries has been considered. Figure 1.3. shows an image representing the new list of TB HBCs.

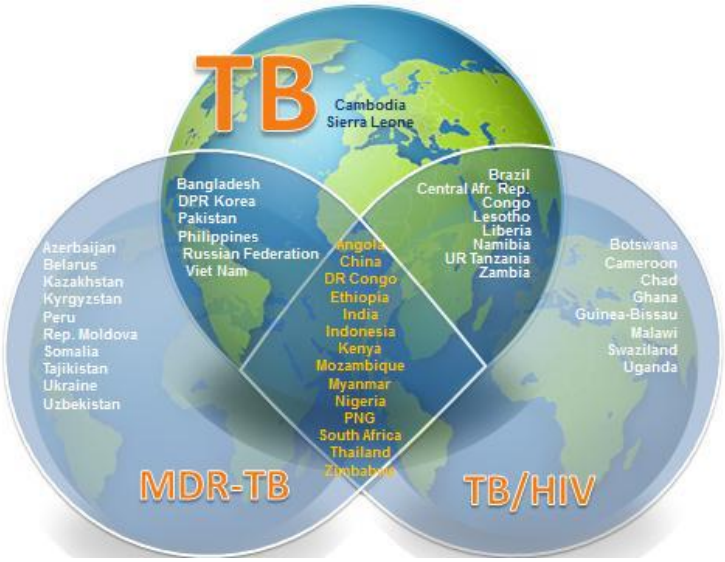


Figure 1.3. The new TB HBCS list for the 2016-2020 (WHO, 2016).

1.2.2. Global tuberculosis epidemiology

Tuberculosis is a major public health concern worldwide: despite a regular, although slow, decline in incidence over the last decade, as many as 9.6 million new cases and 1.5 million deaths were estimated to have occurred in 2014 (WHO, 2016). Globally in 2014 an estimated 480,000 people developed multidrug-resistant TB (MDR-TB) and the global average incidence was 133 cases per 100,000 population. In 2014 there were an estimated 13 million prevalent TB cases and it is estimated that by the end of 2015 the prevalence rate will have fallen 42% globally since 1990. Globally in 2014, approximately 140,000 children died of TB and an estimated 1 million became ill with TB. As tuberculosis is by all means a poverty-related disease it mainly affects the most vulnerable population in the poorest countries. Approximately, over 95% of TB deaths occur in low- and middle-income countries. Figure 1.4. shows the global trends in estimated rates of TB incidence, prevalence and mortality.

The South-East Asia and Western Pacific Regions collectively accounted for 58% of the world's TB cases in 2014 whereas the African Region had 28% of the world's cases. On the other hand, The African Region presented the most severe burden relative to population (281 incident cases per 100,000 population on average). It is important to note that India, Indonesia and China had the largest number of cases (23%, 10% and 10% of the global total, respectively).

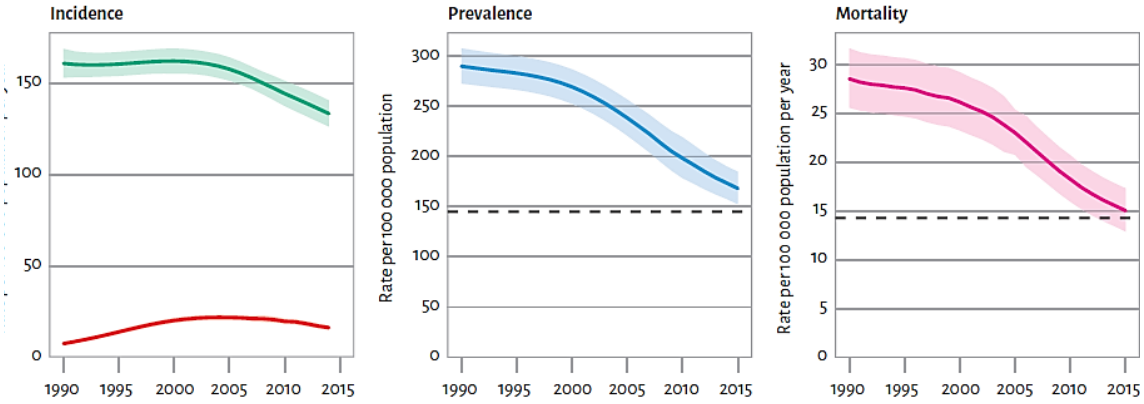


Figure 1.4. Global trends in estimated rates of TB incidence (1990-2014), and prevalence and mortality rates (1990-2015). Left: Estimated incidence rate including HIV-positive TB (green) and estimated incidence rate of HIV-positive TB (red). Centre and right: The horizontal dashed lines represent the Stop TB Partnership targets. Mortality excludes TB deaths among HIV-positive people (WHO, 2016).

Early TB case detection, especially in resource-constrained settings and in marginalized groups remains a challenge. About 3 million people are estimated to remain undiagnosed or not notified and untreated. TB control cannot be carried out without setting up an effective surveillance system with the aim to define the course of the epidemic and also assess the impact of control measures on the disease. This strategy has been used extensively as a control strategy for TB in the developed world (typically with low prevalence) but it is uncommon in developing countries (with high prevalence).

Global strategies against TB

Since 1997, the World Health Organization (WHO) has published a global TB report every year with the aim to provide a comprehensive and up-to-date assessment of the TB epidemic and progress in prevention, diagnosis and treatment of the disease at global, regional and country level. During the period, different strategies against TB have been developed.

For the past decade, the focus has been on progress towards 2015 global targets for reduction in TB disease burden (see table 1.1) set in the context of the Millennium Development Goals (MDGs). The targets are that TB incidence should be falling and that TB prevalence and mortality rates should be halved compared with their 1990 levels. Also, the Stop TB Strategy, developed for the period 2006-2015, has been WHO's recommended approach to achieving these targets.

With the end of 2015 the results of the Stop TB Strategy were analysed. The MDG target of halting and reversing TB incidence by 2015 was achieved globally, in all six WHO regions and in 16 high TB burden countries. The global mortality rate in 2015 was 47% lower than in 1990 and so the target of a 50% reduction was almost met. Globally, the TB prevalence rate in 2015 was 42%slower than in 1990 and the target of a 50% reduction was met in three WHO regions and in nine high burden countries (HBCs). All Stop TB targets were met in the Region of the Americas, the South-East Asia Region and the Western Pacific Region and in nine HBCs: Brazil, Cambodia, China, Ethiopia, India, Myanmar, the Philippines, Uganda and Viet Nam.

The WHO has launched a new global TB strategy for the "post-2015 era" which is aimed at ending the global TB epidemic by 2035 with a 95% reduction in number of TB deaths and a 90% reduction in TB incidence rate, both compared to 2015. The End TB Strategy is based on the three pillars that emphasize patient-centred TB care and prevention, bold policies and supportive systems, and intensified research and innovation. In order to ensure full impact, the fight against TB must build on principles of government stewardship, engagement of civil society, human rights and equity, and adaptation to the unique context of diverse epidemics and settings. Figure 1.5. shows the predicted behaviour of global TB incidence with optimization of current tools combined with progress towards universal health coverage and social protection from 2015 and the additional impact of new tools by 2025.

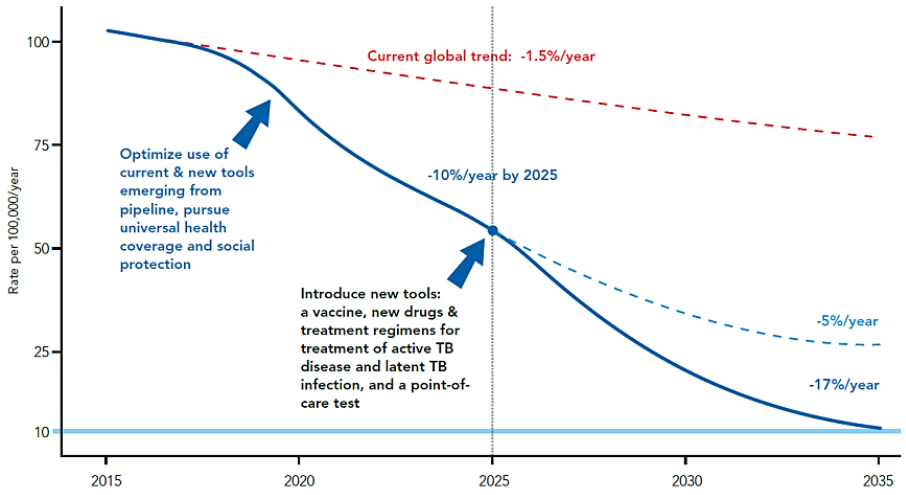


Figure 1.5. Projected acceleration of the decline in global TB incidence rates according to the End TB Strategy (WHO, 2016). Retrieved from the End TB fact sheet.

Surveillance methods and difficulties

Surveillance is by all means a key tool in the fight against TB. Between 2000 and 2014 an estimated 43 million lives were saved through TB diagnosis and treatment. But getting to know the epidemiological situation of tuberculosis is not an easy task. Most cases often go unnoticed or are detected with some delay, when the sick individual has already infected individuals. It is considered that a single untreated individual can infect approximately 2-10 people before it is detected. For this reason, a strong surveillance and disease control system is required.

- **TB incidence**

Measuring TB incidence at national level has never been accomplished because it would require long-term studies among large cohorts of people, involving high costs and challenging logistics.

The ultimate goal of TB incidence surveillance is to directly measure TB incidence from TB notifications in all countries. This requires strengthened surveillance, better quantification of under-reporting and universal access to health care.

In countries that have both high-performance surveillance systems and where few cases are not diagnosed due to the quality of and access to health care, notifications of TB cases provide a good proxy indication of TB incidence. There are a large number of countries where these criteria are not yet met. In these cases, better estimates can be obtained from an inventory study.

On each country social and political circumstances are quite different and so is the surveillance of TB incidence. Methods currently used by WHO in order to estimate TB incidence can be grouped into four major categories. These are:

1. Case notification data combined with expert opinion about case detection gaps.
2. Results from national TB prevalence surveys.
3. Notifications in high-income countries adjusted by a standard factor to account for under-reporting and under-diagnosis.
4. Results from inventory/capture-recapture studies.

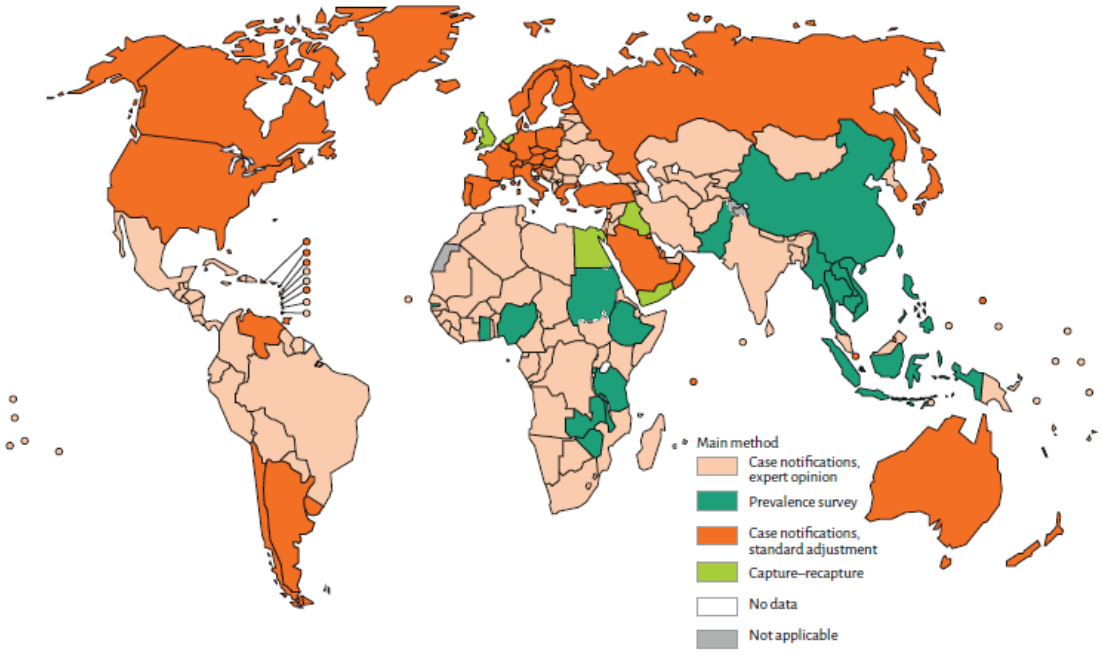
Estimates of TB incidence are often uncertain in many countries, especially on those with a large private sector or political instability where cases may be detected but not reported. With the aim

to stablish a basis for addressing gaps in reporting, an inventory study can be used in order to quantify the number of cases that are detected but not reported to national surveillance systems.

Figure 1.6. contains a world map where the main method used to estimate TB incidence is shown.

- **TB prevalence**

The prevalence of bacteriologically-confirmed pulmonary TB can be directly measured in nationwide population-surveys using sample sizes of around 50,000 people in countries with a relatively high burden of TB (around 100 cases per 100,000 population or above). Usually, the cost of a survey ranges from US$ 1 to 4 million. In order to assess trends in disease burdens, repeat surveys are conducted about every ten years. The WHO has an online technical appendix provides details about the methods used to produce estimates of TB prevalence.



[a] In the first method, case notification data are combined with expert opinion about case detection gaps (under-reporting and under-diagnosis), and trends are estimated using either mortality data, repeat surveys of the annual risk of infection or exponential interpolation using estimates of case detection gaps for three years. For all high-income countries except the Netherlands and the United Kingdom, notifications are adjusted by a standard amount or measure of under-reporting from inventory studies, to account for case detection gaps. For further details about all four methods, see text.

Figure 1.6. World map representing the main method used to estimate TB incidence (WHO, 2016).

In low and medium-burden countries, sample sizes and costs for surveys become prohibitively large. For this reason, alternative methods must be implemented in order to estimate TB prevalence. When survey data is not available, prevalence can be indirectly estimated as the

product of incidence and the average duration of the disease. This estimation, however, presents a considerable uncertainty.

1.2.3. Tuberculosis in Barcelona

According to the report of the Public Health Agency of Barcelona (ASPB), in 2014 were detected 300 cases of tuberculosis in Barcelona residents, equivalent to an incidence rate of 18.6 per 100,000 population (Orcau i Palau, Arcas i Ferré, Caylà i Buqueras, & García de Olalla i Rizo, 2015). It is considered a 9% decline in the number of cases regarding 2013, a year in which were detected 329 case. The disease had a higher incidence (Orcau i Palau et al., 2015) in men, where it reached, in 2014, 24.5 cases per 100,000 population, while in women it was 13.3 cases per 100,000 population. About a 50.3% of the detected cases were immigrants, which represents a decrease of 13% over the previous year. As for the local population, however, there has been an increase of 9% in the number of cases.

Given the country of birth, very different patterns can be observed in the age distribution. In the native population the highest rate occurred in people aged over 64 years while the immigrant population presented a major influence on people between 25 and 39 years.

Highlight the district of Ciutat Vella, which remains the spot with the highest incidence in the city with 60.5 cases per 100,000 population.

Regarding risk factors, 85 individuals (28.3%) were smokers, 25 (8.3%) consumed alcohol in excess, 19 (6.3%) suffered from diabetes, 17 (5.6%) were infected with HIV^{+}, 13 (4.3%) had received prior immunosuppressant therapy and 6 (2%) were drug addicts.

In 2014, Barcelona held the 45th Union World Conference on Lung Health that led to the Barcelona Declaration, which aims for the recognition of the importance of TB by representatives of several countries, and the right of individuals to access public health and the treatment of this disease. Likewise, it also establishes the commitment official organizations and non-governmental organizations work together, through parliaments, in order to achieve a better response against the disease.

As shown in figure 1.7., the evolution of TB in Barcelona from 1990 to 2014 presented a linear decrease of approximately a 10% yearly decrease during a short time. However, on the later years,

the behaviour of TB incidence in Barcelona became less linear, with a worst decreasing tendency. This reduction in the slope decrease may be due to factors that are difficult to control (arrival of immigrants from countries with high incidence ...), but it is also plausible that the observed decline is not really linear, but obeys a different dynamic.

On a first hypothesis, the system's structure was suspected to determine the dynamics on the decrease of the incidence, even when a proper TB control program was being conducted.

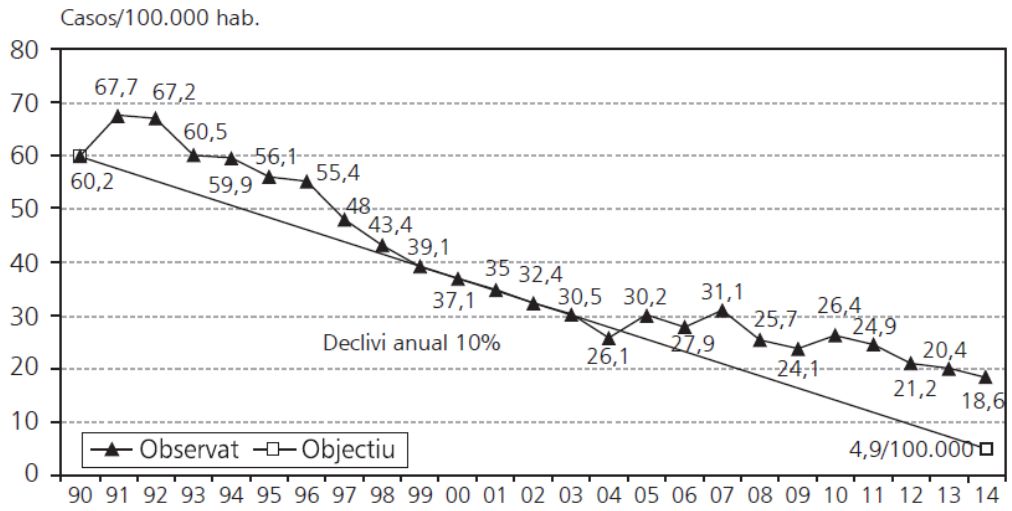


Figure 1.7. Evolution of TB incidence in Barcelona during the period 1991-2014. (Orcau i Palau et al., 2015).

Figure 1.8. shows the adjustment of the parabolic equation to the data of Barcelona retrieved from figure 1.7. After making different types of adjustments of experimental data we found that from 1991 to 2014, epidemiological data fit quite properly to a parabolic function. This suggests that the behavior is typical of the structure of the system. That is, we can make the assumption that the observed behavior is not caused by external factors but, instead, it is intrinsic to the system. We pretend to develop a model that allows us to understand this behavior.

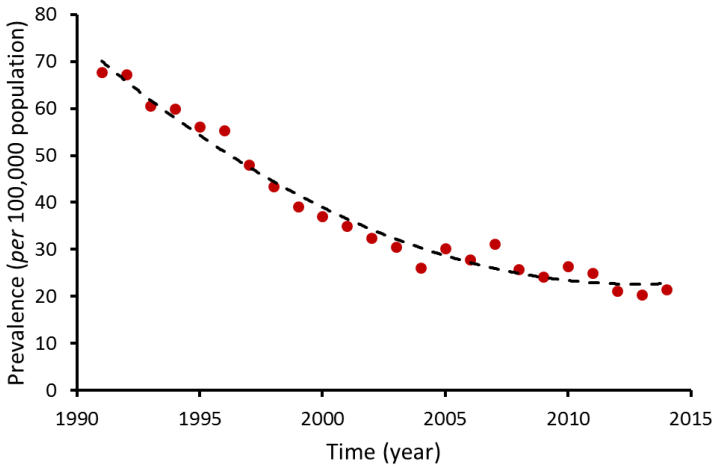


Figure 1.8. Adjustment of the parabolic function to data of TB incidence in Barcelona during the period 1991-2014. (Orcau i Palau et al., 2015).

In many cases, the behaviour of TB epidemiology can be adjusted to a parabola. On the first part of this work, TB epidemiological data will be adjusted to the quadratic function in order to assess the different types of behaviour that can be expected.

1.3. Mathematical models in epidemiology

Mathematical models are a type of scientific models that allow us to describe a system using mathematical concepts and mathematical language. In epidemiology, these models have become a very important tool in the study of the dynamics of disease. Mathematical models in epidemiology can be used for both improving understanding and predictive purposes.

Back in 1760, Daniel Bernoulli published the first known article which defined an explicit mathematical model to describe an infectious disease. Since then, with the participation of epidemiologists and mathematicians such as Ronald Ross or Kermack and McKendrick, mathematical models have evolved to bringing a wide range of models, tailored to specific diseases or other more general models. These models can be as simple as a differential equation or so complex that they require a large support of computer technology. However, the aim of mathematical models is to represent reality using the most simplified language and mathematical tools available. Mathematical models were first used to study the dynamics of TB epidemiology in the 1960s (Waaler, Geser, & Andersen, 1962) and have been used extensively since the mid-1990s.

Continuous models are especially useful for studying the spreading diseases that present a behaviour similar the ecological r-strategy, i.e. with a rapid increase in their incidence. This strategy usually guarantees that the fraction of population simultaneously affected in a small period is large enough to be statistically significant and to support the continuum hypothesis. In contrast, the diseases have a behaviour similar to the ecological K-strategy present slow dynamics and small incidence but greater persistency in a community, usually remaining hidden for years. Tuberculosis may be considered a disease that follows a behaviour similar to the ecological K-strategy (Prats, Montañola-Sales, et al., 2016). The number of people with a simultaneous active disease in a certain community is relative small.

In order to develop mathematical models in epidemiology multiple schemes have been used, from the basic models of differential equations with compartments (SEIR, SIR, SIS, SI...) to stochastic models that may focus on an individual scale. Some variations of models have allowed obtaining more realistic results by taking into account population heterogeneity. An example of this heterogeneity could be the variation within gender or age group.

From an epidemiological point of view, two main types of models are used in TB epidemiology (Aparicio & Castillo-Chavez, 2009). On one hand we can find aggregated models, that include epidemiological states models, cluster models and standard compartmental models; and on the other, age structured models.

Compartment models consist of different pools of individuals characterized by the status of the disease. In the most common models, for instance SEIR, the S stands for susceptible, or persons capable of being infected; E stands for exposed, persons who have been in contact with the disease but are not infectious; I stands for infectious, individuals who might transmit the disease and R for recovered, individuals who have overcome the disease. These are the common foundations of this type of models, but depending on the disease to model the behaviour of these compartments can change or even may differ.

Note the relevance of the SEIR model, in which a person who has acquired the parasite does not develop the disease, automatically but must go through a phase of latency. This model incorporates a compartment exposed (E), infected (I) and recovered (R). The exposed individuals are defined as the fraction of people infected that do not have symptoms of the disease and depending on the model you are adjusting, are either able or unable to transmit the disease.

Individual-based models (IbM) are also used in TB epidemiology since this kind of model can be used to capture emergent phenomena that cannot be studied nor understood with a top-down approach. The output of IbM result interesting since it allows to study the variations in global dynamics emerging from actions over a certain community. Recently, an IbM has been developed with the aim of analysing the evolution of pulmonary TB in a community (Prats, Montañola-Sales, et al., 2016). The model was developed with a bottom-up approach for studying the dynamics of pulmonary TB in a certain population, that was considered constant. The model allowed to study the short-term effect of health-control polices on modifying the structure of the TB infected subpopulation. Results showed that the characteristics of the population are crucial for the local epidemiology of TB.

In the last decade, the use of mathematical modelling in the TB research has increased significantly (Zwerling *et al.*, 2015). Different models have been developed in order to describe and advance in the understanding of some aspects of the dynamics of tuberculosis infection in different scales or levels. For example, spatiotemporal and socio-demographic dynamics aspects of transmission (Kasaie *et al.*, 2013), the dynamics of growth of granulomas and its relation to the evolution of infection (Gong *et al.*, 2015; P.-J. Cardona & Prats, 2016), the effect of antibiotics and vaccines (Linderman *et al.*, 2015), or Cardona & Prats (2016) who describe the behaviour of active disease in an animal model.

Mathematical models are a key element in TB epidemiology since they provide valuable insight into potential impact and cost-effectiveness of strategies to improve both TB diagnosis and treatment. Models of TB not only serve to describe the epidemiology of the disease but they also provide a better understanding of the key drivers of impact and allow us to describe the role of population structure on the dynamics of an epidemic (Begun, Newall, Marks, & Wood, 2013).

1.4. Objectives

The fact that in most countries where there is a decrease in the incidence obeys the same kind of dynamic reinforces the idea that the behaviour of the TB epidemiological dynamic is structural. Therefore, it is reasonable to define a mathematical model that correctly describes the behaviour of this type of system.

Noting this regularity in real systems, this work aims to:

(1) Analyse tuberculosis epidemiology from real epidemiological data. Determining the main dynamics observed in the processes which progressively reduce the incidence.
(2) Develop the simplest possible (*Principle of parsimony*) compartment-based mathematical model that allows the study of the evolution of TB disease, properly describing the types of behaviour observed. It is expected that the model can explain the behaviours that are not strictly a parabolic decline.
(3) Make possible the use of model for predictive purposes in the medium and long term.
(4) Design a system for evaluating the quality of TB control programs

1.5. Context of this work

Biological Systems Engineering and Public Health

Biological Systems Engineering (BSE) can be defined as the discipline that applies concepts of biology, chemistry and physics, along with engineering and design principles, to solve problems in biological systems.

Thus, BSE requires the knowledge of the basics physical, chemical and biological processes that constitute the skeleton and metabolism of biological systems, but it also requires other disciplines such as mathematics, computer science and engineering, which allow the modelling of their behaviour, or the design of structures to produce some kind of benefit. Among the applications of this discipline we can highlight the design of facilities for biological systems, genomics and proteomics analysis, and also the modelling and simulation of biological systems.

This last discipline, computer modelling, allows biological systems engineering to provide an essential tool for public health. With multidisciplinary knowledge, mathematical models can be established in order to computerize and try to analyse, understand, and even, in some cases, to predict the behaviour of different diseases and the effect of drugs in the body.

Scientific context

This work has been developed under the supervision of the *Computational Biology and Complex Systems* (*BIOCOMSC*) research group (*http://biocomsc.upc.edu/en*), *Agència de Salut Pública de Barcelona* (*Barcelona Public Health Agency, ASPB, http://www.aspb.cat/*) and the *Unitat de Tuberculosi Experimental* (*Experimental Tuberculosis Unit, UTE, https://unitatdetuberculosiexperimental.wordpress.com*), the later belonging to the *Institute for Health Science Research Germans Trias i Pujol (IGTP)*.

As shown in figure 1.9., research on TB can be conducted in very different scales that can go from epidemiological (global, country or city) to molecular and genomic level. The dynamics of the disease can also be studied at an individual level or at the involved organs and tissues level. At a cellular level, the dynamics of macrophages and bacillus inside the alveolus can be studied under different conditions. The different scales in TB research are often individually and independently addressed. However, an integrating view of the different levels of study can procure a better understanding of the different mechanisms of the disease and greatly contribute to the scientific knowledge.

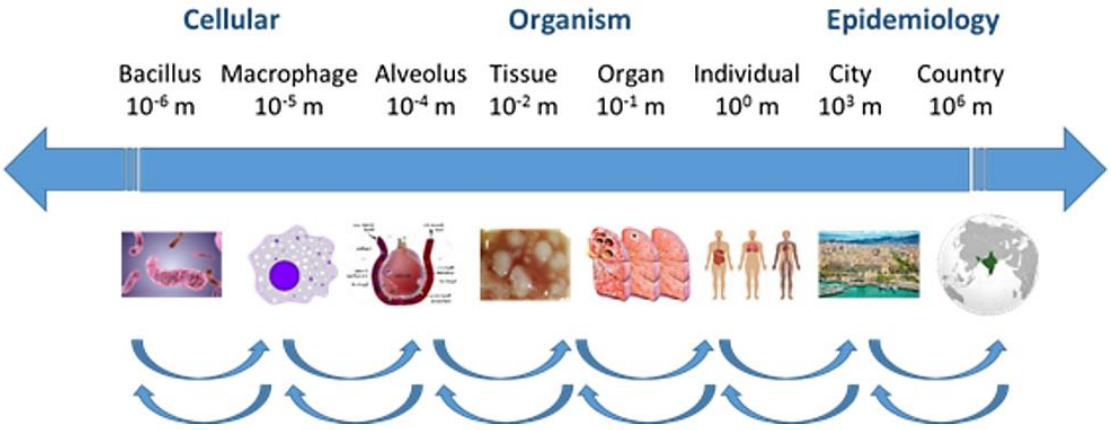


Figure 1.9. Different levels in the study of TB. The lower arrows represent the relationship between the different levels.

1.6. Structure of this work

In this work, two main parts can be distinguished: the analysis of tuberculosis epidemiology from real epidemiological data and the development and implementation of the mathematical model.

This first chapter has served as a brief, but sufficient, introduction to the main characteristics in TB epidemiology and to settle the objectives of this work. The second chapter will talk about the analysis of the tuberculosis epidemiology that was conducted with the TB surveillance data retrieved from the WHO's data base. On the other hand, the third chapter will be dedicated to the description, adjustment and analysis of the compartment-based mathematical model here proposed. The fourth chapter will summarize the results and discussions of this work whereas the fifth chapter will summarize the conclusions of the work.

2. Analysis of the tuberculosis epidemiology dynamics

2.1. Surveillance data

TB surveillance data were obtained from the WHO's database, which provides a vast information relative to the tuberculous infection such as statistical data and prevalence and incidence world maps. Data is collected in annual rounds of global TB data collection from countries and territories, including 194 WHO's Member States. For that purpose, a web-based system is used (Extranet.who.int, 2016). The system is usually opened by mid-March and remains available for a few months to allow the countries to make modifications on the data submitted. In 2015, 205 countries and territories that accounted for more than 99% of the world's population and estimated TB cases reported data. The data collected in 2015 included a wide range of topics such as TB case notifications and treatment outcomes, including breakdowns by TB case type, age, sex and HIV status, or laboratory diagnostic services.

Data submission is different on each country, even at local level of a country. For instance, countries in the European Union submit notification and data to the TESSy system managed by the European Centre for Disease Prevention and Control (ECDC) and then data is uploaded from the TESSy system to the WHO's database. A clear example could be considering on one hand, Spain, which notified 4,818 new cases and had an estimated case detection rate of 84% while, on the other hand, Nigeria notified 86,464 new cases and had an estimated detection rate of 15% (WHO, 2016).

Once the system is closed, all data is processed and updated in the public WHO's database.

The data downloaded from the WHO's database consists on CSV file with data structured on 47 columns of information, containing both numeric values and strings. Figure 2.1. shows a small fragment of the data contained in the file downloaded from the WHO's database, while figure 2.2. shows an example of the information that can prove more relevant in order to adjust the compartment-based model.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	country	iso2	iso3	iso_numeric	g_whoregion	year	e_pop_num	e_prev_100k	e_prev_100k_lo	e_prev_100k	e_prev_num	e_prev_num_lo	e_prev_num_hi
2	Afghanistan	AF	AFG	4	EMR	1990	12067570	307	160	502	37000	19000	61000
3	Afghanistan	AF	AFG	4	EMR	1991	12789374	344	180	562	44000	23000	72000
4	Afghanistan	AF	AFG	4	EMR	1992	13745630	373	191	613	51000	26000	84000
5	Afghanistan	AF	AFG	4	EMR	1993	14824371	392	194	657	58000	29000	97000
6	Afghanistan	AF	AFG	4	EMR	1994	15869967	409	198	694	65000	31000	110000
7	Afghanistan	AF	AFG	4	EMR	1995	16772522	425	201	731	71000	34000	120000
8	Afghanistan	AF	AFG	4	EMR	1996	17481800	436	202	758	76000	35000	130000
9	Afghanistan	AF	AFG	4	EMR	1997	18034130	444	204	776	80000	37000	140000
10	Afghanistan	AF	AFG	4	EMR	1998	18511480	446	204	781	83000	38000	140000
11	Afghanistan	AF	AFG	4	EMR	1999	19038420	443	203	774	84000	39000	150000
12	Afghanistan	AF	AFG	4	EMR	2000	19701940	431	203	743	85000	40000	150000
13	Afghanistan	AF	AFG	4	EMR	2001	20531160	410	198	697	84000	41000	140000
14	Afghanistan	AF	AFG	4	EMR	2002	21487079	393	196	658	84000	42000	140000
15	Afghanistan	AF	AFG	4	EMR	2003	22507368	376	191	622	85000	43000	140000
16	Afghanistan	AF	AFG	4	EMR	2004	23499850	359	186	586	84000	44000	140000
17	Afghanistan	AF	AFG	4	EMR	2005	24399948	346	181	563	84000	44000	140000
18	Afghanistan	AF	AFG	4	EMR	2006	25183615	338	177	548	85000	45000	140000
19	Afghanistan	AF	AFG	4	EMR	2007	25877544	332	175	538	86000	45000	140000
20	Afghanistan	AF	AFG	4	EMR	2008	26528741	328	173	531	87000	46000	140000
21	Afghanistan	AF	AFG	4	EMR	2009	27207291	326	172	528	89000	47000	140000

Figure 2.1. Caption of a small part of the data obtainable through the WHO's database.

	A	B	C	D	E	F	G
1	country	iso3	g_whoregion	year	e_pop_num	e_prev_100k	e_inc_100k
2	Afghanistan	AFG	EMR	1990	12067570	307	189
3	Afghanistan	AFG	EMR	1991	12789374	344	191
4	Afghanistan	AFG	EMR	1992	13745630	373	191
5	Afghanistan	AFG	EMR	1993	14824371	392	189
6	Afghanistan	AFG	EMR	1994	15869967	409	188
7	Afghanistan	AFG	EMR	1995	16772522	425	188
8	Afghanistan	AFG	EMR	1996	17481800	436	188
9	Afghanistan	AFG	EMR	1997	18034130	444	189
10	Afghanistan	AFG	EMR	1998	18511480	446	189
11	Afghanistan	AFG	EMR	1999	19038420	443	190
12	Afghanistan	AFG	EMR	2000	19701940	431	190

Figure 2.2. Example of the information that can prove relevant for our epidemiological data analysis.

2.2. Treatment of the epidemiologic data for its analysis

The mathematical model aims to predict the behaviour of tuberculosis epidemiology in different populations. Before developing the mathematical model, a previous analysis of epidemiological data is required.

As seen in the previous chapter, a parabolic equation can be adjusted to most of the different types of behaviour observed in epidemiological data (for instance, to the data of TB incidence in Barcelona). For this reason, the main objective of the treatment of the epidemiological data is to adjust a second degree polynomic equation (2.1) to the behaviour of the prevalence and incidence of the different countries with the data provided by the WHO, as described on the previous section.

$$y = Ax^2 + Bx + C \tag{2.1}$$

An initial analysis showed that it is possible to achieve a good fit with a quadratic equation. The goodness of the parabolic adjustment is not achieved with an exponential function. The parabolic decrease properly describes the decline of incidence in most of the world's countries, with a fairly regular behaviour. As shown in figure 2.3., the parabolic decreasing behaviour provides a good adjustment for countries with both a high and low TB incidence.

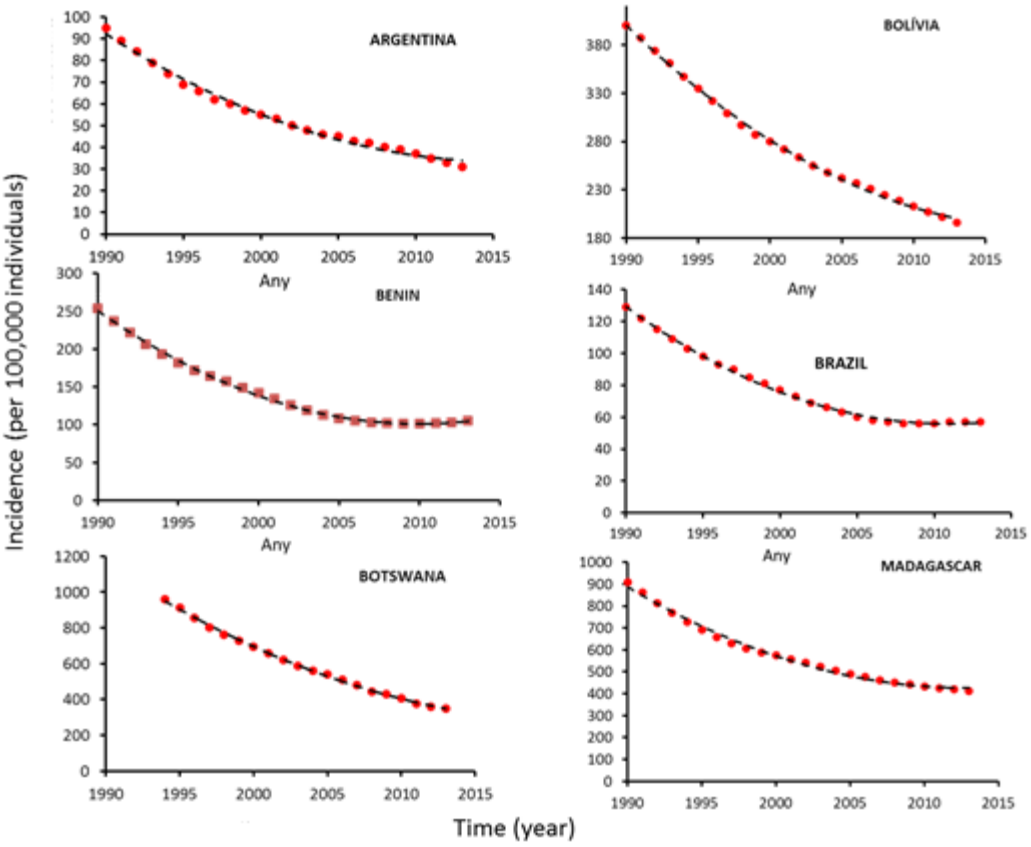


Figure 2.3. Adjustments of the parabolic function to TB incidence of several countries.

The graph of the quadratic function represents a parabola which location, size and how it opens depend on the values of *A*, *B* and *C*. The value of *a* can be used to distinguish two types of openings. If $A > 0$, the parabola has a minimum point and opens upward. On the other hand, if $A < 0$, the parabola has a maximum point and opens downward. Whether minimum or maximum, the extreme point of the parabola corresponds to its vertex and its x-coordinate can be located at $x = \frac{-B}{2A}$. These concepts will later be used on the description of the mathematical model. Figure 2.4.

shows the four types of behaviour that can be expected from the adjustment of the epidemiological data.

At first, data was analysed in an Excel spreadsheet but it required more time and resources. With the aim to decrease the time required to adjust and treat the data, a script to automate the analysis of the data was developed in Matlab. The use of a Matlab script to analyse data allows a quick and objective analysis that is often necessary because the WHO data are updated periodically. Thus, although the creation of the script requires a period of preparation and verification of the results with a rudimentary analysis, once implemented it allows to update the data with minimal effort.

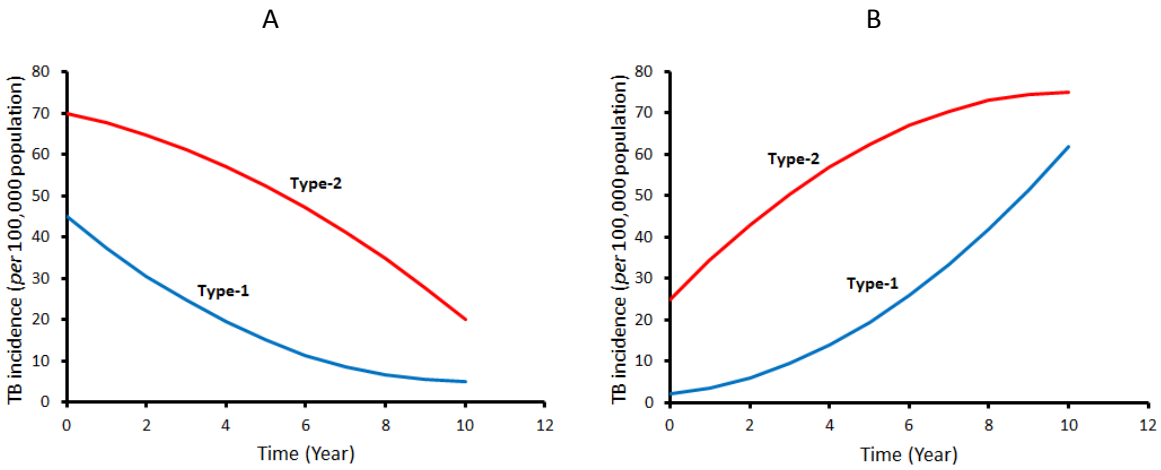


Figure 2.4. The four types of behaviour expected in WHO's data.
A: Decreasing behaviours. B: Increasing behaviours.

In this section, the main considerations on the development of the script as well as the flow diagrams used for the purpose will be presented. Figure 2.5. shows the work-flow followed for the treatment of epidemiological data.

The script created imports the data from the Excel sheet, screens the information and repeats different adjustments starting from a minimum number of points (10 years) to the maximum number of points available for each country. The best adjustment, considering as the best the one with at least the minimal length and the highest r-squared value, is then selected and the process is repeated for each country. Right after the first adjustment is completed for a given country, a second adjustment is conducted and stored. On every successful adjustment, only if all conditions are met, the figure representing the adjustment of data is automatically stored. After all adjustments are finished, data is filtered to reduce size and then exported to be analysed later on.

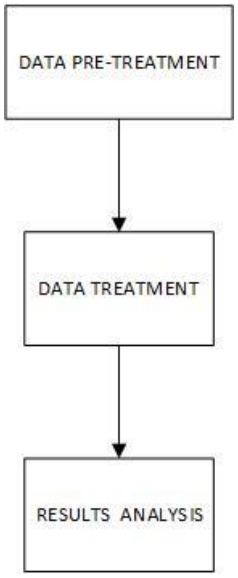


Figure 2.5. Main scheme of the work-flow to analyse epidemiological data.

Data pre-treatment

Before starting to work on the script to treat the data retrieved from the database, a pre-treatment is required in order to prepare it for an optimal operation. The first step will consist in converting the CSV file into an Excel (*xlsx*) file with the help of the software's incorporated function. It is important to note that the data from the CSV file has comma separated valued and it has to be converted into different rows, one for each variable. Once the data is properly converted into the *xlsx* file, the replacement assistant must be used to replace the dots (.) in the numerical values for the decimal operator (,). Another precaution is to change the names of certain countries in the data file because the software does not support certain characters. For example, Côte d'Ivoire must be replaced as Cote d'Ivoire, without any special characters. If some of this considerations are not met, Matlab can crash due to incompatibilities on the structure of the data.

Data treatment

Data treatment is divided in many different steps with a varying degree of complexity. Figure 2.7. shows the main flow diagram of the script developed for data treatment, that is executed through the Matlab interface once the data pre-treatment is accomplished. The script has been designed to adjust the quadratic equation to prevalence or incidence data on each execution, depending on the

value of a switch variable, but not both at the same time. The reason for such condition is the long execution time and that, depending on the circumstances, it is only require to work or re-analyse

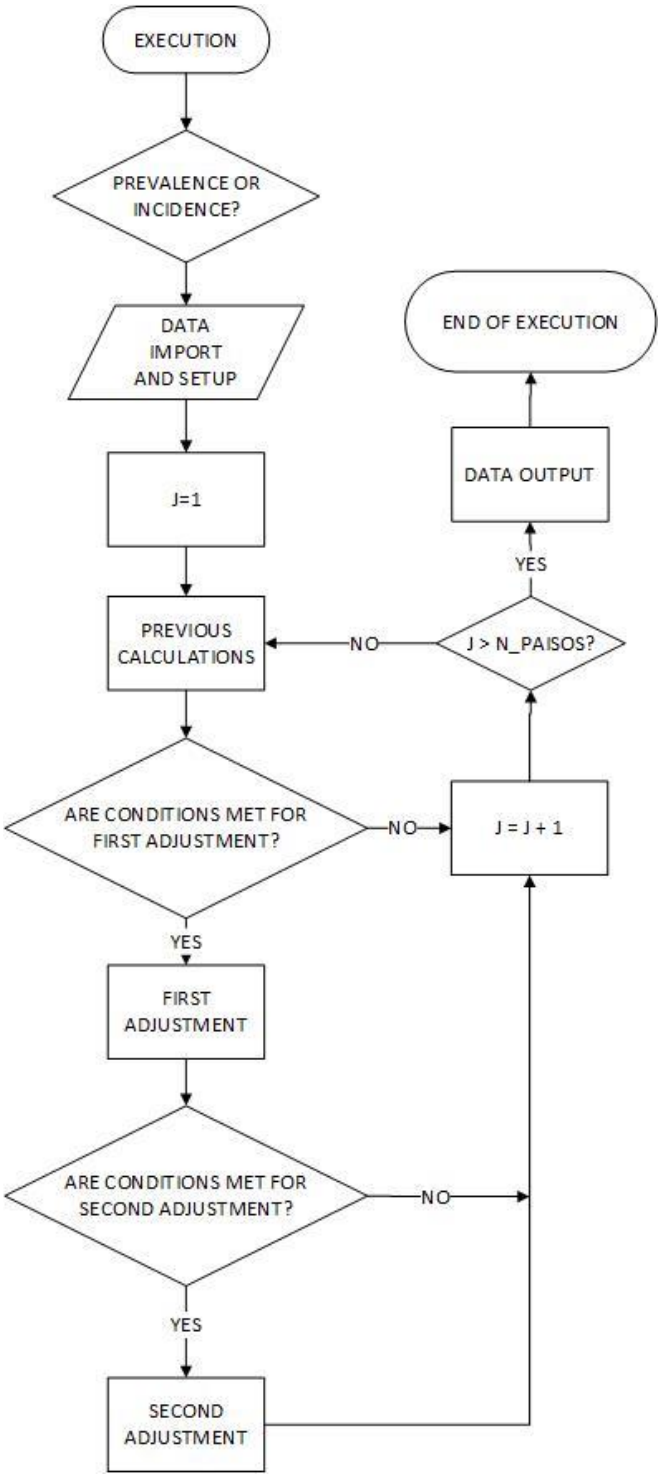


Figure 2.7. Main diagram of the script to adjust epidemiological data.

data with for one parameter (incidence or prevalence). Before executing the script, the user must change the value of the variable *prev_inc* that will serve as a switch, this corresponds to the *"PREVALENCE OR INCIDENCE?"* box in figure 2.7. If the value of the variable is 0, the program will adjust prevalence data. For any other value, the program will adjust incidence data. It is important to note that the current section aims to cover only the main characteristics of the script. In order to have a more complete knowledge of the script, the reader can check the script with annotations in the annex of this work.

Data import and setup

The Matlab script developed imports the data retrieved from the WHO's database once it has been pre-treated, with the use of the *xlsread* function (*Matlab,* 2016) with the specification of the location of the file, and assigns different vectors, one for each variable. The read data is divided into two different matrices, *num* that contains the numerical values of data and *txt* that contains the characters or strings. Once the data is read it is assigned into different vectors as shown on table 2.1. The naming of this vectors is similar as the variable names used by the WHO.

Table 2.1. Vectors defined to store the imported data.

Vector name	Stored data	Vector name	Stored data
country	Country name	*e_pop_num*	Population (individuals)
iso3	ISO3 country code	*e_prev_100k*	TB prevalence (per 100k population)
g_whoregion	WHO Region	*e_inc_100k*	TB incidence (per 100k population)
year	Year		

On the script, both matrices and vectors are created. Sometimes, working with vectors is more practical than working with matrices when we want to define specific names, make smaller groups of data and work more independently, but we have to be careful and be aware at all times of the created vectors in order to avoid overlaps that could cause a bad script execution.

A list containing the non-repeated values for the vector containing the countries column is created using the order *unique*. This information is contained in the vector *country_list*. It is important to assign the *stable* property in order to avoid the restructuring of the data into an alphabetical order, what would cause a mismatch with the other vectors. The data retrieved from the WHO is already in alphabetical order base on the country name. The order used results *country_list = unique(country, 'stable')*. Other lists containing the non-repeated ISO3 codes and the WHO Regions of data are created using the same function, and will later be used to export data.

The number of countries is found and stored in a variable using the command *n_paisos = length(country_list)*. Figure 2.8. shows an example of how the *unique* function works.

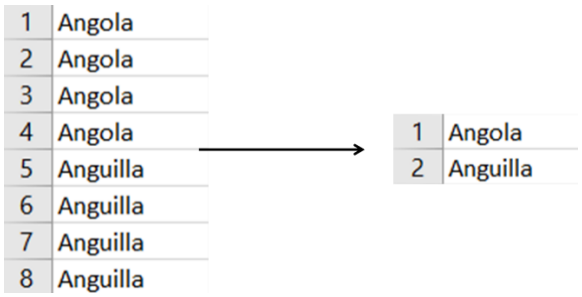


Figure 2.8. Example of the utilization of the Matlab function *unique*.

Another important step consists in creating a vector, named *country_data*, containing the number of times that each country is repeated in the data entry. This vector will have a length corresponding to *n_paisos* and will serve to find the data limits for each country when screening and adjusting data. It is created with the help of the following two commands:

```
[idx, label] = grp2idx(iso3);
country_data = hist(idx, unique(idx, 'stable'));
```

The maximum value of *country_data* is stored as the variable *lgth_ajustos*. Both variables, n_paisos and lgth_ajustos are then used to set the dimensions of the matrices that will contain data from the adjustment. Those are the matrices which name starts with *ajustos*, that will store the value of incidence or prevalence for each adjustment, and *mat_derivada*, that will contain the values of the derivative on each point for every adjustment. A set of matrices with name starting by *param_ajustos* will also be created, with a dimension *n_paisos x 6.* Table 2.2. contains the names of the matrices that will be used to store different adjustment data once the behaviour of the adjustment is classified for every country analysed.

Table 2.2. Matrices used to store information of different adjustments and classifications.

	Adjustment parameters	**Adjusted data**	**Derivatives**
Not clear (*1)	*param_ajustos*	*ajustos*	*mat_derivada*
Not clear (*2)	*param_ajustos_2*	*ajustos_2*	*mat_derivada_2*
Decreasing type-1 (*1)	*param_ajustos_decre*	*ajustos_decre*	*mat_derivada_decre*
Decreasing type-1 (*2)	*param_ajustos_decre_2*	*ajustos_decre_2*	*mat_derivada_decre_2*
Decreasing type-2 (*1)	*param_ajustos_decre2*	*ajustos_decre2*	*mat_derivada_decre2*
Decreasing type-2 (*2)	*param_ajustos_decre2_2*	*ajustos_decre2_2*	*mat_derivada_decre2_2*
Increasing type-1 (*2)	*param_ajustos_cre*	*ajustos_cre*	*mat_derivada_cre*
Increasing type-1 (*2)	*param_ajustos_cre_2*	*ajustos_cre_2*	*mat_derivada_cre_2*
Increasing type-2 (*1)	*param_ajustos_cre2*	*ajustos_cre2*	*mat_derivada_cre2*
Increasing type-2 (*2)	*param_ajustos_cre2_2*	*ajustos_cre2_2*	*mat_derivada_cre2_2*

**1: First adjustment; *2: Second adjustment*

The variable *min_numpoints* is set as 10, the minimal number of points required for the adjustments. Another variable, *min_pop,* is also created to stablish the minimal population required in order to proceed to the adjustment. A matrix named *dades_ajustar* is created. Depending on the value of the variable *prev_inc*, incidence or prevalence data is stored. The relative routes for storing data and a variable to store part of the string appearing on the figures, label, are also set depending on the value of the variable. For instance, the code used to store the prevalence data, setting up the label fragment and the relative routes is the following:

```
dades_ajustar(:,1)= e_prev_100k(:,1);
str_eix_y=strcat('prevalence');
dir_ajustos = strcat('ajustos_matlab/prevalence/');
dir_imatges = strcat('img/prevalence/');
```

Main loop

Before entering the main loop, the variable *initial* is set with a value of 1 and the variable *fn* is set as 0. This variable *initial* is used to indicate where to start the data screening in order to make the different adjustments for data whereas *fn* is used to count the number of figures created.

During the main loop the counter *j* is used in a *for loop* in order to screen all countries. The command used, *for j = 1:n_paisos*, will make the program to change the value of *j* one by one for the value of *n_paisos* which will be the exact number of countries.
In each loop, two auxiliary matrices are set to 0, one that will contain the data for the first adjustment and the other for the second. The variable *final* will also be set as the value of *initial*, that will be updated at the end of each loop, plus the number of data points for the current country, minus one. This is, for instance, if after one loop the value of initial is updated to 25, the data points for the next country will start on the 25^{th} position of the data vector. Then, if the current country has data for 15 years, the value of *final* will be 49. The value of final is defined with the command *final = initial + country_data(j)-1*, where *country_data* is the number of data points for the country on the *j* position in the country list.

For each country, the mean value of its population during the data period is calculated and stored in a temporary variable. If this variable is equal or bigger than the population threshold and If the data points for the corresponding country are also equal or bigger than the minimal number of points, the program starts the first adjustment.

Data adjustment

During the first and second adjustment, the temporary variables shown in table 2.3. are created.
The differences between the first and second adjustment are slight. There are but a few changes relevant on the names of the variables, vectors and matrices used, that intend to keep the data of both adjustments separated. The major rests in a sub-index, *_2*, used to differentiate between first and second adjustment.

The first step consists in creating a variable to store the maximum number of points of the adjustment (*max_numpoints*), the variables *n_fitting*, *ini* and *fin*, and creating the vectors that contain the current data points. The variable *n_fitting* is calculated as the number of different possible adjustments that can be obtained in the current data range, defined by the number of

points available from the value of the variable *initial* to the value of *final* plus two, considering that the first adjustment will be calculated two times.

A previous adjustment is obtained using the command *[coef_fitting, god] = fit(xvalues,yvalues,'poly2')*. This command is used to adjust the quadratic equation to the points contained between the values of *fin* and *ini* and retrieve the parameters of the adjustment that are then stored into the corresponding variables.

After the previous adjustment is stored, a series of loops are created. The first one will change the number of points one by one inside the range compressed between the minimal and maximum points value for the current country. During this loop, the value of *fin* is checked and if its value is equal or lesser than the final point of the data, another previous adjustment is conducted. This will make sure that the previous adjustment is different for every value of the number of points used to adjust. The parameters of the adjustments are then stored in auxiliary variables and if the r^2 value of the adjustment is higher than the previous one, the parameters are stored on their regular variables.

Then, a second loop is created inside the previous loop. Once the second previous adjustment is completed, the new loop will screen data, moving one point from left to right on each iteration. During this loop, the values of *ini* and *fin* are also updated: *ini* will increase +1 on each iteration while *fin* will be defined as *ini+numpoints+1*. The value of *fin* is checked on every iteration and if it is inside the data range, the adjustment will be conducted as previously described. If conditions are met, the parameters will be stored. This way, after the program has screened all data points, only the adjustment containing the highest r-squared parameter will be stored for that country.

Classification

Once the adjustment (first or second) is completed and before moving to the next country, the adjustment is automatically classified. Prior to the classification, the first derivative is calculated for each one of the adjustment's points with the aid of a small loop. The second derivative is also calculated. A few variables are defined: one to store the number of times that the first derivative is positive, one to store the number of times it is negative and another that will count the number of points of the adjustment and serve as a criteria of classification. Table 2.4. shows the five different types of behaviour on which adjustments can be classified and its classification criteria.

Table 2.3. Different variables used to store temporary variables during the adjustment process.

Temporary variable/s	Description
max_numpoints	Stores the maximal number of points available on the adjustment.
n_fitting	Stores the number of possible adjustments (in different points) that can be done in the current data range defined.
Ini	Auxiliary variable used to store the value of *initial*. It is also used during a loop to move the points from left to right, screen and adjust data.
Fin	Auxiliary variable used to store the final point of an adjustment.
xvalues	Stores the year values to be used on each adjustment.
yvalues	Stores the incidence or prevalence values on each adjustment.
ini_fit_point	Stores the value of the initial point of the best adjustment.
fin_fit_point	Stores the value of the final point of the best adjustment.
a; a_aux	Used to store the second degree coefficient of the quadratic equation.
b; b_aux	Used to store the first degree coefficient of the quadratic equation.
c; c_aux	Used to store the independent term of the quadratic equation.
r2; r2_aux	Used to store the r-squared of the adjustment.

Table 2.4. Different types of adjustment classification and its respective criteria for classification.

Classification	Classification criteria
Decreasing type-1	All first derivatives are negative and the second derivative is positive.
Decreasing type-2	All first derivatives are negative and the second derivative is negative.
Increasing type-1	All first derivatives are positive and the second derivative is positive.
Increasing type-2	All first derivatives are positive and the second derivative is negative.
Not clear	If the adjustment does not meet the conditions for any of the other types of behaviour, it is classified as *not clear*.

Once the adjustment is classified in one of the five types of behaviour, a graphic containing the adjustment and the epidemiological data is created. The parameters of the adjustment and other data are stored in the corresponding country's row of the *param_ajustos* matrix corresponding for the type of adjustment (see table 2.2.). The different columns are used to store a different parameter each, as shown in table 2.5.

This procedure is used in both first and second adjustment. Once the first adjustment is completed, the program will try to make a second adjustment starting five positions before the ending point of the best first adjustment. This is done with the intention of finding the closest adjustment that could manifest a different behaviour. Sometimes both adjustments can be overlapped. If the second adjustment is successful, the program will store the data and graphic from the best second adjustment, as it had done previously with the first one, and also the graphic containing both adjustments in a different image folder.

Data output

To save the adjustment parameters (*A*, *B*, *C* and R^2), different matrices named with different subscript have been used (see table 2.2.). The resulting data had a lots of 0 and also contained the names of the countries that had no adjustment or other non-relevant because the requirements for starting the adjustment were not met. In order to reduce the size of the matrices and facilitate

the reading of the exported data, different filters were created and applied to the different matrices, thus clearing all the rows were there was no relevant data and removing the empty spaces.

Once filters were applied, the data coming from the different adjustments was exported to different Excel files using the Matlab function *xlswrite*, one set for each type of behaviour considered in the classification. Each set consists in one excel file for the values of the adjustment points, another for the values of the derivative on every adjusted point and a third containing the parameters of the adjustment.

Table 2.5. Different information contained on each column of the matrix to store adjustment parameters.

Column	Stored information
1	Second degree coefficient of the quadratic equation.
2	First degree coefficient of the quadratic equation.
3	Independent term of the quadratic equation.
4	R-squared of the adjustment.
5	Type of adjustment. Takes value 1 or 2 for the corresponding adjustment type of *increasing* or *decreasing*, and 0 when it corresponds to *not clear*.
6	Initial value of incidence or prevalence.

2.3. Results

Outcome from the Matlab data adjustment

The execution time was estimated with a timer incorporated on the later versions of the script. Depending on whether Matlab had to overwrite the files generated by the script or not, the execution time was found between 25-30 minutes (using a computer with the processor Intel® Core™ i3-4005U) and approximately 700 files were produced for every parameter, including mostly images of the different adjustments and Excel spreadsheets containing different information.

Images are saved two times, in two different formats, one as the generic image format *png* and other as Matlab's figure format (*fig*).

For the incidence, 634 images and 15 spreadsheets were obtained. 162 corresponding to decreasing type-1, 42 to decreasing type-2, 174 containing double adjustments, 16 increasing type-1, 44 increasing type-2 and 196 corresponding to not-clear behaviour.

In the case of prevalence, 675 images and 15 spreadsheets were obtained. 166 corresponding to decreasing type-1, 54 to decreasing type-2, 194 containing double adjustments, 12 increasing type-1, 24 increasing type-2 and 224 corresponding to not-clear behaviour.

Table 2.6. shows the distribution of the different adjustments for both incidence and prevalence. In both cases, data from 143 countries was adjusted and the decreasing type-1 was dominant. A total of 87 double adjustments were conducted for incidence whereas 97 were obtained for prevalence. Overall, the classification of the adjustments is slightly different between incidence and prevalence. The total distribution of the different adjustments obtained for TB incidence by type of behaviour is shown in table 2.7. As seen in the table, besides most of the adjustments were classified in the category not-clear, when comparing between the four types of adjustments (increasing and decreasing types 1 and 2), the decreasing type-1 behaviour is the most frequent (35.68% over the total number of adjustments).

The number of adjustments in the increasing category is smaller when compared with the number of adjustments in the decreasing category. However, the number of adjustments classified into the *not clear* category is, in some circumstances, higher or nearer to the number of decreasing type-1 adjustments. The reason for this high number of *not clear* behaviour is because the data retrieved from the WHO is not consistent on every country in the whole period since not all countries have the resources to undergo an effective TB control and some of them do not submit valuable information. After reviewing the adjustments one by one, some of them were found not to be on the proper classification due to their overall behaviour or because they were adjusted to a small proportion of the data set and the real situation proved to have another behaviour. Thus, restrictions for the classification must be upgraded to improve adjustment classification and provide a more reliable tool for the automated analysis of epidemiological data.

The current version of this program has proved as a good tool to have a first assessment of the overall behaviour of TB epidemiological data, demonstrating that in most cases, TB has a decreasing tendency that can be adjusted to the quadratic function.

Table 2.6. Distribution of the different adjustments obtained for TB incidence and prevalence. 143 countries were adjusted from a total of 219.

	Incidence	Prevalence
Countries adjusted	**143**	**143**
First adjustments	**143**	**143**
Decreasing type-1	59	54
Decreasing type-2	21	27
Increasing type-1	6	6
Increasing type-2	16	10
Not clear	42	47
Second adjustments	**87**	**97**
Decreasing type-1	22	29
Decreasing type-2	0	0
Increasing type-1	2	0
Increasing type-2	1	1
Not clear	58	67

Table 2.7. Total distribution of the different adjustments obtained for TB incidence by type of behaviour. Absolute value (left) and % over total adjustments.

	Adjustments	% over total
Decreasing type-1	81	35.7
Decreasing type-2	21	9.2
Increasing type-1	8	3.5
Increasing type-2	17	7.5
Not clear	100	44.1
Total	227	100

Types of behaviour observed

Figure 2.9. shows two of the decreasing type-1 adjustments generated with Matlab. As seen in the figure, the script works correctly and only saves the best adjustment possible. On the particular case of Bolivia (figure 2.9., left) and China (figure 2.9., right), the behaviour of the data points is consistent and all the data could be adjusted. Instead, the Matlab adjustment stored adjustments that had not the entire data of the corresponding countries due to the restrictions imposed, i.e. that the R^2

coefficient was larger than the previous adjustment's one. Both countries show a similar behaviour even though the magnitude of their incidence is different. It is important to take into account the magnitude of the variables (incidence or prevalence) when comparing the behaviour of two different countries.

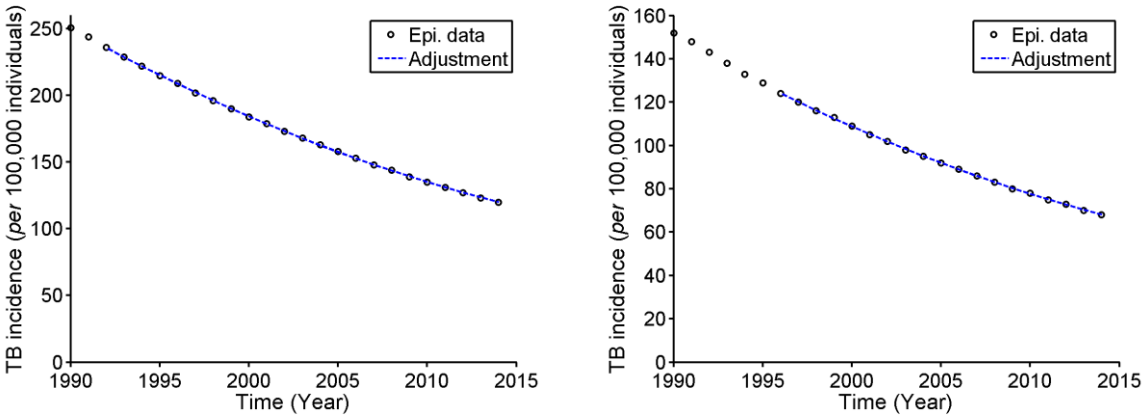


Figure 2.9. Two decreasing type-1 adjustments of TB incidence. Left: Bolivia. Right: China

In some cases, the program adjusted two different types of behaviour, as is the case of Portugal, shown in figure 2.10., where the two different types of decreasing behaviour were adjusted. In the particular case of this adjustment, if it had been done manually, the adjustment would have covered the whole dataset but it would have been difficult to distinguish between the two types of decreasing behaviour since the difference can be slightly appreciated in the image. Figure 2.11. shows a clear example of the overlapping that is observed in some cases of double adjustment. This happens due to the condition imposed for the second adjustment to start five positions before the last position of the first adjustment.

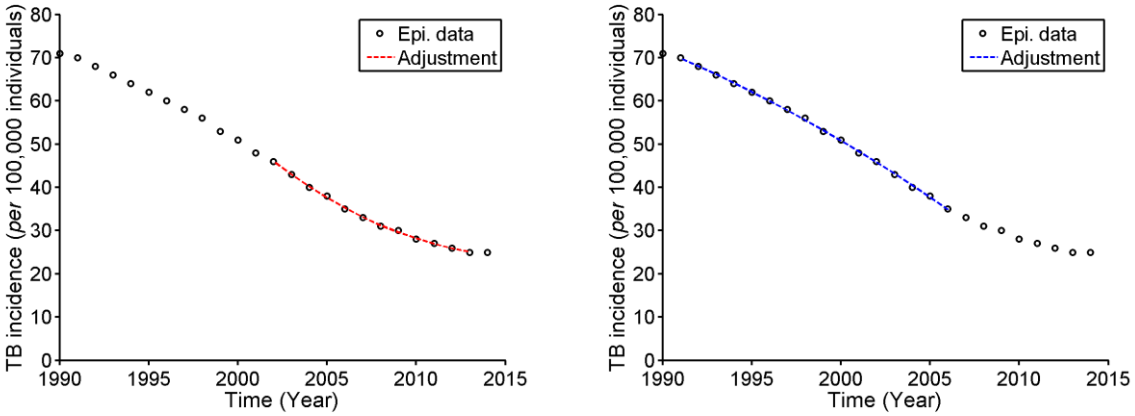


Figure 2.10. Examples of the two types of decreasing adjustments retrieved from the Matlab outcome. Both adjustments belong to Portugal. Left: Decreasing type-1. Second adjustment (red). Right: Decreasing type-2. First adjustment (blue).

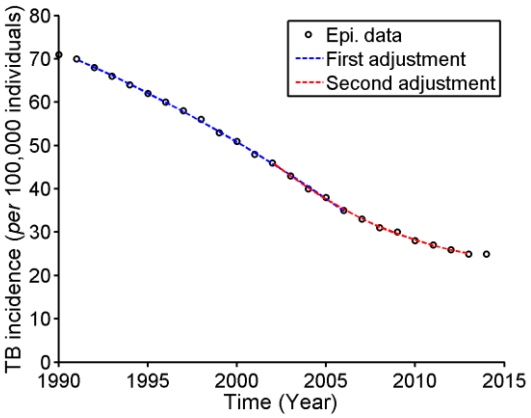


Figure 2.11. Both first and second adjustments on Portugal TB incidence.

In some cases, the behaviour of all data points proved to have not a clear tendency as can be seen in the adjustments in figure 2.12. In the particular case of Singapore, two completely different behaviours were observed and classified in the category *not clear*: while the first adjustment was similar to a decreasing type, the second one resembled an increasing one. This particular country seems to reflect a behaviour closest to a sinusoidal function. On the other hand, Serbia & Montenegro data has been adjusted to a behaviour similar to a decreasing type but the data points prove inconsistent since there is no information available after 2004.

Singapore and Serbia & Montenegro are just an example of the countries that have a behaviour in TB incidence or prevalence that is still not a clear one. This might happen for different conditions leading to less consistent data, lesser data quality or other causes.

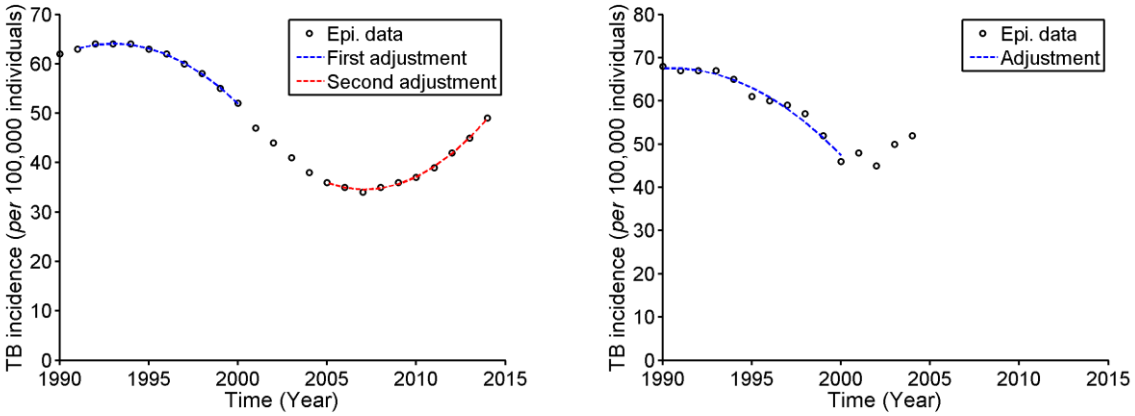


Figure 2.12. Different *not clear* adjustments. Left: Singapore. Right: Serbia & Montenegro.

Figure 2.13. illustrates two examples of the two different types of increasing behaviour in Uruguay (left) and Turkmenistan (right). In both cases, the general behaviour of the data is difficult to determine. While on Uruguay incidence starts decreasing and then increasing it does exactly the inverse process in Turkmenistan. In Uruguay the difference between the first data entry (1990) and the final data entry (2015) is slightly different because the magnitude of TB incidence is very small. On contrast, the difference between the two possible behaviours that could be adjusted in Turkmenistan is quite larger when compared with the previous case, since the magnitude of TB incidence is larger in the later. In both cases, the best classification possible would be in the category *not clear* and we would need future data to analyse the real situation of TB in the two countries.

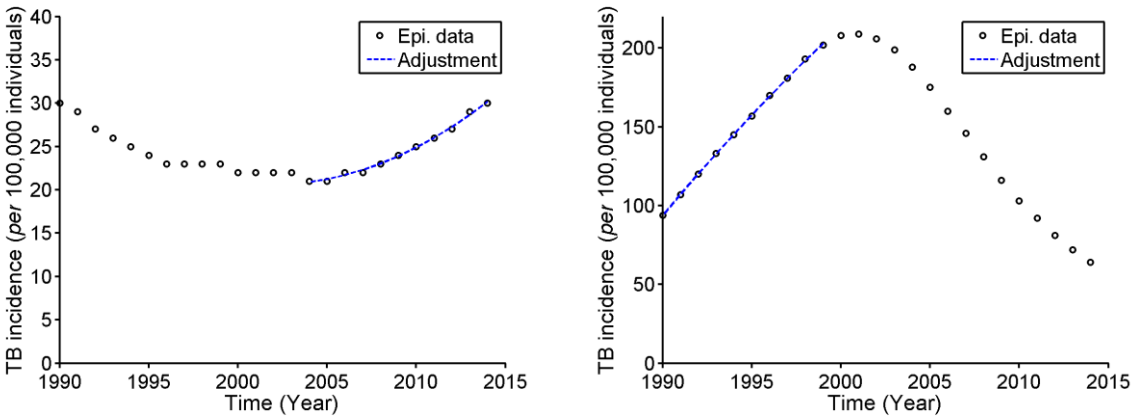


Figure 2.13. Examples of the two types of increasing adjustments.
Left: Increasing type-1 adjustment for Urugay.
Right: Increasing type-2 adjustment for Turkmenistan.

Another difficulty that the automated data adjustment needs to overcome is illustrated in figure 2.14. In some countries with a low level of TB incidence, such as Canada, the variations in the data can be very small and in some cases even unnoticeable, resembling a straight horizontal line. In many cases, data is only adjusted in a small portion of the data of a country and, sometimes, a different classification could be applied when considering the whole data range. This is the case of Cuba, were data was adjusted in a small region but it could have been adjusted for more data points.

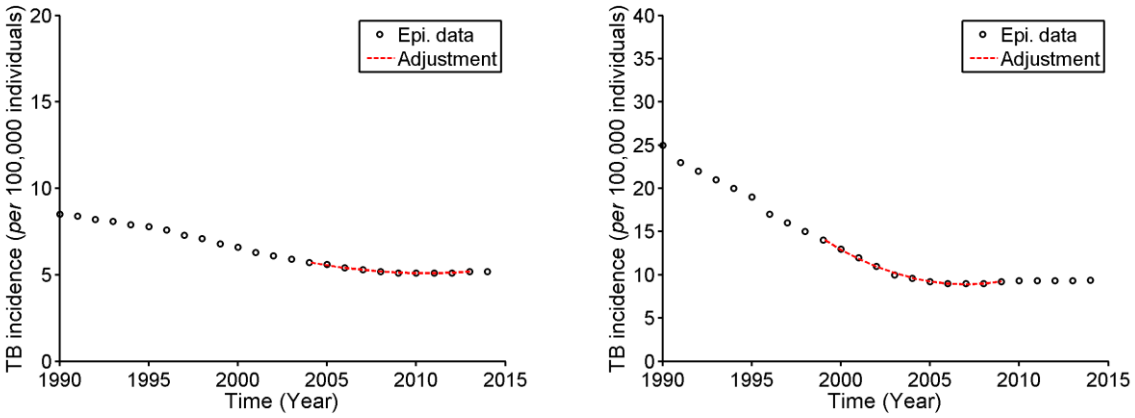


Figure 2.14. Different adjustments of incidence. Left: Canada. Right: Cuba.

Figures 2.14., 2.15. and 2.16. show a small part of the output data that is stored in the different files created with the program. The information showed in all three images was obtained for each type of adjustment considered by the program. Countries can be easily identified in the first column by their name, by their ISO-3 code in the second column or by their WHO Region in the third. The first image shows some of the points of the different adjustments of the decreasing type-1 behaviour of incidence while the second one shows the derivatives of the corresponding adjustments and the third one shows the parameters of the adjustments.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Country	ISO3	WHO Region										
2	Austria	AUT	EUR	17,8885449	16,9528896	16,0577399	15,203096	14,3889577	13,6153251	12,8821981	12,1895769	11,5374613	10,9258514
3	Azerbaijan	AZE	EUR	487,909091	410,824242	342,178788	281,972727	230,206061	186,878788	151,990909	125,542424	107,533333	97,9636364
4	Belgium	BEL	EUR	17,6628376	17,2229744	16,788845	16,3604496	15,9377882	15,5208606	15,109667	14,7042074	14,3044816	13,9104898
5	Benin	BEN	AFR	126,709091	120,072727	114,057576	108,663636	103,890909	99,7393939	96,2090909	93,3	91,0121212	89,3454545
6	Bolivia (Plur	BOL	AMR	235,76087	228,766798	221,937041	215,271598	208,770469	202,433653	196,261152	190,252964	184,409091	178,729531
7	Brazil	BRA	AMR	83,8545455	81,0121212	78,3060606	75,7363636	73,3030303	71,0060606	68,8454545	66,8212121	64,9333333	63,1818182
8	Bulgaria	BGR	EUR	54,4818182	50,9787879	47,5742424	44,2681818	41,0606061	37,9515152	34,9409091	32,0287879	29,2151515	26,5
9	Burundi	BDI	AFR	270,745455	249,048485	229,275758	211,427273	195,50303	181,50303	169,427273	159,275758	151,048485	144,745455
10	Cambodia	KHM	WPR	511,954545	495,342424	479,480303	464,368182	450,006061	436,393939	423,531818	411,419697	400,057576	389,445455
11	Canada	CAN	AMR	7,09642857	6,81900599	6,56091409	6,32215285	6,10272228	5,90262238	5,72185315	5,56041459	5,41830669	5,29552947
12	Central Afric	CAF	AFR	842,1	760,948485	687,228788	620,940909	562,084848	510,660606	466,668182	430,107576	400,978788	379,281818
13	China	CHN	WPR	124,06015	120,092732	116,225711	112,459089	108,792865	105,227038	101,76161	98,3965797	95,1319475	91,9677134
14	China, Hong	HKG	WPR	103,043956	100,43956	97,8651349	95,3206793	92,8061938	90,3216783	87,8671329	85,4425574	83,047952	80,6833167

Figure 2.14. Small part of the output containing the values of the points of the different adjustments for incidence decreasing type-1.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Country	ISO3	WHO Region										
2	Austria	AUT	EUR	-0,95590815	-0,91540248	-0,8748968	-0,83439112	-0,79388545	-0,75337977	-0,7128741	-0,67236842	-0,63186275	-0,59135707
3	Azerbaijan	AZE	EUR	-81,3045455	-72,8651515	-64,4257576	-55,9863636	-47,5469697	-39,1075758	-30,6681818	-22,2287879	-13,7893939	-5,35
4	Belgium	BEL	EUR	-0,44273021	-0,43699628	-0,43126236	-0,42552843	-0,4197945	-0,41406057	-0,40832664	-0,40259272	-0,39685879	-0,39112486
5	Benin	BEN	AFR	-6,9469697	-6,32575758	-5,70454545	-5,08333333	-4,46212121	-3,84090909	-3,21969697	-2,59848485	-1,97727273	-1,35606061
6	Bolivia (Plur	BOL	AMR	-7,07622812	-6,91191417	-6,74760023	-6,58328628	-6,41897233	-6,25465839	-6,09034444	-5,92603049	-5,76171654	-5,5974026
7	Brazil	BRA	AMR	-2,91060606	-2,77424242	-2,63787879	-2,50151515	-2,36515152	-2,22878788	-2,09242424	-1,95606061	-1,81969697	-1,68333333
8	Bulgaria	BGR	EUR	-3,55227273	-3,45378788	-3,35530303	-3,25681818	-3,15833333	-3,05984848	-2,96136364	-2,86287879	-2,76439394	-2,66590909
9	Burundi	BDI	AFR	-22,6590909	-20,7348485	-18,8106061	-16,8863636	-14,9621212	-13,0378788	-11,1136364	-9,18939394	-7,26515152	-5,34090909
10	Cambodia	KHM	WPR	-16,9871212	-16,2371212	-15,4871212	-14,7371212	-13,9871212	-13,2371212	-12,4871212	-11,7371212	-10,9871212	-10,2371212
11	Canada	CAN	AMR	-0,28708791	-0,26775724	-0,24842657	-0,2290959	-0,20976523	-0,19043457	-0,1711039	-0,15177323	-0,13244256	-0,11311189
12	Central Afric	CAF	AFR	-84,8674242	-77,4356061	-70,0037879	-62,5719697	-55,1401515	-47,7083333	-40,2765152	-32,844697	-25,4128788	-17,9810606
13	China	CHN	WPR	-4,01761757	-3,91721952	-3,81682147	-3,71642341	-3,61602536	-3,5156273	-3,41522925	-3,3148312	-3,21443314	-3,11403509
14	China, Hong	HKG	WPR	-2,61938062	-2,58941059	-2,55944056	-2,52947053	-2,4995005	-2,46953047	-2,43956044	-2,40959041	-2,37962038	-2,34965035

Figure 2.15. Small part of the output containing the values the first derivative of the different adjustments for incidence decreasing type-1.

	A	B	C	D	E	F	G	H	I
1	Country	ISO3	WHO Region	a coeff.	b coeff.	c coeff.	r2	Adjustment type	Initial incidence
2	Austria	AUT	EUR	0,02025284	-81,8862487	82777,2032	0,9954644	1	21
3	Azerbaijan	AZE	EUR	4,21969697	-16985,4106	17092803,1	0,99978529	1	319
4	Belgium	BEL	EUR	0,00286696	-11,8532467	12252,1599	0,9875342	1	18
5	Benin	BEN	AFR	0,31060606	-1243,15909	1243982,24	0,99914072	1	127
6	Bolivia (Plur	BOL	AMR	0,08215697	-334,38961	340335,736	0,9999451	1	251
7	Brazil	BRA	AMR	0,06818182	-274,274242	275882,779	0,99915614	1	84
8	Bulgaria	BGR	EUR	0,04924242	-201,014394	205132,565	0,99733065	1	35
9	Burundi	BDI	AFR	0,96212121	-3873,06818	3897945,88	0,9997094	1	164
10	Cambodia	KHM	WPR	0,375	-1520,73712	1542080,51	0,99980878	1	584
11	Canada	CAN	AMR	0,00966533	-38,9097652	39164,7527	0,99897292	1	8,5
12	Central Afric	CAF	AFR	3,71590909	-14970,7992	15079092,3	0,99996923	1	864
13	China	CHN	WPR	0,05019903	-204,412133	208136,951	0,99979524	1	152
14	China, Hong	HKG	WPR	0,01498501	-62,6193806	65407,044	0,99879375	1	127

Figure 2.16. Small part of the output containing the different parameters of the different adjustments for incidence decreasing type-1.

Global analysis

Another Matlab script was implemented to analyse the output of the *adjustment program*. The script retrieved the adjustment and parameter data (figures 2.14. and 2.16.) and generated different plots showing fragments of the same length (10 years or data points) of the different adjustment types and comparing the second degree coefficient (*A*) in front of the initial incidence or prevalence. Since in the mathematical model only incidence data with a behaviour corresponding

to the decreasing type-1 will be properly adjusted, only the most relevant results of the incidence are shown. Similar results are obtained for the prevalence but the differences are slight and not relevant for this work. Figure 2.17. shows the different decreasing type-1 adjustments of the incidence. On the left image, some of the adjustments cannot be distinguished due to the high amount of countries with a low-range TB incidence that have been adjusted. The image on the right is an enlargement of the previous image, only showing the countries that had an initial incidence lower than 100 (*per* 100,000 individuals). On the other hand, figure 2.18. shows a clearer detail of the adjustments with initial incidence lower than 50 (left) and 25 (right). Although the magnitude of the incidence is different between countries, when the images are enlarged, it is observed that, except for some countries, the most part of the adjustments show a similar behaviour.

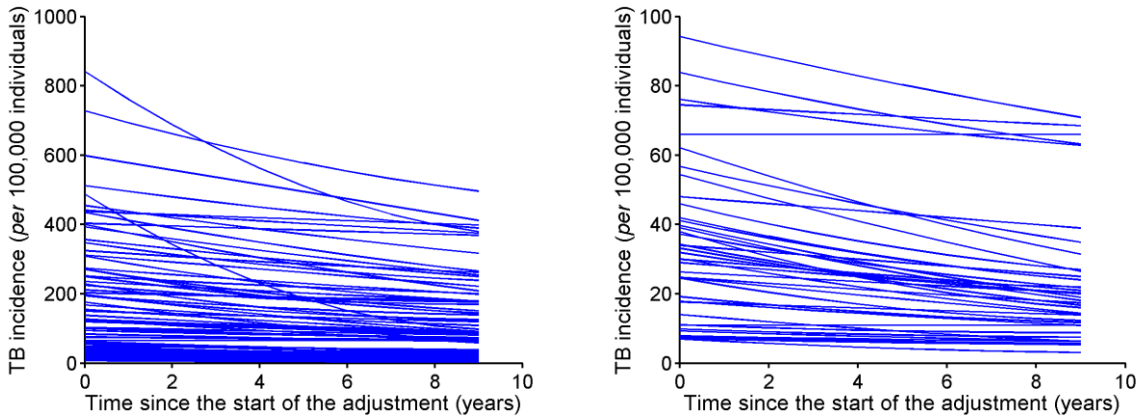


Figure 2.17. Different adjustments of incidence decreasing type-1. Left: all adjustments. Right: adjustments that had an initial incidence (of the adjustment) lesser than 100 (*per* 100,000 individuals).

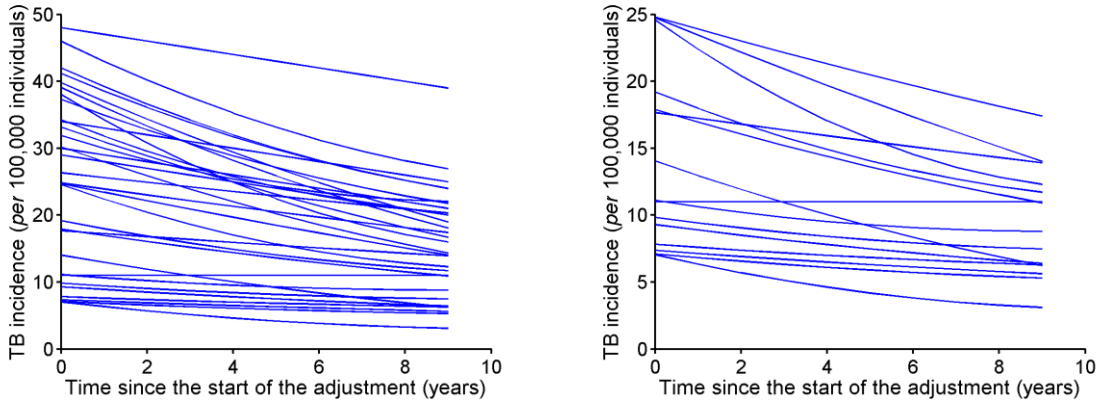


Figure 2.18. Different adjustments of incidence decreasing type-1. Left: adjustments with an initial incidence lesser than 50. Right: adjustments that had an initial incidence lesser than 25.

As shown in figure 2.19. (left), the number of adjustments in the decreasing type-1 is so large that an enlargement (figure 2.19., right) is required. In this graphic, the second grade coefficient of every adjustment is represented in front of its corresponding *A* coefficient. The presence of countries with a strange behaviour compared with the rest of countries, that are located in the extremes of the graphic, such as Azerbaijan or the Central African Republic (CAR), difficult data analysis. The behaviour of both Azerbaijan and the CAR is characterised by a high TB incidence and a fast decline over the later years. Most of the countries of the decreasing type-1 adjustment are comprised in the range between 0 to 200 initial incidence (*per* 100,00 individuals) and 0 to 0.2 for the second degree coefficient. The values of the second degree coefficient are positive because the second derivative in this classification is negative.

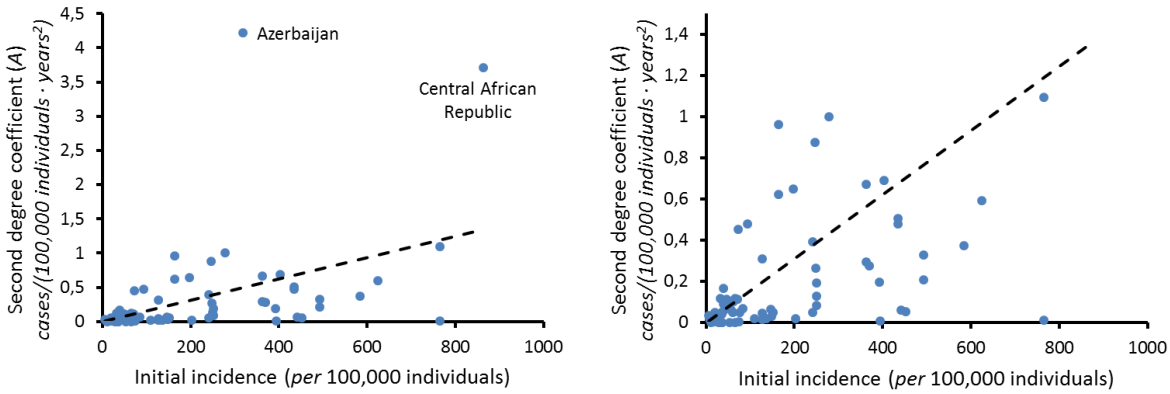


Figure 2.19. Second degree coefficient (*A*) in front of the initial incidence for decreasing type-1. Left: all adjusted countries. Right: adjusted countries with a second degree coefficient lesser than 1.2. The positive a coefficient grants a positive second derivative.

In figure 2.20. the different decreasing type-2 adjustments for TB incidence (left) and the second degree coefficient (*A*) is shown in from of initial incidence (right) are shown. Similar to the previous situation in figure 2.17., there are many adjustments in countries with a lower incidence. However, the number of adjustments is considerably lesser than in the decreasing type-1, as can be observed comparing both figures. In the particular case of the decreasing type-2 adjustments, the second degree coefficient was equal to 0 or negative as shown in the figure 2.20. (right). In the case of this type of adjustment, the second degree coefficients are negative because the second derivative of the quadratic function ($2 \cdot A$) must be negative.

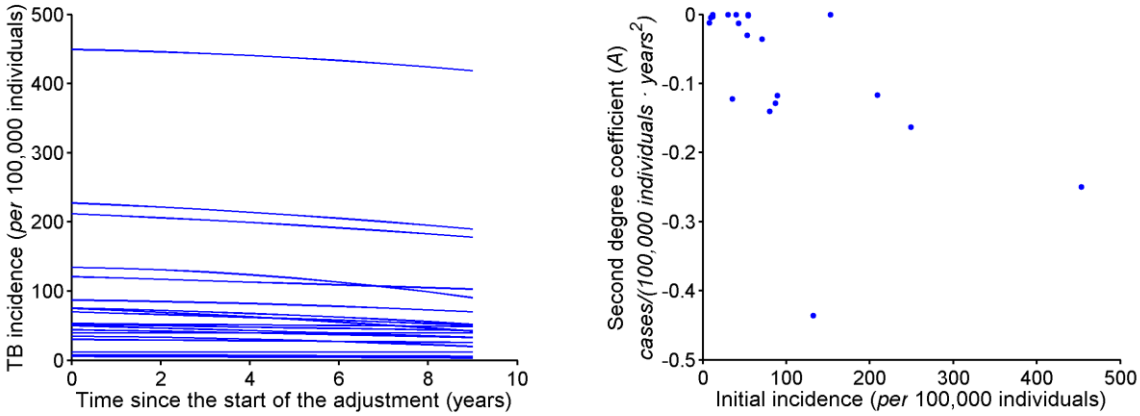


Figure 2.20. Different results for the decreasing type-2 behaviour. Left: all incidence adjustments for this type of behaviour. Right: second degree coefficient (*A*) in front of the initial incidence for decreasing type-2. The negative a coefficient grants a negative second derivative.

In figure 2.21. the second degree coefficient is shown in front of the initial incidence for the increasing type-1 (left) and type-2 adjustments (right). The number of adjusted countries is, in both cases, smaller than the number of adjusted countries of the decreasing type-1. The points of the plots are dispersed, suggesting very different situations: countries with a small magnitude of the second degree coefficient and low initial incidence, or with a bigger magnitude of the second degree coefficient and higher initial incidence.

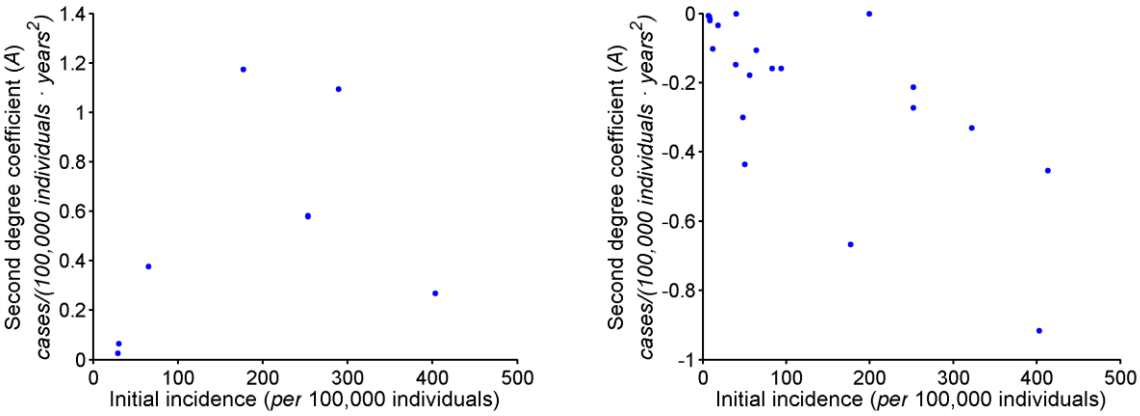


Figure 2.21. Second degree coefficient (*A*) in front of the initial incidence.
Left: increasing type-1. The positive *A* coefficient grants a positive second derivative. Right: increasing type-2. The negative a coefficient grants a negative second derivative.

3. Mathematical model

3.1. Description of the mathematical model

Introduction

The main objective of this work is to develop a simple model to help understand the temporal dynamics of TB incidence that are observed in the epidemiological data available. The model must allow to detect which are the strategic aspects that determine the behaviour of the system. This should finally allow to define a methodology for assessing the quality of public programs of disease control. The work published by the BIOCOMSC research group that presents an IbM model to study the epidemiological behaviour of the disease at city level (Prats, Montañola-Sales, et al., 2016) is taken as a starting point.

Based on previous work, and with the aim of developing a model able to study the evolution of the disease at regional or national level, a compartment-based model, with a time step of one year, was developed.

Variable definition

Based on the classical mathematical modelling methods in epidemiology, similarly to SEIR models, a system where two types on individuals is considered:

- Individuals with latent infection (E).
- Individuals with active disease (I).

The number of susceptible individuals is considered not limiting, this way there is no need to control them in the mathematical model. However, in some situations with a high incidence it may be necessary to include it. Considering the presence of a large number of susceptible people allows to avoid the need of controlling the individuals that have gained some immune capacity. However, this is a perspective with many questions.

For infected individuals (E), the probability of getting sick is considered to decrease progressively during seven years (P. J. Cardona & Ruiz-Manzano, 2004). For this reason, in a previous work (Prats,

Montañola-Sales, et al., 2016), people with latent infection were distributed into seven compartments, corresponding to the first seven years since the individuals became infected.

In the model, the population considered is distributed in 9 compartments, 8 of which will include the people with a latent infection (E_i). The ninth compartment will correspond to the individuals with an active disease (I).

In figure 3.1., the schematic diagram of the whole model structure:

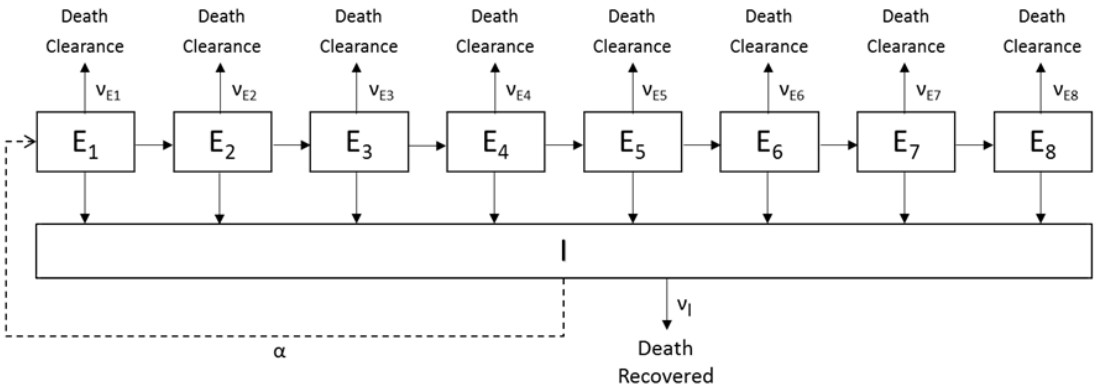


Figure 3.1. Diagram of the structure of the compartment-based mathematical model.

The time from initial infection expressed in years (t_{inf}) serves to define the *i* index corresponding to the 8 compartments of the population with latent infection.

$$i = Integer(t_{inf}) + 1 \quad if \quad t_{inf} < 7\ years \tag{3.1}$$

$$i = 8 \quad if \quad t_{inf} \geq 7\ years \tag{3.2}$$

Thus, the number of people with latent infection, depending on the elapsed time and the initial infection evolve over time *t* can be expressed:

$$E_i(t) \quad for \quad i = 1 - 8 \tag{3.3}$$

Where time *t* is a natural number expressed in years. t_0 will be the year corresponding to the initial moment. The value of *t* will increase discrete unit every year.

The number of people with active disease will be:

$$I(t) \tag{3.4}$$

Structure of the model

The population of each compartment will vary from year to year (*t*). Considering the population $E_{i-1}(t)$ for $1 < i < 8$, part of the population will die (*mortality of the population*) or will spontaneously remove TB infection (*clearance*). The factor that determines the speed with which both processes happen is named $\nu_{E_{i-1}}$. Initially considering:

$$\nu_E = \nu_{E_1} = \nu_{E_2} = \nu_{E_3} = \cdots \nu_{E_7} \tag{3.5}$$

The speed with which *clearance* is performed does not need to necessarily be constant over the years, perhaps it can be growing.

Moreover, a part of the population $E_{i-1}(t)$ will become ill (will pass into compartment I) with a probability p_{i-1}. Figure 3.2. schematically shows this behaviour:

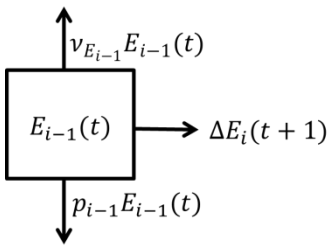


Figure 3.2. Scheme of the outflows in the compartment $E_{i-1}(t)$. The upper arrow represents the outflow corresponding to the death/clearance whereas the lower arrow corresponds to the individuals that become infected. The arrow to the right refers to the number of individuals that will move to the next compartment (advance one year).

The population that does not die, does not eliminate the parasite nor becomes ill, will become part of the next compartment.

$$\Delta E_i(t+1) = E_{i-1}(t) - \nu_{E_{i-1}} E_{i-1}(t) - p_{i-1} E_{i-1}(t) \quad for \quad 1 < i < 8 \tag{3.6}$$

The population of the compartments $i = 2$ to $i = 7$ is determined by:

$$E_i(t+1) = \Delta E_i(t+1) \tag{3.7}$$

The population of the compartment $i = 8$ will be:

$$E_8(t+1) = E_8(t) + \left(E_7(t) - \nu_{E_7} E_7(t) - p_7 E_7(t)\right) - \nu_{E_8} E_8(t) - p_8 E_8(t) \tag{3.8}$$

Considering:

$$\nu_{E_8} > \nu_E \tag{3.9}$$

For the eighth compartment the model considers, on one hand, that the number of people that eliminate the parasite is not negligible and, on the other hand, that the average age of this compartment is substantially higher than on any other and, thereby, it is conceivable that it will have a higher mortality.

People with latent infection that become sick determine the population with active disease:

$$I(t+1) = I(t) + \sum_{i=1}^{8} p_i E_i\,(t) - \nu_I I(t) \tag{3.10}$$

Where ν_I includes the fraction of people that as a consequence of the treatment stop being sick, the fraction of people that are still sick (for instance, with resistant strains) but stop transmitting the infection because they are isolated, and the people that die as a result of the disease or other causes. Initially, the model considers:

$$\nu_I = 1 \tag{3.11}$$

Thus, the mathematical expression to determine the population with active disease will simplify to:

$$I(t+1) = \sum_{i=1}^{8} p_i E_i\,(t) \tag{3.12}$$

Therefore, the model considers that the compartment of sick people will be renewed annually. This hypothesis will not be correct in those environments where the health system is not conducting a minimally correct diagnosis, treatment and monitoring (TB control), where $\nu_I < 1$.

People with active disease are those that transmit the disease. We consider that each sick person can infect an average of α_0 persons. This value depends on several factors: (1) the social structure (average number of people living in the same house, the organization of schools or workplaces, features of leisure activities, transport systems…), (2) the diagnostic delay and (3) the continuity and efficiency of treatments.

If we have a good system of epidemiological surveillance, studies will be conducted to identify persons who have been infected. We consider that $\Delta\alpha$ is the number of infected persons that are detected and start and end treatment. Thus, if we consider the difference between the potentially infected and the persons detected and treated, the average effective number of people infected by a single sick individual will be:

$$\alpha = \alpha_0 - \Delta\alpha \tag{3.13}$$

This value determines how many people constitute the compartment E_1.

$$E_1(t+1) = \alpha\, I(t) \tag{3.14}$$

It can also be expressed:

$$E_1(t+1) = \alpha \sum_{i=1}^{8} p_i E_i\,(t) \tag{3.15}$$

Probability of becoming ill

The function that defines the probability that a person with latent infection has to become ill is a key feature in the model. We know that the probability decreases with the time elapsed since the initial infection, but we do not know which function would be a good approximation of the real behaviour. We have used a decreasing parabolic function since it seems reasonable to assume that the probability is much higher in the first year and then it declines rapidly, but we also know that even after ten years from the initial infection, a person can become ill. This indicates that the decline

is probably not exponential. For this reason, a parabolic decrease is taken as a reasonable behaviour, considering:

$$p_i = a\, i^2 + b\, i + c \quad for \quad i = 1 - 7 \tag{3.16}$$

$$p_8 = c_8 \approx \frac{c}{20} \tag{3.17}$$

Taking a low value for p_8 seems reasonable given that the eighth compartment includes people with several years of infection. So, in addition to the progressive parabolic decrease, this probability will be divided by the average number of years. This will only serve to assess the order of magnitude of c_8 since it is difficult to know the exact number of years that people remain in the eighth compartment.

Brief review

The structure of the model is really very simple, however the model uses 10 variables $(E_1, \dots, E_8, I, t)$ and 13 parameters $(\nu_1, \dots, \nu_8, a, b, c, c_8, \alpha)$. This means that we can adjust, for example, a parabolic decline using different sets of parameters. The values for the parameters must be reasonable. Yet, the model is valuable not for the ability to assess the values of the parameters that will be used but because it provides an understanding of the dynamic behaviour of the systems.

The model can be implemented in various programing languages (C++, Python...), with calculation tools (Matlab…) or even with Excel spreadsheets.

3.2. Adjustment of the model

The results obtained from the program for the automated adjustment of epidemiological data showed that it was possible to adjust a parabolic equation to the data. Furthermore, the most common behaviour on data was a decreasing type. In order to adjust the mathematical model to a decreasing behaviour, an Excel spreadsheet was used. On a first approximation, the ratio at which the individuals moved from the compartments by the cause of death or clearance were considered equal for all countries adjusted. The parameters responsible of the probability of becoming ill (*a*, *b*, *c* and c_8) were also assumed equal for all countries. The values of *clearance* are not known. This is

why they are estimated altogether with mortality. The initial parameters for all adjustments are presented in table 3.1.

Table 3.1. Initial parameters considered for the model adjustment.

$\boldsymbol{a}$	$0.001035\ (\frac{cases}{100{,}000\ ind.\cdot\ years^2})$
$\boldsymbol{b}$	$-0.0152\ (\frac{cases}{100{,}000\ ind.\cdot\ years})$
$\boldsymbol{c}$	$0.06\ (\frac{cases}{100{,}000\ ind.})$
$\boldsymbol{c_8}$	$0.000288\ (\frac{cases}{100{,}000\ ind.})$
$\boldsymbol{\nu_{E_{1-7}}}$	0.01 (1%)
$\boldsymbol{\nu_8}$	0.35 (35%)
$\boldsymbol{\nu_I}$	1 (100%)

The first step in the adjustment procedure consisted in finding the stationary state of the system, which is achieved when the number of exposed (infected) individuals and the number of sick people (infectious patients) remains constant over time. This is achieved creating a structure similar to the one shown at table 3.2. The first row represents the duration of the infection (T_i), in years. The second row contains the initials conditions of the system, which must be guessed since there is not certain information. A first approach is conducted for the initial conditions of every compartment, considering that for compartments 2 to 7 the number of infected will decline when compared with the previous compartment, and that the number of infected in the eighth compartment is much larger. This is because the infected elapse more time in the eighth compartment. The remaining rows showed in figure 3.3. consist in the implementation of the model for only one time-step, using the parameters from table 3.1. and taking the guessed initial conditions as the data from the previous year in order to calculate the number of infected at the first year of evolution.

The model is implemented for several years retrieving the initial conditions from the previous approximation. A plot is created in order to compare the evolution of the model in contrast of the

experimental data, thus allowing a visual support that aids in the process of adjustment. The value of α in the stationary state (α_{sta}) is also guessed, modifying its value in order to look for a horizontal straight line (figure 3.3., right) . This is, that the difference of sick individuals between the successive years of simulation of the model must be minimal or 0 if possible. However, the initial incidence is different than on the model and at the end we want to adjust the model to epidemiological data. Thus, the values of the initial conditions are modified taking as a reference the ones obtained for the first year of evolution in the model implementation. After some iterations the stationary state is finally achieved in the initial incidence of the epidemiological data. This is illustrated in figure 3.3., that shows the first (left) and the final (right) approach of the stationary state.

Table 3.2. Basic structure used to find the stationary level and adjust the model.

Duration of the infection (years)	Initial conditions (number of infected)	Number of infected (first year of the evolution)	Probability of becoming ill	Sick (new ill)	New infections (new infected)	Death / clearence (People with 8 or more years of infection that are no longer infected)
t_{inf}	E_i (t=0)	E_i (t=1)	p_i	$p_i E_i$	$\alpha p_i E_i$	vi Ei
1	215,00	199,45	0,05	9,89	67,94	2,15
2	203,11	202,96	0,03	6,91	47,44	2,03
3	196,13	194,17	0,02	4,71	32,34	1,96
4	193,24	189,47	0,02	3,09	21,24	1,93
5	185,03	188,21	0,01	1,85	12,71	1,85
6	176,35	181,33	0,01	1,06	7,27	1,76
7	172,19	173,53	0,00	0,69	4,73	1,72
≥7	4200,00	2898,94	0,00	0,84	5,77	1470,00

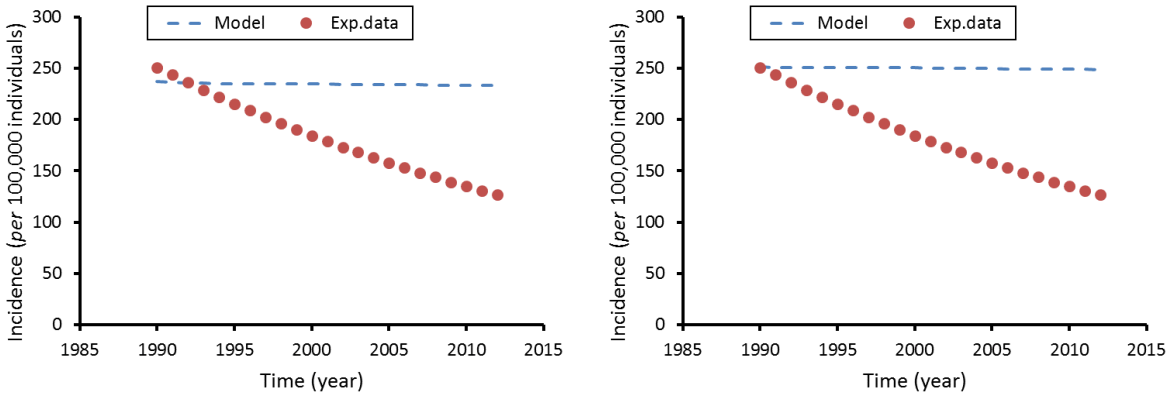


Figure 3.3. The stationary state in Bolivia. Left: a first approach, not adjusted to the initial incidence. Right: a final approach, adjusted to the initial incidence.

The final step to achieve the adjustment of the model consists in using a solver methodology in order to minimize the difference between the model and the epidemiological data (18) by changing the value of α, since all previous parameters have already been fixed. Figure 3.4. shows an example of the adjustment of the model for the epidemiological data of Bolivia.

$$Minimze\left(\sum|epidemiological\ data - model\ data|\right) \tag{3.18}$$

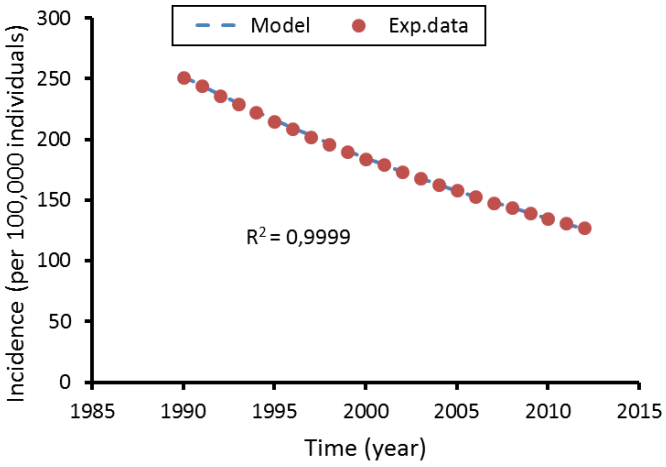


Figure 3.4. Final model adjustment achieved for Bolivia. R stands for the correlation between the epidemiological data and the model.

3.3. Analysis of the model behaviour

The aim of this section is to assess the effect of various parameters used in the model. To do this, we have found a set of values to adjust the behaviour of Barcelona (table 3.1.) and we have changed each parameter one by one to discuss their effects.

Effects of the parameters *a*, *b* and *c*

The probability of becoming sick for the first seven compartments is defined by a parabolic function that depends on parameters *a*, *b* and *c*. On the other hand, the probability of becoming sick is much smaller for the eighth compartment and depends on the value of c_8. The effect of the three parameters (*a*, *b* and *c*) is analysed, taking the adjustment of the mathematical model in Barcelona, shown in figure 3.5., as a contrasting reference. Figures 3.6., 3.7. and 3.8. show the effects of the increase (left) and decrease (right) of parameters *a*, *b* and *c* respectively.

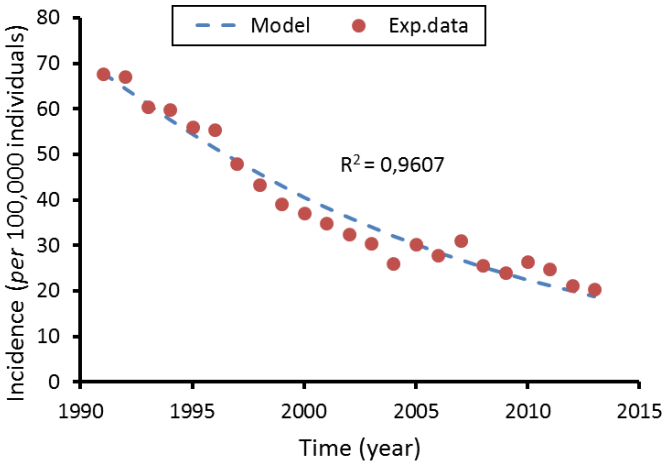


Figure 3.5. Adjustment of the mathematical model for TB incidence in Barcelona for the period 1991-2014.

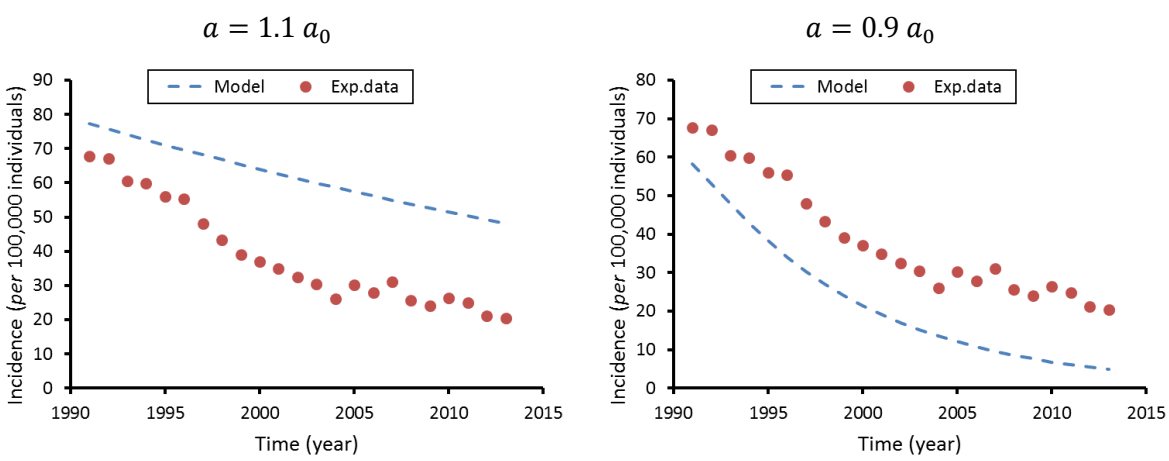


Figure 3.6. The effects of parameter *a* on the model. a_0 corresponds to the initial value of the parameter, presented in table 3.1. Left: increases a 10% from its initial value. Right: decreases a 10% from its initial value.

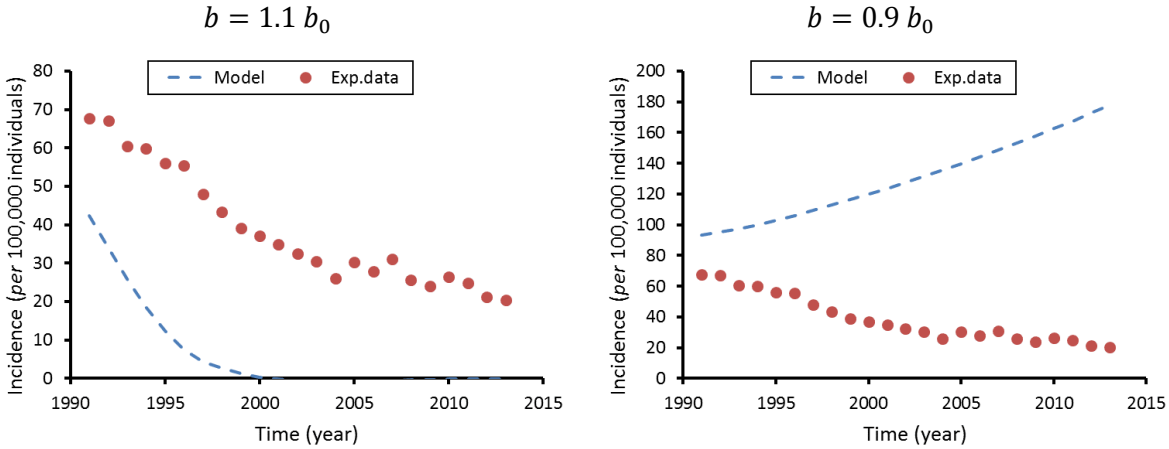


Figure 3.7. The effects of parameter *b* on the model. b_0 corresponds to the initial value of the parameter, presented in table 3.1. Left: increases a 10% from its initial value. Right: decreases a 10% from its initial value.

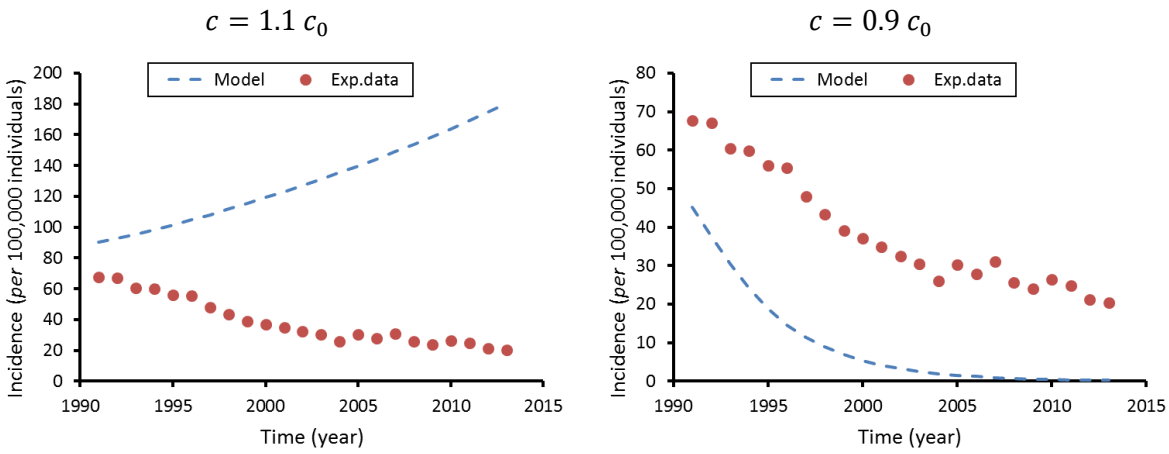


Figure 3.8. The effects of parameter *c* on the model. c_0 corresponds to the initial value of the parameter, presented in table 3.1. Left: increases a 10% from its initial value. Right: decreases a 10% from its initial value.

The behaviour observed states that the mathematical model is highly sensible to the probability of becoming ill (p_i) that is determined by the values of *a*, *b* and *c*. In the adjustment obtained, the probability was about 14%, perhaps a too high value. If the probability of becoming ill declines, the population with latent infection must be increased in order to keep the adjustment.

The probability of becoming ill varies depending on certain risk factors (HIV, smokers, diabetes…), but outside of these factors it seems reasonable to take a similar value for all locations.

Effects of the *α* parameter

The *α* parameter strongly depends on the organization in public health. There is a value (α_{sta}) at which the incidence remains constant over time (stationary state). Figure 3.9. shows the final adjustment (left) and the stationary state (right) for the city of Barcelona. On the other hand, figure 3.10. the behaviour of the system when α is increased (left) and decreased (right) 15% from the value of α_{sta}.

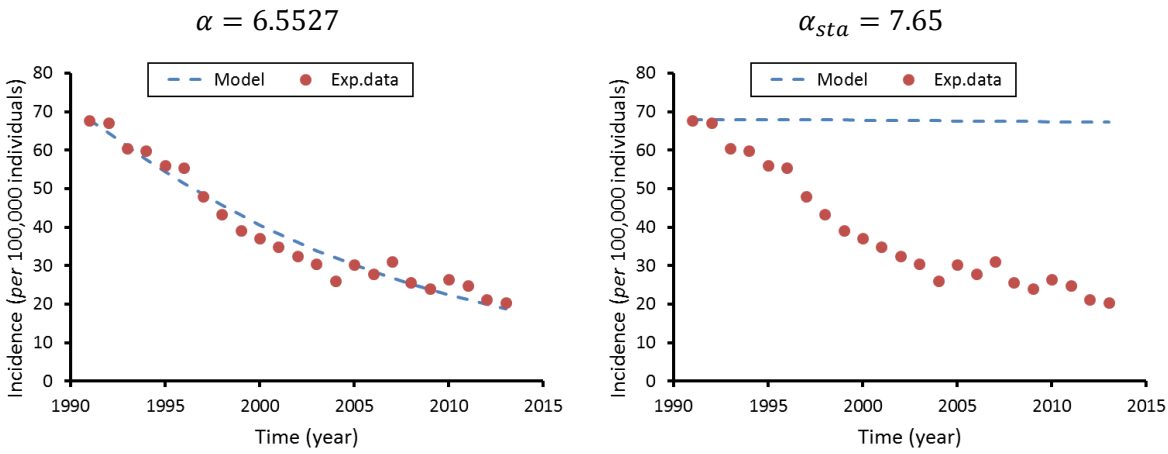


Figure 3.9. Both best adjusted model (left) and its stationary state (right). α is the average effective number of people infected by a single sick individual. α_{sta} corresponds to the value of α that is necessary to achieve the stationary state of the system.

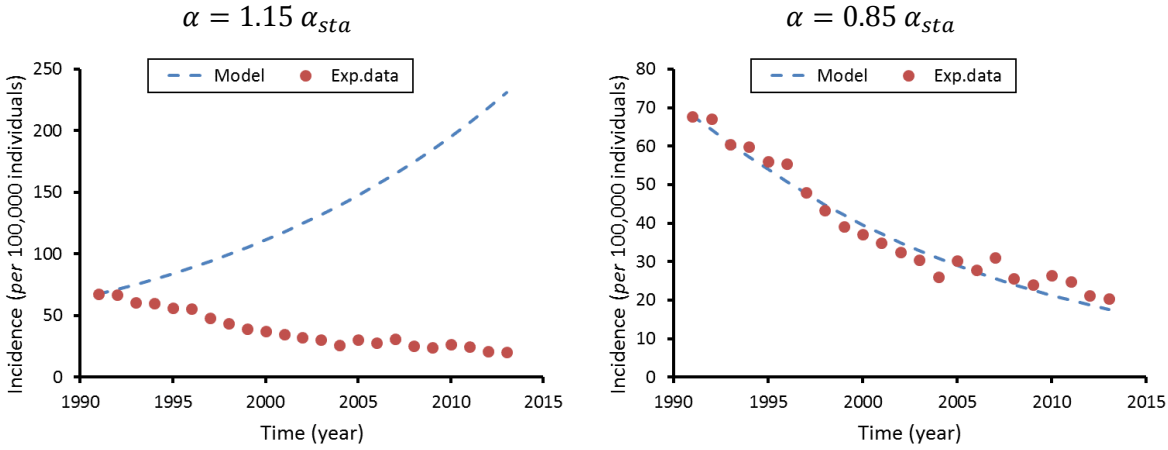


Figure 3.10. The effects of parameter α on the model. Left: increases a 15% from the value of α_{sta}. Right: decreases a 15% from α_{sta}.

If we consider a value 15% higher, the growth of the incidence is considerably higher. In contrast, if we consider a value 15% lower, the behaviour of the model is more approximated to the conditions in the city of Barcelona. It is noted that the value of α necessary to reproduce the evolution of the incidence in certain populations is something to take into consideration to assess the quality of the work in public health.

Effect of *mortality/clearance*

To evaluate the effects of mortality and clearance, the behaviour of the model was tested multiplying and dividing by two the values of the initial mortality/clearance (ν_0) of the eight compartments. Figure 3.11. shows the changes observed in the model behaviour when ν of the 8 compartments is multiplied by 2 (left) and when it is divided by 2 (right).

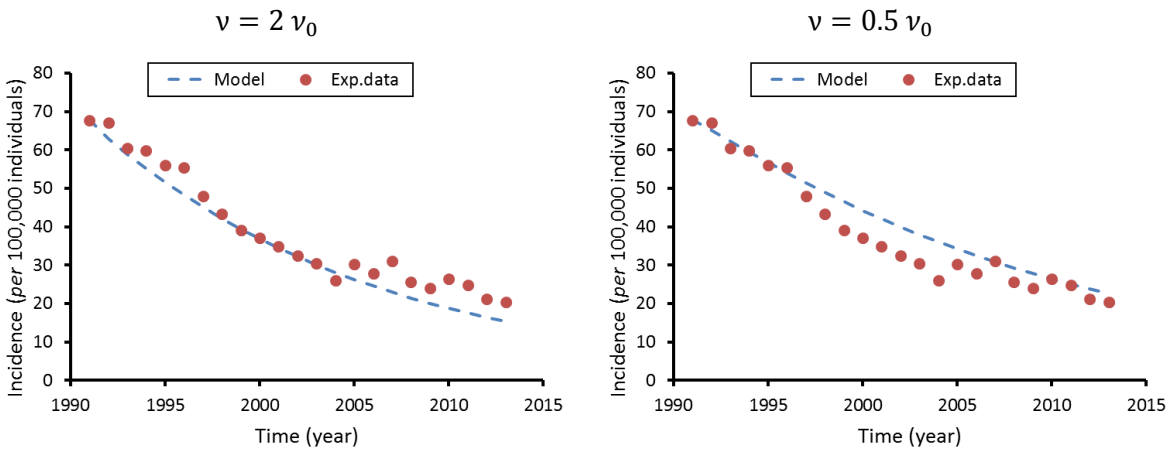


Figure 3.11. The effects of *mortality/clearance (ν)* on the model. Left: ν is doubled for all 8 compartments. Right: ν is halved for all 8 compartments.

The effect of *mortality/clearance* (ν) is insignificant. Their values must be greatly changed (increased or decreased) in order to observe significant changes on the model behaviour. Certainly mortality is different in each town, so reasonable values can be included in order to be able to adjust each location.

Effects of ν_8

The effects of *mortality/clearance* in the eighth compartment (ν_8) are correlated with the parabolic behaviour. Figure 3.12. shows the changes on the behaviour of the model that are obtained by duplicating (left) and halving (right) the value of ν_8.

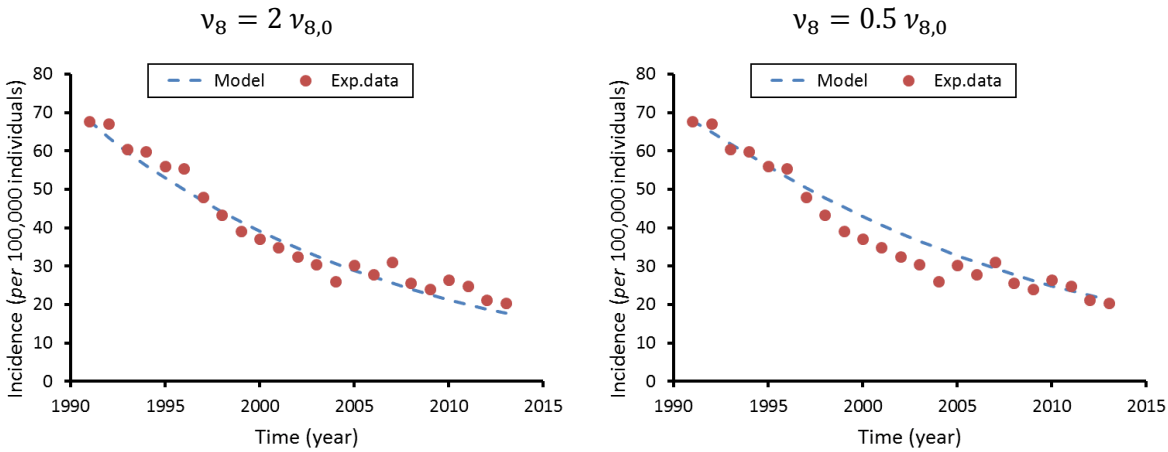


Figure 3.12. The effects of *mortality/clearance* of the eight compartment *(ν_8)* on the model. Left: ν_8 is double. Right: ν_8 is halved.

As the value of ν_8 increases, the opening of the parabolic function decreases. On the other hand, if the value decreases, the value of incidence on the model will increase slightly. The effects of ν_8 on the system are not very relevant.

Figure 3.13. shows the experimental data from Barcelona and the adjusted model (left) and a comparison of the model adjustment with an exponential and a parabolic adjustment (right).

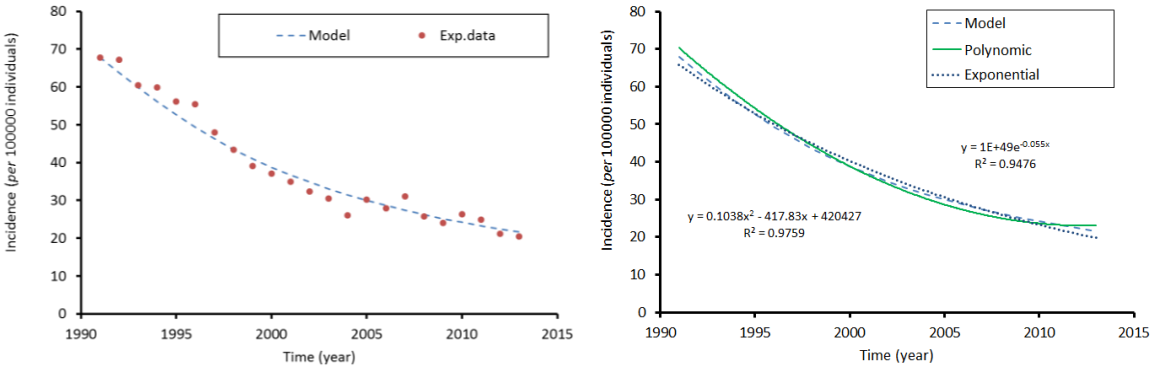


Figure 3.13. The effects of ν_8 on the model, compared with a polynomial adjustment and an exponential adjustment. Left: adjustment of the model considering $\nu_8 = 0.0263$. Right: comparison between the model adjustment and the exponential (dotted, dark blue) and parabolic adjustment (green).

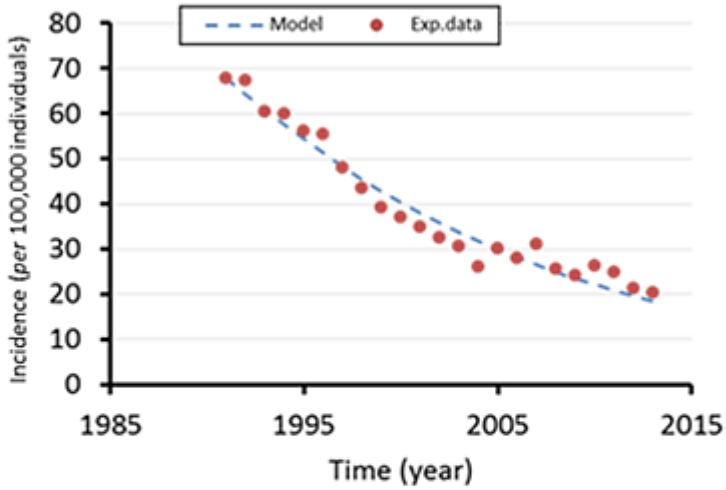
The parabolic behaviour of the system is obtained with lower ν_8 values. The quadratic coefficient (R^2), which determines the correlation between the adjusted function (parabolic, exponential…) and the epidemiological data, is higher in the case of the parabolic function, i.e. a better correlation

between the epidemiological data and the adjusted function is found. In the case of the exponential adjustment R^2 was found to have a value of 0.9476, while it was 0.9759 for the parabolic adjustment

Effects of the initial distribution of the population

Two different evolutions are compared, assuming that in the second case the system has reached a hundred more infected people in the first year (figure 3.14.).

Compartiment	Condicions inicials (nombre d'infectats)
T_i	$E_i(t=0)$
1	500
2	472
3	451
4	436
5	425
6	416
7	409
≥8	10000



1	600
2	472
3	451
4	436
5	425
6	416
7	409
≥8	10000

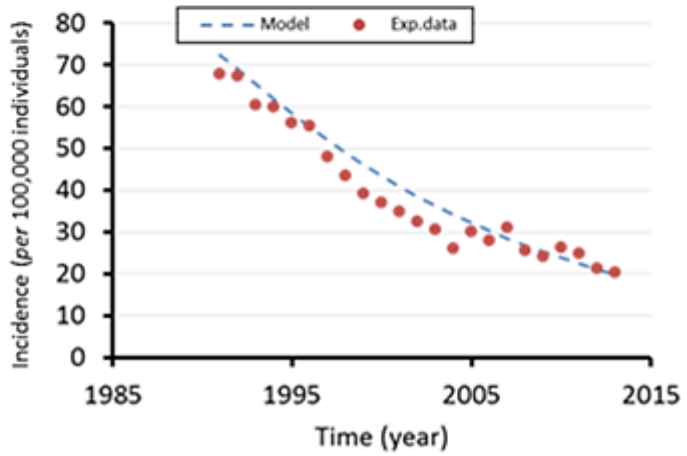


Figure 3.14. Two different evolutions achieved with the model, for TB incidence in Barcelona. Left: initial conditions for both cases. Right: graphic with the epidemiological data and the model behaviour. In the second case (down) the infected individuals in the first year of the initial conditions is increase by 100.

The dynamics of the system varies very substantially, i.e. changes are small, therefore we can say that the system is robust to disturbances. This means for example that the occasional arrival of infected people does not vary the dynamics of the system.

When making major disturbances, note the robustness of the system behaviour. As shown in figure 3.15., the population on compartments 2-4 was removed, while the population in compartments 5-7 was increased. After a transitional period, the system recovers the dynamics of the undisturbed system. Figure 3.16. shows the effect of multiplying by 1.5 the initial population in the 8 compartments.

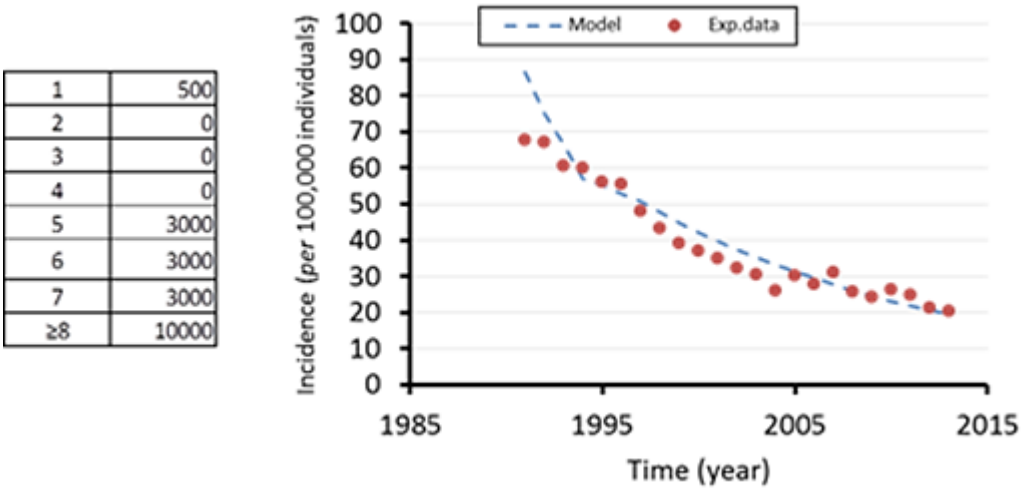


Figure 3.15. Evolution achieved with the model, for TB incidence in Barcelona. The population from compartments 2-4 is 0 while the population of compartments 5-7 is increased. Left: initial conditions for both cases. Right: graphic with the epidemiological data and the model behaviour.

1	750
2	708
3	676.5
4	654
5	637.5
6	624
7	613.5
≥8	15000

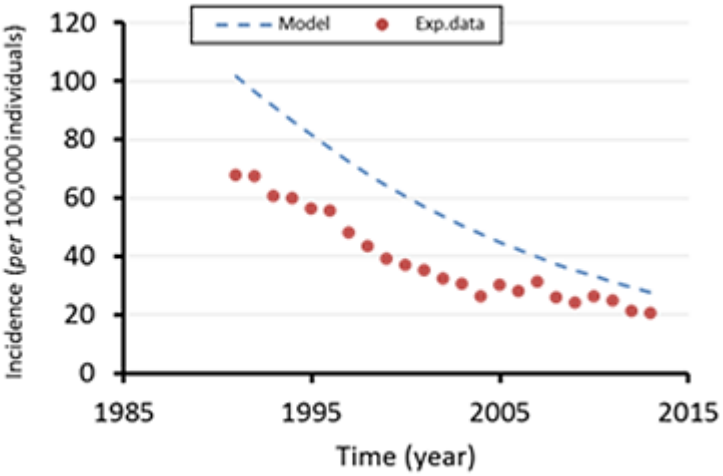


Figure 3.16. Evolution achieved with the model, for TB incidence in Barcelona. The population from all compartments is multiplied by 1.5. Left: initial conditions for both cases. Right: graphic with the epidemiological data and the model behaviour.

Multiplying by 1.5 the initial population of each compartment, it is verified that, effectively, even though the incidence is higher, the dynamic remains the same.

4. Results and discussion

4.1. Results of the model adjustment

The mathematical model we have developed has been positively adjusted to some countries. We have adjusted the model for countries that had clear parabolic decline behaviour. In tables 4.1. and 4.2. the parameters for the adjustments we have conducted are shown.

Table 4.1. Parameters used for the adjustment of several countries and Barcelona (initial distribution, α...). The % of infected a also estimated.

	T_i	Laos	Mauritania	Philippines	Madagascar	Niger	Viet Nam
Initial distribution of the population in the compartments, $E_i(t=0)$	**1**	3505	3200	3215	2838	2600	1833
	2	3590	3024	3056	2680	2552	1794
	3	3429	2891	2944	2598	2535	1715
	4	3313	2878	2866	2619	2448	1657
	5	3227	2801	2813	2548	2377	1614
	6	3165	2824	2729	2498	2330	1581
	7	3115	2814	2790	2458	2292	1600
	≥8	70000	65000	64000	57000	52000	38000
	% infected	93	85	84	75	69	50
	$\alpha_{adjusted}$	6.8667	6.5491	7.3006	7.1504	6.4855	7.1673
	$q\ (\alpha_{adj}-\alpha_{sta})$	-0.7833	-1.1009	-0.3494	-0.4996	-1.1645	-0.4827
	R^2	0.9999	0.9995	0.9998	0.9762	0.9947	0.9979

Table 4.2. Parameters used for the adjustment of several countries and Barcelona (initial distribution, α…). The % of infected is also estimated.

	T_i	Bolivia	China	Barcelona	Syria	Costa Rica
Initial distribution of the population in the compartments, $E_i(t=0)$	**1**	1808	1095	500	440	215
	2	1732	1086	473	425	203
	3	1730	1038	452	407	196
	4	1657	1002	437	366	193
	5	1615	1020	425	374	185
	6	1616	999	417	506	176
	7	1589	983	410	493	172
	≥8	36000	22000	10000	8000	4200
	% infected	48	29	13	11	6
	$\alpha_{adjusted}$	7.0570	6.9915	6.5527	6.6008	6.8700
	$q\ (\alpha_{adj}-\alpha_{sta})$	-0.5930	-0.6585	-1.0973	-1.0492	-0.7800
	R^2	0.9999	0.9998	0.9607	0.9994	0.9974

The parameter q is defined as the difference between α and α_{sta} and it will be discussed later on this chapter. All adjustments have been conducted with many parameters in common (p_i, ν…) as was shown in table 3.1. of the previous chapter. An important remark is that the value for the α parameter for the stationary state was found to be equal for all the countries, since this is the parameter that determines the system dynamics.

The percentage of infected individuals was calculated according to (4.1).

$$\%\ of\ infected = \frac{\sum_{i=1}^{8}(E_{i,\ initial})}{100{,}000\ individuals} \cdot 100 \qquad (4.1)$$

In figure 4.1., the probability of becoming ill depending on the time of infection is showed. The behaviour of this parameter shows a clear parabolic decline that is almost negligible on the eighth

compartment (after 7 or more years of infection). Figures 4.2. to 4.12. show the different adjustments (left) and a prediction of the evolution of TB incidence considering α constant.

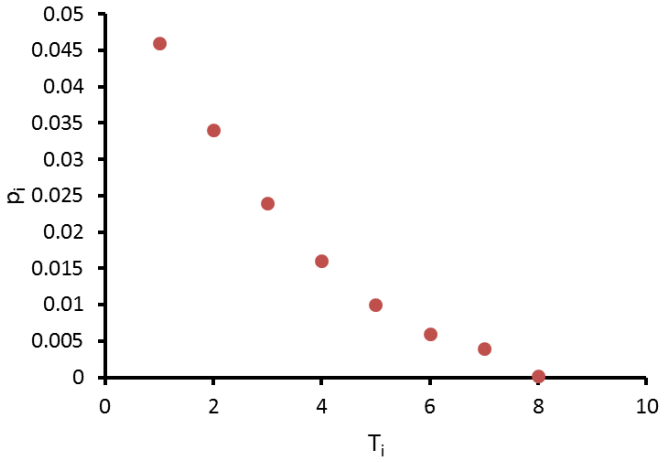


Figure 4.1. The probability of becoming ill (p_i) depending on the time of infection (T_i). This is common for all countries adjusted.

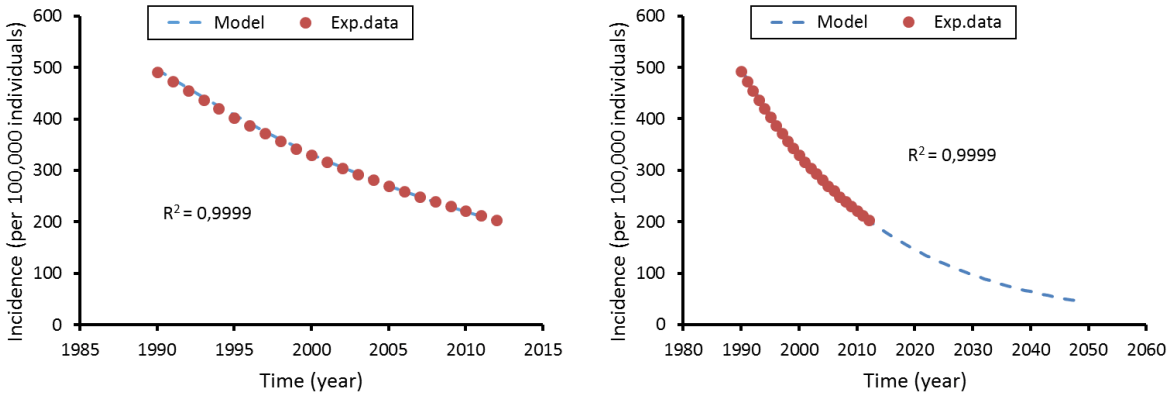


Figure 4.2. Laos. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

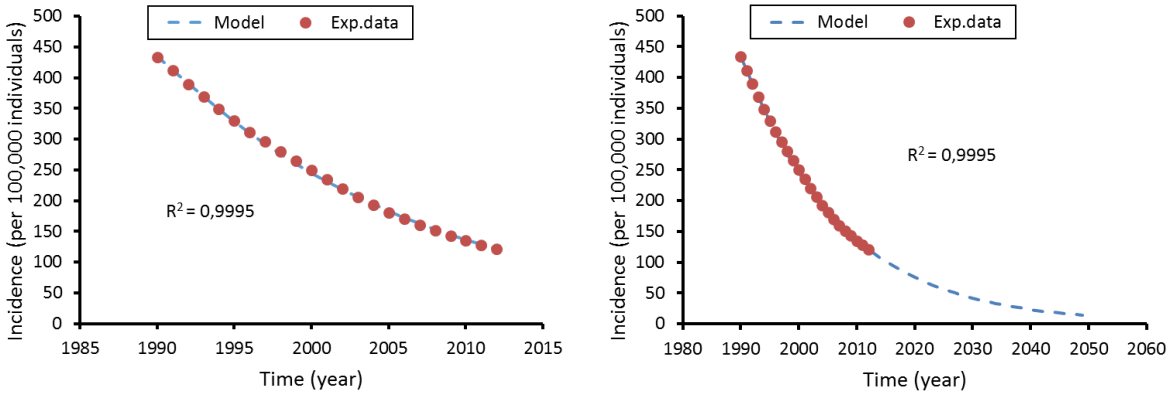


Figure 4.3. Mauritania. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

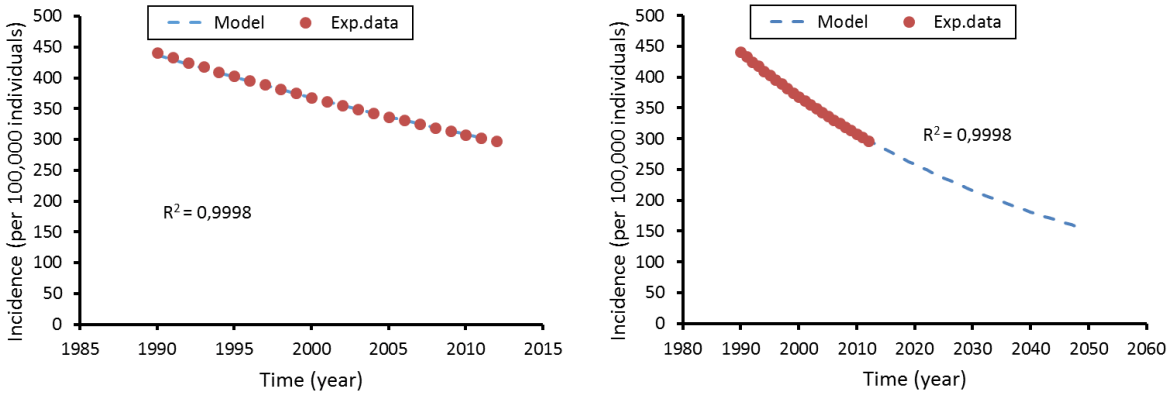


Figure 4.4. Philippines. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

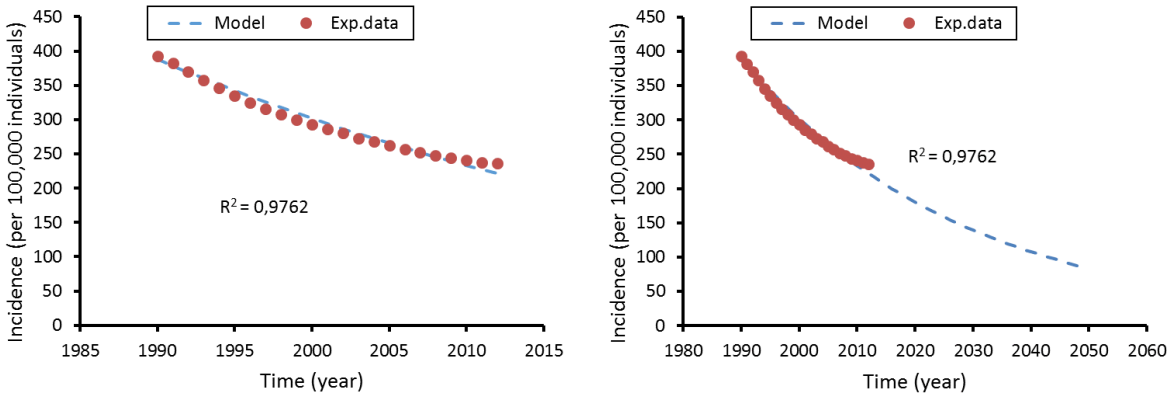


Figure 4.5. Madagascar. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

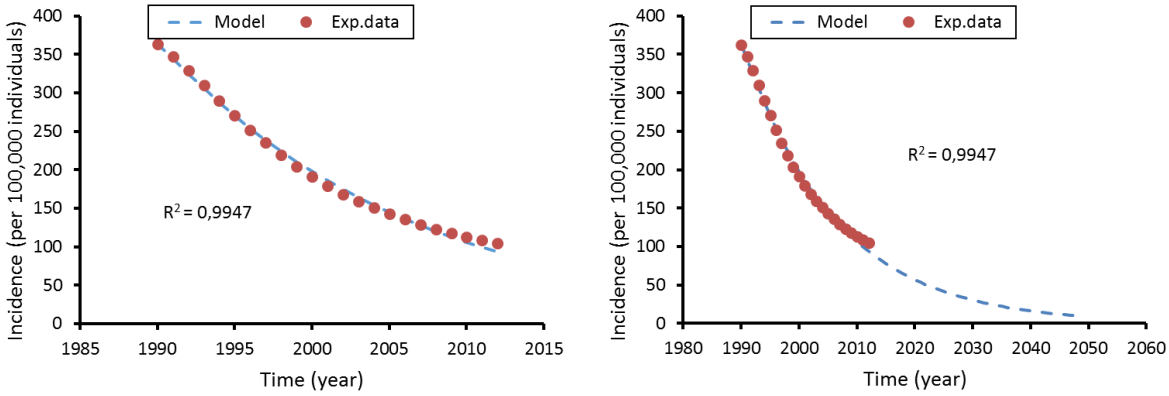


Figure 4.6. Niger. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

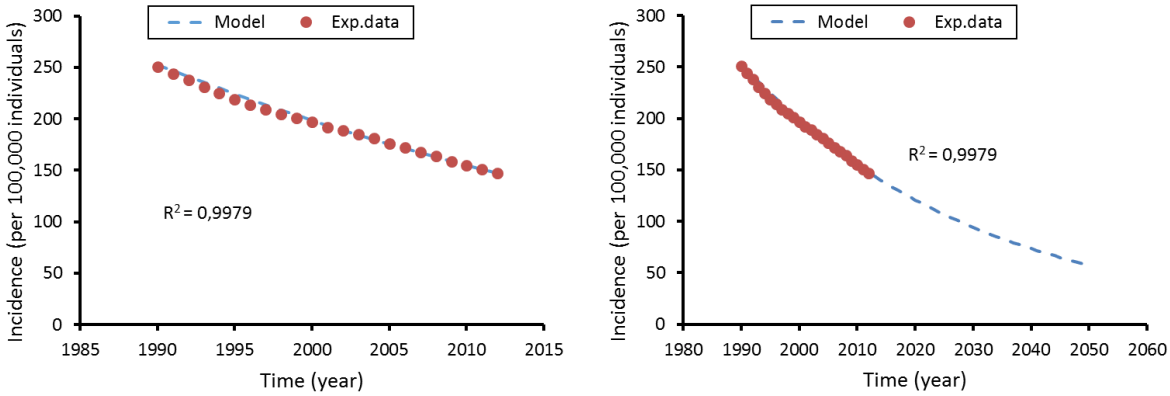


Figure 4.7. Viet Nam. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

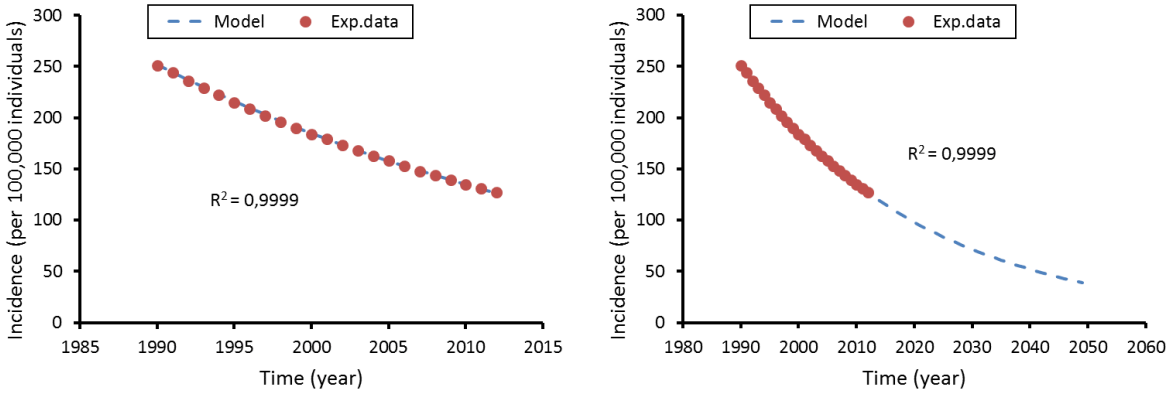


Figure 4.8. Bolivia. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

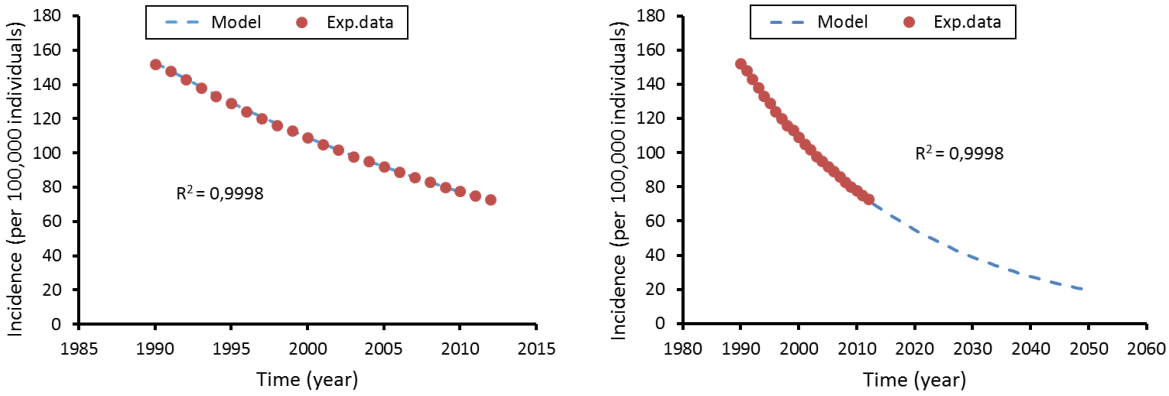


Figure 4.9. China. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

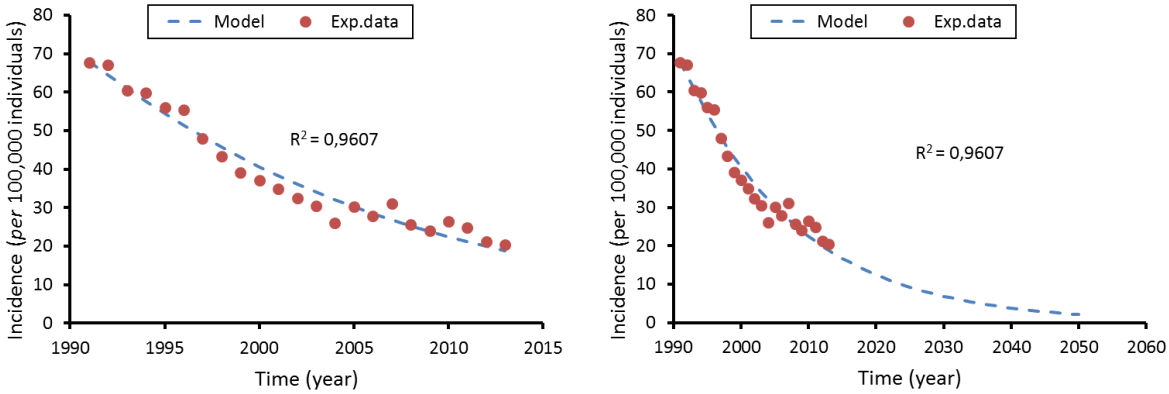


Figure 4.10. Barcelona. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

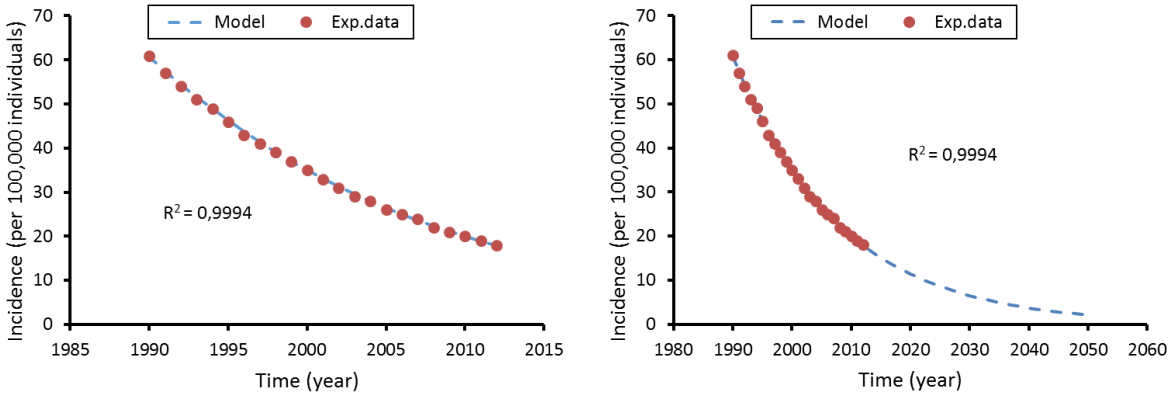


Figure 4.11. Syria. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

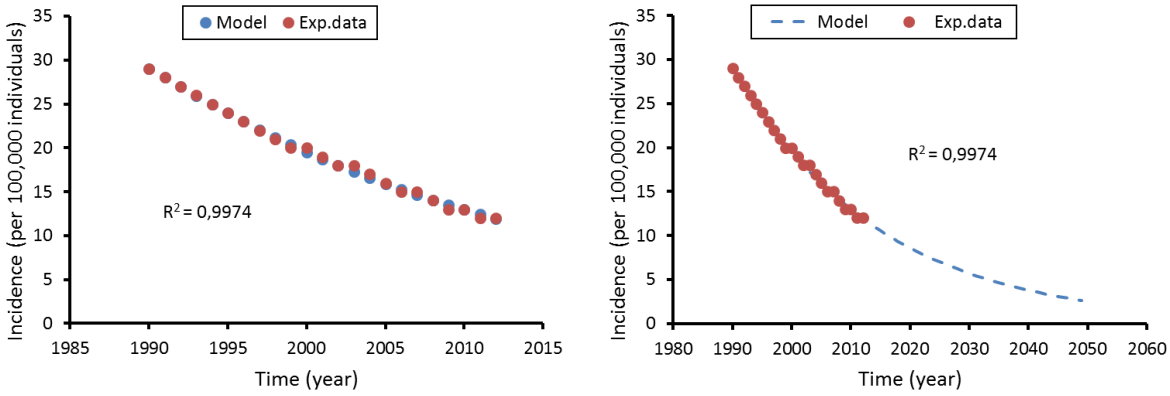


Figure 4.12. Costa Rica. The model adjustment (left) and the prediction of TB incidence evolution until 2050 assuming that α remains constant.

The results show that the model can be adjusted to a wide range of circumstances: countries with a high TB incidence, countries with a low TB incidence, cities…The adjustment is excellent when communities are big (tens of millions of people, or more), but in systems with a smaller dimension, for instance, at city level, some differences (noise) are observed between the adjusted model and the epidemiological dat. Such is the case of Barcelona.

4.2. Discussion

Analysis of the quality of epidemic control programs

Note that the parameter that determines the evolution of the system is:

$$q = \alpha - \alpha_{sta} \tag{4.2}$$

If $q < 0$, the epidemic remains under control, while if $q > 0$, the number of infected increases exponentially.

The defined parameter, q, determines the speed of improvement in epidemic control. Figure 4.13. shows the evolution of TB incidence in the model consider, considering that we decline by 1.0 the value of α ($\alpha = 5.52\,; q = -2.13$) respect the value corresponding to the adjustment in the city of Barcelona ($\alpha = 6.52\,; q = -1.13$). In consequence, the rate of improvement is much higher.

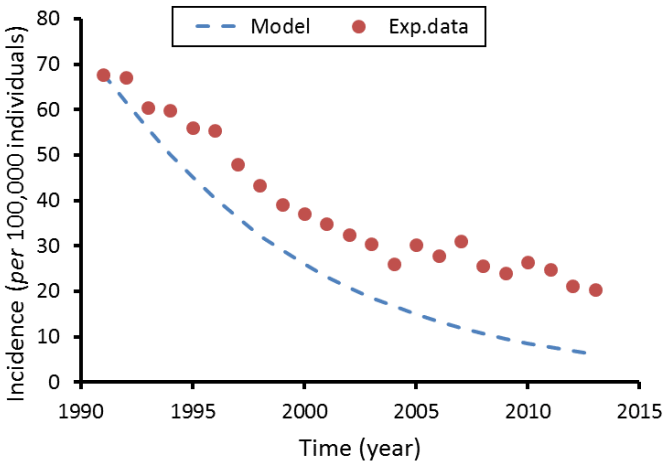


Figure 4.13. The behaviour of the model in Barcelona, taking α=5.52.

Long-term behaviour

Considering a system where there is a good control of the epidemic $(q < 0)$, the model allows to predict the time required to achieve a substantial reduction in the incidence. Taking as a target value an incidence of 4 cases per 100,000 inhabitants. In the case of Barcelona, if conditions remain the same as in the current control policy, the objective is reached in 2039 (figure 4.14).

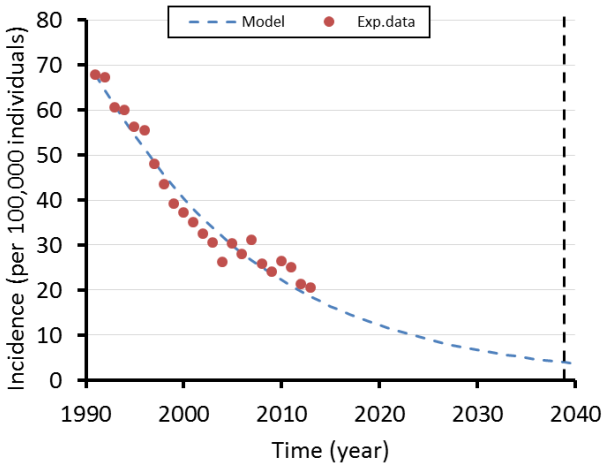


Figure 4.14. Predicted evolution of TB incidence in Barcelona using the mathematical model.

A second index to determine the quality of the control program is the amount of time (t_{obj}) required in order to achieve a certain objective (for example, an incidence of 4/100,000). In the case of Barcelona, it would be:

$$t_{obj} = 25\ years \tag{4.3}$$

Behaviour prediction

The model does not serve to design modifications of the control programs. This goal can be achieved by using models that take into consideration the internal structure of the population, such as Prats *et al.* (2016). However, the model allows to assess the effects of a decrease in α. Figure 4.15. shows what the behaviour of TB in Barcelona would be if, in 2015, had been an improvement in the epidemic control program, assuming $\alpha = 4.54;\ q = -3.11;\ t_{obj} = 9\ years$. In this particular case, the objective would be accomplished by 2023.

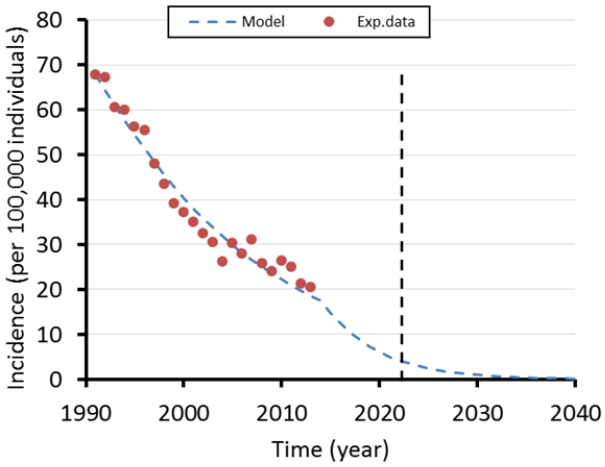


Figure 4.15. Predicted evolution of TB incidence in Barcelona if in 2015, the epidemic control program had improved, assuming $\alpha = 4.54;\ q = -3.11;\ t_{obj} = 9\ years$.

Analysis of more complex situations

If we consider that α is a variable, α (t), that may vary over time we can adjust any epidemic evolution observed. We have not developed an automated adjustment method, method of automatic adjustment, but on figures 4.16. to 4.18. some examples that consider α with few variations are shown to illustrate this concept.

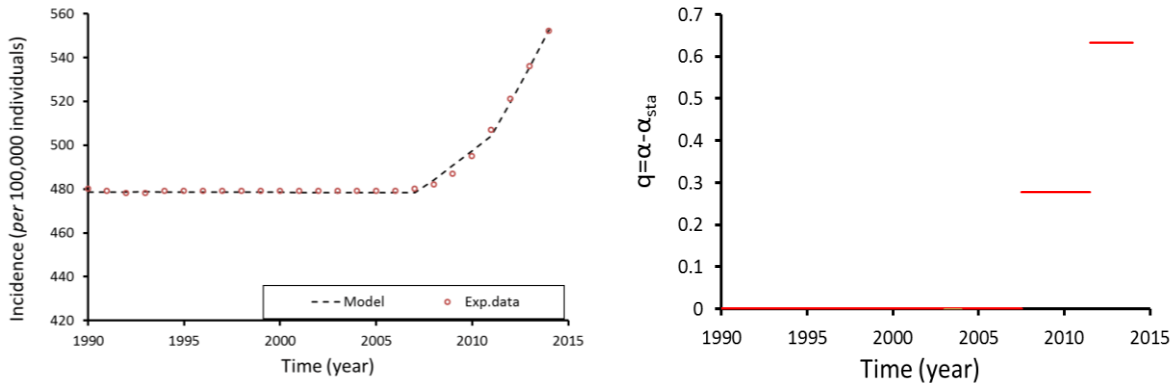


Figure 4.16. TB incidence in North Korea. Left: adjustment with α changing over time. Right: evolution of the *q* parameter over time.

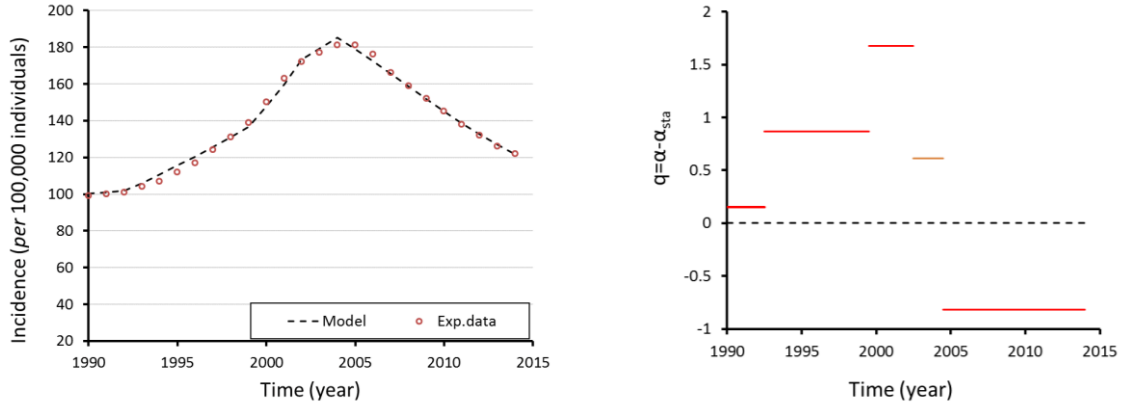


Figure 4.17. TB incidence in Uzbequistan. Left: adjustment with α changing over time. Right: evolution of the *q* parameter over time.

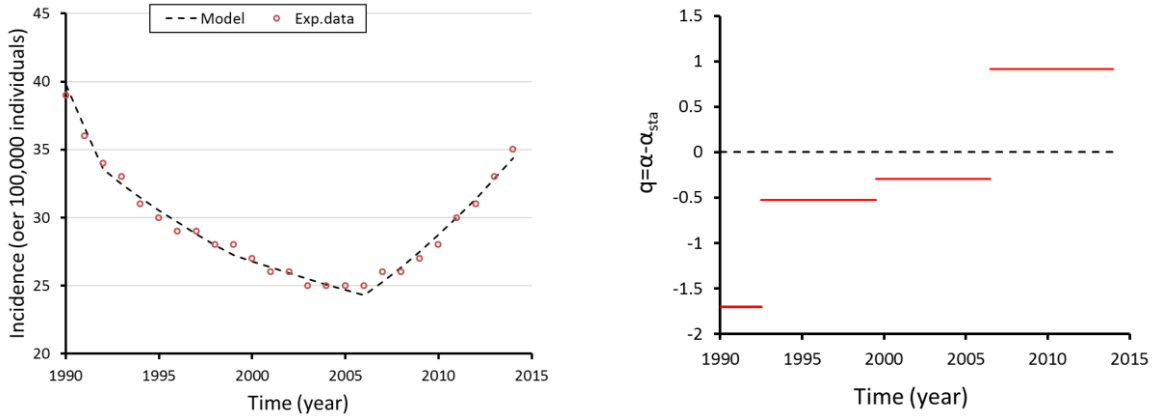


Figure 4.18. TB incidence in Uruguay. Left: adjustment with α changing over time. Right: evolution of the *q* parameter over time.

At the light of the exposed results, we can say that the mathematical model developed allows to make a long-term evaluation of the epidemic control programs.

5. Conclusions

1. The dynamics of tuberculosis incidence has been analysed using data provided by the WHO. It has been found that in countries where there is undergoing a progressive control of the disease, the incidence declines over the years following a parabolic shape. Given the different circumstances of each country (with different social structures, different models of public health, with diversity in the degree of impact), the fact that the temporal epidemiologic dynamic is the same, clearly shows that the dynamic is intrinsic of the epidemiological behaviour of the disease.

2. A compartmentalized mathematical model has been developed considering the population with latent infection depending on the time from initial infection. The model correctly describes the behaviours observed in experimental data. Using the same set of parameters to define the epidemiologic behaviour, and changing the initial conditions and the effective number of infections caused by each sick individual, the model has been adjusted to the behaviour of different countries, thus demonstrating the usefulness of the developed model. We have developed a mathematical model that can describe the dynamics of TB epidemiology and allows to make medium and long-term predictions of the evolution of TB disease under different circumstances. The model uses 10 variables $(E_{1_,} \dots, E_8, I, t)$ and 13 parameters $(\nu_1, \dots, \nu_8, a, b, c, c_8, \alpha)$, and it can be easily implemented in different programming languages. However, we have not developed an automated adjustment for the model, this must be a consideration for a future improvement and continuation of this work.

   In the model, the dynamic of the system is defined by the system structure. The different dynamics observed are very stable. The changes produced are long-term effects, therefore, sporadic alterations do not significantly affect the system The model developed is robust in front of changes in the initial distribution of population but especially sensible to the parameter α (average number of people infected by a single individual), that defines the dynamics of the system. Defining α as a variable depending on the time allows to adjust the model to most of the behaviours observed in epidemiological data. Furthermore, it allows to make a long-term evaluation of the epidemic control programs.

3. The model must allow to predict the behaviour of the disease in the long term, both constantly keeping the situation in each county and improving epidemiological monitoring and control processes.

4. Two indices have been designed to assess the quality of the processes of epidemiological monitoring and control processes. The parameter q, defined as the difference between α and α_{sta}, can be used as an index to evaluate the quality of TB control programs. This parameter will decline as α decreases, indicating that control measures are taken by public health systems. If TB control programs are improved, the average number of people infected by a single individual, α, will decrease, and so will q. In order to ensure a good TB control policy, α must be as small as possible. The second index, t_{obj}, is defined as the amount of time required to achieve a certain objective, i.e. a certain value of the disease incidence.

5. A second index to determine the quality of the control program is the required amount of time in order to achieve a certain objective (for example, an incidence of 4/100,000).

Bibliography

[1] Aparicio, J. P., & Castillo-Chavez, C. (2009). Mathematical modelling of tuberculosis epidemics. *Mathematical Biosciences and Engineering : MBE*, *6*(2), 209–237. http://doi.org/10.3934/mbe.2009.6.209

[2] Begun, M., Newall, A. T., Marks, G. B., & Wood, J. G. (2013). Contact tracing of tuberculosis: a systematic review of transmission modelling studies. *PloS One*, *8*(9), e72470. http://doi.org/10.1371/journal.pone.0072470

[3] Cardona, P. J. (2010). Revisiting the natural history of tuberculosis : tthe inclusion of constant reinfection, host tolerance, and damage-response frameworks leads to a better understanding of latent infection and its evolution towards active disease. *Archivum Immunologiae et Therapiae Experimentalis*, *58*(1), 7–14. http://doi.org/10.1007/s00005-009-0062-5

[4] Cardona, P. J., & Ruiz-Manzano, J. (2004). On the nature of Mycobacterium tuberculosis-latent bacilli. *European Respiratory Journal*, *24*(6), 1044–1051. http://doi.org/10.1183/09031936.04.00072604

[5] Cardona, P.-J., & Prats, C. (2016). The Small Breathing Amplitude at the Upper Lobes Favors the Attraction of Polymorphonuclear Neutrophils to Mycobacterium tuberculosis Lesions and Helps to Understand the Evolution toward Active Disease in An Individual-Based Model. *Frontiers in Microbiology*, *7*, 354. http://doi.org/10.3389/fmicb.2016.00354

[6] Extranet.who.int. (2016). The WHO global TB data collection system. [online] Available at: https://extranet.who.int/tme [Accessed 26 Apr. 2016].

[7] Gong, C., Linderman, J. J., & Kirschner, D. E. (2015). A population model capturing dynamics of tuberculosis granulomas predicts host infection outcomes. Math Biosci Eng 12(3), 625-642

[8] Kasaie, P., Dowdy, D. W., & Kelton, W. D. (2013). An Agent-based simulation of a tuberculosis epidemic: understanding the timing of transmission. In E. R. Pasupathy, et al. (Eds.), Proceedings of the 2013 Winter Simulation Conference (pp. 2227–2238).

[9] Linderman, J. J., Cilfone, N. A. et al. (2015). A multi-scale approach to designing therapeutics for tuberculosis. Integrative Biology, 7(5), 591–609.

[10] MATLAB Documentation, 2016. http://es.mathworks.com [online].

[11] Michel De La Rosa, F. J., Michel De La Rosa, F. J., Bermejo, M. C., Bermejo, M. C., Clavera, I., Clavera, I., … Mar, B. (2007). Epidemiología de la tuberculosis. *An. Sist. Sanit. Navar.*, *30*(2), 7–20. http://doi.org/10.4321/S1137-66272007000400002

[12] Neyrolles, O., & Quintana-Murci, L. (2009). Sexual inequality in tuberculosis. *PLoS Medicine*, *6*(12). http://doi.org/10.1371/journal.pmed.1000199

[13] Orcau i Palau, À., Arcas i Ferré, M., Caylà i Buqueras, J. a., & García de Olalla i Rizo, P. (2015). La tuberculosi a Barcelona. Informe 2013, 65. Retrieved from http://www.aspb.cat/quefem/docs/Tuberculosi_2013.pdf

[14] Perrin, P. (2015). Human and tuberculosis co-evolution: An integrative view. *Tuberculosis*, *95*, S112–S116. http://doi.org/10.1016/j.tube.2015.02.016

[15] Porta, M. (2014). *A Dictionary of Epidemiology*. (M. Hernán, S. Greenland, I. dos Santos Silva, & J. M. Last, Eds.) (6th Editio). New York (USA): Oxford University Press. Retrieved from https://books.google.cl/books?id=3Dr8dyuzvTkC

[16] Prats, C., Montañola-Sales, C., Gilabert-Navarro, J. F., Valls, J., Casanovas-Garcia, J., Vilaplana, C., … López, D. (2016). Individual-based modeling of tuberculosis in a user-friendly interface: Understanding the epidemiological role of population heterogeneity in a city. *Frontiers in Microbiology*, *6*(JAN), 1–13. http://doi.org/10.3389/fmicb.2015.01564

[17] Prats, C., Vilaplana, C., Valls, J., Marzo, E., Cardona, P.-J., & López, D. (2016). Local Inflammation, Dissemination and Coalescence of Lesions Are Key for the Progression toward Active Tuberculosis: The Bubble Model. *Frontiers in Microbiology*, *7*, 33. http://doi.org/10.3389/fmicb.2016.00033

[18] Waaler, H., Geser, A., & Andersen, S. (1962). The use of mathematical models in the study of the epidemiology of tuberculosis. *American Journal of Public Health and the Nations Health*, *52*(6), 1002–1013. Retrieved from http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.52.6.1002

[19] World Health Organization. (2013). Global Strategy and Targets for Tuberculosis Prevention, Care, and Control After 2015. Retrieved from: http://apps.who.int/gb/ebwha/pdf_files/EB134/B134_12-en.pdf?ua=1

[20] World Health Organization. (2016). Global Tuberculosis Report 2015. Retrieved from: http://www.who.int/tb/publications/global_report/en/

[21] Zwerling, A., Shrestha, S., & Dowdy, D. W. (2015). Mathematical Modelling and Tuberculosis: Advances in Diagnostics and Novel Therapies. Advances in Medicine 2015:907267.

Annex: Script for automated adjustment of epidemiological data

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%% ADJUSTMENT FOR WHO DATA V 7.2 %%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%% By Edgar Sánchez Prados %%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%% Supervised by Clara Prats and Daniel López %%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%% Just as a precaution, we close and clear all data from Matlab. %%%%%%
close all
clear all

disp('On calculation process')
warning('off','all'); % This function will avoid the appearence of warnings
% during the data adjustment process, thus increasing the performance speed
C0=fix(clock);% The program will start counting the execution time until the TOC
% function is found.
tic
% We will use the same names used by WHO in order to create the vectors to
% contain the data that will be analysed.

prev_inc=0; % This variable will be used as a switch. It takes much time to
% adjust that many data and sometimes we will only want to adjust incidence
% or prevalence. If this variable is equal to 0, the program will adjust
% all prevalence data. For any other value, it will adjust incidence data.

filename = 'TB_burden_countries_2015-11-13.xlsx'; % Here we will input the
% directory where our file containg the data retrieved from WHO will be
% placed. It is important to previously have some country names reviewed in
% order to avoid strange characters that might not be recognised by Matlab
% thus conducting to a malfunction of the program.

[num,txt] = xlsread(filename); % This function will read all the data in
% the file and store it in two different matrix, num containing all the
% numerical data and txt containing all the text and strings.

Ndades=length(num); % This will read and store the length of num matrix.
txt_lgth=length(txt);
min_numpoints=10; % We set the minimal number of data entries required to
% adjust the second grade equation. In our case we consider 10 entries.
country = txt(2:txt_lgth,1); % Contains the first column of the WHO data,
% the country name on each data entry.
iso3 = txt(2:txt_lgth,3); % Contains the iso3 code for each data entry. It
% can be used in a future version of the script.
g_whoregion = txt(2:txt_lgth,5); % Contains the WHO Region of the data.
year = num(:,3); %Contains the year value for every data entry.
e_pop_num = num(:,4); %Contains the population value for every data entry.
e_prev_100k = num(:,5); %Contains the prevalence (for each 100k population)
% value for every data entry.
e_inc_100k = num(:,25); %Contains the incidence (for each 100k population)
% value for every data entry.
min_pop=2000000; % The minimal value for the mean of a country for
% adjusting the data is set.

country_list = unique(country, 'stable'); % This vector will erase the
% repeated values for the country names in the data. It is important to
% use the 'stable' property in order to keep the same data structure.
% If not used, the information will be put in alphabetical order and since
% it will just change one data structure, mismatch can happen when using
% the rest of the data vectors and matrices, leading to a failure of
```

```
% the program.
iso3_list = unique(iso3, 'stable');
n_paisos=length(country_list); % Stores the number of countries.(219)
% iso3_mat = cell2mat(iso3);
% iso3_list_mat=cell2mat(iso3_list);

[idx,label] = grp2idx(iso3); %Creates index vector from iso3. Starting from
% the first country, it will assign to all the data entry of that same
% country the value '1'. For the next country it will be '2' and so on.
country_data=hist(idx,unique(idx, 'stable')); % Counts the number of times
% that each index is repeated. unique(idx, 'stable') being the number of
% bins (nbins). -> hist(x,xbins)
lgth_ajustos=max(country_data); % Saves the length of the maximum value of
% country_data. It will be used to set the size of various matrices.

% All the matrices with the _2 index on their name, will be used to store
% data from a second adjustment process that will start, for each country
% that undergoes data adjustment, 5 data entries (or points) before the
% last point of the better first adjustment (fin_fit_point).
% The following matrices will store the parameters of every adjustment.
param_ajustos=zeros(n_paisos,6);
param_ajustos_decre=zeros(n_paisos,6);
param_ajustos_decre2=zeros(n_paisos,6);
param_ajustos_cre=zeros(n_paisos,6);
param_ajustos_cre2=zeros(n_paisos,6);
param_ajustos_2=zeros(n_paisos,6);
param_ajustos_decre_2=zeros(n_paisos,6);
param_ajustos_decre2_2=zeros(n_paisos,6);
param_ajustos_cre_2=zeros(n_paisos,6);
param_ajustos_cre2_2=zeros(n_paisos,6);
% The following matrices will store the value of the incidence or
% prevalence for each different adjustment.
ajustos=zeros(n_paisos,lgth_ajustos);
ajustos_decre=zeros(n_paisos,lgth_ajustos);
ajustos_decre2=zeros(n_paisos,lgth_ajustos);
ajustos_cre=zeros(n_paisos,lgth_ajustos);
ajustos_cre2=zeros(n_paisos,lgth_ajustos);
ajustos_2=zeros(n_paisos,lgth_ajustos);
ajustos_decre_2=zeros(n_paisos,lgth_ajustos);
ajustos_decre2_2=zeros(n_paisos,lgth_ajustos);
ajustos_cre_2=zeros(n_paisos,lgth_ajustos);
ajustos_cre2_2=zeros(n_paisos,lgth_ajustos);
% The following matrices will store the value of the derivative for each
% different adjustment.
mat_derivada = zeros(n_paisos,lgth_ajustos);
mat_derivada_decre = zeros(n_paisos,lgth_ajustos);
mat_derivada_decre2 = zeros(n_paisos,lgth_ajustos);
mat_derivada_cre = zeros(n_paisos,lgth_ajustos);
mat_derivada_cre2 = zeros(n_paisos,lgth_ajustos);
mat_derivada_2 = zeros(n_paisos,lgth_ajustos);
mat_derivada_decre_2 = zeros(n_paisos,lgth_ajustos);
mat_derivada_decre2_2 = zeros(n_paisos,lgth_ajustos);
mat_derivada_cre_2 = zeros(n_paisos,lgth_ajustos);
mat_derivada_cre2_2 = zeros(n_paisos,lgth_ajustos);

% The following vectors can be used to manually store the countries
% that we might consider not properly classified by the automated system
% and that we would like to classify them 'manually'. After a first
% execution of this program, we can analyze data and decide which ones
```

```
% should fit better on each kind of behavior. For now, we will just set up
% two vectors that can be used to reclassify the non-clear behavior
% adjustments into either increasing or decreasing ones.
nc_to_decre=zeros(2,1);
nc_to_cre=zeros(2,1);

% The dades_ajustar vector is created to store the data to be adjusted,
% either incidence or prevalence, depending on the value of prev_inc.
% We'll also add the lines for the manual reclassification but as a comment
% since, for now, they are not used.
dades_ajustar = zeros(Ndades,1);
if prev_inc == 0 % If the condition is met, prevalence will be adjusted.
   dades_ajustar(:,1)= e_prev_100k(:,1);
   % Setting of the data and image storage directories.
   str_eix_y=strcat('prevalence');
   dir_ajustos = strcat('ajustos_matlab/prevalence/');
   dir_imatges = strcat('img/prevalence/');
  % After reviewing the adjustments of a first execution, we can manually
  % add countries to be reclassified in a second execution. We need to
  % identify the position of each country on the list (j) and input it in
  % one of the two command lines commented below. Previously, delete* the
  % data and images that might be affected by the reclassification,
  % uncomment the two commands and execute. *Deleting the files might
  % increase the program speed and reduce execution time since no
  % overwritting is required.

  %nc_to_decre=cell2mat({valor1, valor2...});
  %nc_to_cre=cell2mat({valor1, valor2...});

else
   dades_ajustar(:,1)= e_inc_100k(:,1);
   str_eix_y=strcat('incidence');
   dir_ajustos = strcat('ajustos_matlab/incidence/');
   dir_imatges = strcat('img/incidence/');
  % After reviewing the adjustments of a first execution, we can manually
  % add countries to be reclassified in a second execution. We need to
  % identify the position of each country on the list (j) and input it in
  % one of the two command lines commented below. Previously, delete* the
  % data and images that might be affected by the reclassification,
  % uncomment the two commands and execute. *Deleting the files might
  % increase the program speed and reduce execution time since no
  % overwritting is required.

  %nc_to_decre=cell2mat({13, 20, 42, 72, 77, 83, 93, 94, 95, 150, 182,
  % 213,...});
  %nc_to_cre=cell2mat({116, 209});

end

% The upper and lower limits for the X axis, and the lower limit
% for the Y axis are defined.
lim_inf_x=min(year);
lim_sup_x=max(year)+1;
lim_inf_y=0;
% Default plot axes styles
set(0,'DefaultAxesLineWidth',2)
set(0,'DefaultAxesFontSize',20)
set(0,'DefaultLineLineWidth',2)
set(0, 'DefaultAxesTickDir', 'out')
```

```
initial=1; % The initial position, from where the script will start the
% data screening in order to make the different adjustments for the data.
fn=0; % fn will serve as a counter for the number of images/figures.
counter_1_adjustments=0;
counter_2_adjustments=0;
counter_decre_1=0;
counter_decre_2=0;
counter_cre_1=0;
counter_cre_2=0;
counter_nc=1;
counter_decre_1_2=0;
counter_decre_2_2=0;
counter_cre_1_2=0;
counter_cre_2_2=0;
counter_nc_2=1;
  for j=1:n_paisos
      region_list (j,1) = g_whoregion(initial);  % Stores the region code
      % for every data entry.
      ajust_aux=zeros(min_numpoints,1); % This auxiliar matrix will
      % store the fist adjustment. It will be used to plot.
      ajust_aux_2=zeros(min_numpoints,1); % It does the same but
      % for the second adjustment.
      final=initial+country_data(j)-1; %Stablishes the final point of
      % the data that belongs to the 'j' country.
      promig=mean(e_pop_num(initial:final)); % Calculates and stores
      % the mean of population for the 'j' country.

    if promig >= min_pop && (final-initial+1) >= min_numpoints
      counter_1_adjustments=counter_1_adjustments+1;
      % The program checks if the mean value of the population for
      % the 'j' country is equal or higher than the threshold
      % previously established and if also has the minimal number of
      % points (equal or higher). If neither of both conditions is
      % met, the program updates the value of 'initial' as the start
      % point of the next country and takes jumpts back to line 169
      % to continue with the iteration, assuming the next value for
      % 'j'.

      %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
      %%%%%%%%%%%%%%%%%%%% FIRST ADJUSTMENT %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
      %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
          max_numpoints=final-initial+1;
          n_fitting=final-initial-min_numpoints+2;
          % The first adjustment is calculated and its data is
          % stored before entering the fitting loop. Once inside
          % the loop, each adjustment will be comparaed with
          % this previous adjustment and if the value of rsquare
          % is higher on the new adjustment, it will be stored,
          % replacing the previous value for the 'j' country.

          ini=initial; % Temporal variable to store the starting
          % point for the adjustment.
          fin=ini+min_numpoints-1; % Temporal variable to store
          % the final point for the adjustment.
          xvalues=year(ini:fin); % Stores the year values to be
          % used for the adjustment.
          yvalues=dades_ajustar(ini:fin); % Stores the prevalence
          % or incidence data to be used for the adjustment.
```

```
    [coef_fitting, gof]=fit(xvalues,yvalues,'poly2');
    % This function is used calculate the parameters of the
    % second grade equation that better fits the data
    % specified by 'xvalues' and 'yvalues'. The adjusted
    % parameters are stored in coef_fitting and the
    % goodness of fit parameters are stored in gof.
    fitted_parameters=coeffvalues(coef_fitting); % Extracts
    % the fit parameters (a, b and c) and stores them into
    % a vector. Then, they are assigned into the
    % corresponding variable.
    a=fitted_parameters(1);
    b=fitted_parameters(2);
    c=fitted_parameters(3);
    r2=gof.rsquare;
    % ini_fit_point and fin_fit_point are used to store the
    % value of the initial and final point of the
    % adjustment so that it can be used later and exported
    % with other relevant data into the Excel spreadsheets.
    ini_fit_point=ini;
    fin_fit_point=fin;

  for numpoints=min_numpoints:max_numpoints
    % This loop will increase the number of points for the
    % adjustment from the minimal value to the maximal
    % value of the 'j' country. This will allow to do
    % multiple adjustments while screening all data points
    % for different adjustment lengths.
      ini=initial;
      fin=ini+numpoints-1;

    if fin <= final
    % The program checks if the temporal value for the
    % final data is lesser or equal than the real final
    % point. If contrary, data adjustment will be nonsense
    % and it will not be adjusted.
          xvalues=year(ini:fin);
          yvalues=dades_ajustar(ini:fin);

          [coef_fitting, gof]=fit(xvalues,yvalues,'poly2');
          fitted_parameters=coeffvalues(coef_fitting);

          a_aux=fitted_parameters(1);
          b_aux=fitted_parameters(2);
          c_aux=fitted_parameters(3);
          r2_aux=gof.rsquare;


        if r2_aux > r2
          % Only when the rsquare of the current
          % adjustment is higher (better fitting) than a
          % previous value, the adjustment parameters are
          % stored on the corresponding variables.
            a=fitted_parameters(1);
            b=fitted_parameters(2);
            c=fitted_parameters(3);
            r2=gof.rsquare;
            ini_fit_point=ini;
            fin_fit_point=fin;
```

```
            end
            % The first adjustment with the minimal number of
            % points is donce two times. After that, all
            % adjustments will only be done one time each.
            % The program needed to adjust the first country
            % data twice in order to be able to do a comparison
            % This will happen for each country the data of
            % which is adjusted.
            for k=2:n_fitting
              % In this loop, the program will fit the data
              % starting from the data point next and ending
              % at the number of points determined by the
              % status of the mother loop which counts the
              % the number of points starting from the
              % minimal number of points and ending at the
              % maximum number of points for each country.
              % (see line 231)
                ini=ini+1;
                fin=ini+numpoints-1;
              % This procedure is identical as with the first
              % adjustment.
              if fin <= final
                  xvalues=year(ini:fin);
                  yvalues=dades_ajustar(ini:fin);

                  [coef_fitting, gof]=fit(xvalues,yvalues,'poly2');
                  fitted_parameters=coeffvalues(coef_fitting);

                  a_aux=fitted_parameters(1);
                  b_aux=fitted_parameters(2);
                  c_aux=fitted_parameters(3);
                  r2_aux=gof.rsquare;

                if r2_aux > r2
                    a=fitted_parameters(1);
                    b=fitted_parameters(2);
                    c=fitted_parameters(3);
                    r2=gof.rsquare;
                    ini_fit_point=ini;
                    fin_fit_point=fin;
                end
              end
            end
          end
        end
      %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
      %%%%%%%%% CLASSIFICATION OF THE FIRST ADJUSTMENT %%%%%%%%%%
      %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% We need to classify the best adjustment for each country in one of three
% different categories: increasing, decreasing and not_clear. The ones
% with a growing (increasing) behavior can be classified into two different
% types dependingon the shape of the slope. In order to proceed to the
% classification, we need to calculate the first and second grade
% derivatives. We will also create two counters that will count, one the
% number of positive values and the other the negative values of the
% derivative in the range of the best adjustment for the current country
% ('j'). This information will serve as a bypass to save the image into a
% determined folder, as classified, and to save the data in a vector or
% matrix that will allow a later data export to different Excel Worksheets.
```

```
            % Calculation of the first derivative
                derivada=zeros((fin_fit_point-ini_fit_point+1),1);
                l=1;
                for kk=ini_fit_point:fin_fit_point
                  derivada(l)=2*a*year(kk)+b;
                  l=l+1;
                end
                segona_derivada=2*a;
            % The program now proceeds to count the positive and
            % negative derivatives and store the numbers in two
            % different variables.
                comptador_decre_der=sum(derivada < 0);
                comptador_cre_der=sum(derivada >= 0);
                % The values of the adjustment are stored in an
                %auxiliar vector.
                    jj=1;
                    for ii=ini_fit_point:fin_fit_point
                        ajust_aux(jj)=a*year(ii)^2+b*year(ii)+c;
                        jj=jj+1;
                    end

                % The program checks if the value of 'j'
                % belongs to a country that needs to be
                % manually reclassified. This will only work
                % after we input the 'j' value of desired
                % countries after analyzing and surveying the
                % data manually.
                con_nc_decre=0; % For precaution, the variable
                % of this condition is set to 0 before checking
                for r=1:(length(nc_to_decre))
                  if j == nc_to_decre(r)
                    con_nc_decre=1;
                  end
                end
                % The same process, with the other category.
                con_nc_cre=0;
                for r=1:(length(nc_to_cre))
                  if j == nc_to_cre(r)
                    con_nc_cre=1;
                  end
                end

% criteri_class is used to store the number of points used for the fit.
% We will consider that only the countries with all the derivative
% positive or negative an be automatically classified.
                criteri_class=(fin_fit_point-ini_fit_point+1);
               % First,the program checks if the adjustment
               % corresponds to a decreasing type-1 behavior.
                if comptador_decre_der == criteri_class && segona_derivada > 0
                  counter_decre_1=counter_decre_1+1;
                  % DECREASING TYPE-1
                % If the conditions are met, it then stores the
                % adjustment data in the corresponding position
                % of the matrix.
                    param_ajustos_decre(j,1)=a;
                    param_ajustos_decre(j,2)=b;
                    param_ajustos_decre(j,3)=c;
                    param_ajustos_decre(j,4)=r2;
```

```
                param_ajustos_decre(j,5)=1;
                param_ajustos_decre(j,6)=dades_ajustar(initial);
            % In this small loop, the program finds
            % the values of each derivative of the fit.
                for m=1:length(derivada)
                    mat_derivada_decre(j,m)= derivada(m);
                end
                % It now stores the data into the
                % corresponding matrix.
                    for n=1:(fin_fit_point-ini_fit_point+1)
                        ajustos_decre(j,n)= ajust_aux(n);
                    end

                    fn=fn+1; % Updates the image index.
                    max_dades=max(dades_ajustar(initial:final));
                    if max_dades < 1000
                        lim_sup_y=roundn(max_dades+10,1);
                    else
                        lim_sup_y=roundn(max_dades+50,2);
                    end
                    figure
                    hold on
                    axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                    plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                    plot(year(ini_fit_point:fin_fit_point),ajust_aux(:), 'b--')
%                       title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                    xlabel('Time (Year)', 'FontSize', 22)
                    ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                    legend('Epi. data', 'Adjustment')
                    set(fn,'visible','off')
                    saveas(fn,[dir_imatges, 'decreasing/tipus_1/', char(country_list(j)),'.png'])
                    saveas(fn,[dir_imatges, 'decreasing/tipus_1/', char(country_list(j)),'.fig'])
                    hold off
            else
                % If previous conditions are not met, the
                % program checks if the criteria for
                % decreasing type-2 are met.
                if comptador_decre_der == criteri_class && segona_derivada < 0
                    counter_decre_2=counter_decre_2+1;
                    % DECREASING TYPE-2
                    param_ajustos_decre2(j,1)=a;
                    param_ajustos_decre2(j,2)=b;
                    param_ajustos_decre2(j,3)=c;
                    param_ajustos_decre2(j,4)=r2;
                    param_ajustos_decre2(j,5)=2;
                    param_ajustos_decre2(j,6)=dades_ajustar(initial);

                    for m=1:length(derivada)
                        mat_derivada_decre2(j,m)= derivada(m);
                    end

                        for n=1:(fin_fit_point-ini_fit_point+1)
                            ajustos_decre2(j,n)= ajust_aux(n);
                        end

                        fn=fn+1;
                        max_dades=max(dades_ajustar(initial:final));
                        if max_dades < 1000
                            lim_sup_y=roundn(max_dades+10,1);
```

```
                else
                    lim_sup_y=roundn(max_dades+50,2);
                end
                figure
                hold on
                axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                plot(year(ini_fit_point:fin_fit_point),ajust_aux(:), 'b--')
%                   title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                xlabel('Time (Year)', 'FontSize', 22)
                ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                legend('Epi. data', 'Adjustment')
                set(fn,'visible','off')
                saveas(fn,[dir_imatges, 'decreasing/tipus_2/', char(country_list(j)),'.png'])
                saveas(fn,[dir_imatges, 'decreasing/tipus_2/', char(country_list(j)),'.fig'])
                hold off
            else
            % If the adjustment doesn't fulfill the
            % previous criteria,the program now checks
            % if it meets the criteria to be classified
            % as Incresing/Growing Type-1.
            if comptador_cre_der == criteri_class && segona_derivada > 0
              counter_cre_1=counter_cre_1+1;
              % INCREASING TYPE-1
              param_ajustos_cre(j,1)=a;
              param_ajustos_cre(j,2)=b;
              param_ajustos_cre(j,3)=c;
              param_ajustos_cre(j,4)=r2;
              param_ajustos_cre(j,5)=1;
              param_ajustos_cre(j,6)=dades_ajustar(initial);

              for m=1:length(derivada)
                mat_derivada_cre(j,m)= derivada(m);
              end

                for n=1:(fin_fit_point-ini_fit_point+1)
                  ajustos_cre(j,n)= ajust_aux(n);
                end

                fn=fn+1;
                max_dades=max(dades_ajustar(initial:final));
                if max_dades < 1000
                  lim_sup_y=roundn(max_dades+10,1);
                else
                    lim_sup_y=roundn(max_dades+50,2);
                end
                figure
                hold on
                axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                plot(year(ini_fit_point:fin_fit_point),ajust_aux(:), 'b--')
%                   title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                xlabel('Time (Year)', 'FontSize', 22)
                ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                legend('Epi. data', 'Adjustment')
                set(fn,'visible','off')
                saveas(fn,[dir_imatges,'increasing/tipus_1/', char(country_list(j)),'.png'])
                saveas(fn,[dir_imatges,'increasing/tipus_1/', char(country_list(j)),'.fig'])
                hold off
```

```
else
   % Checks if it belongs to
   % Incresing/Growing Type-2.
 if comptador_cre_der == criteri_class && segona_derivada < 0
    counter_cre_2=counter_cre_2+1;
   % INCREASING TYPE-2
   param_ajustos_cre2(j,1)=a;
   param_ajustos_cre2(j,2)=b;
   param_ajustos_cre2(j,3)=c;
   param_ajustos_cre2(j,4)=r2;
   param_ajustos_cre2(j,5)=2;
   param_ajustos_cre2(j,6)=dades_ajustar(initial);

   for m=1:length(derivada)
      mat_derivada_cre2(j,m)= derivada(m);
   end

      for n=1:(fin_fit_point-ini_fit_point+1)
         ajustos_cre2(j,n)= ajust_aux(n);
      end

      fn=fn+1;
      max_dades=max(dades_ajustar(initial:final));
      if max_dades < 1000
         lim_sup_y=roundn(max_dades+10,1);
      else
           lim_sup_y=roundn(max_dades+50,2);
      end
      figure
      hold on
      axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
      plot(year(initial:final),dades_ajustar(initial:final), 'ok')
      plot(year(ini_fit_point:fin_fit_point),ajust_aux(:), 'b--')
%        title(['TB ', str_eix_y, ' at ', char(country_list(j))])
      xlabel('Time (Year)', 'FontSize', 22)
      ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
      legend('Epi. data', 'Adjustment')
      set(fn,'visible','off')
      saveas(fn,[dir_imatges,'increasing/tipus_2/', char(country_list(j)),'.png'])
      saveas(fn,[dir_imatges,'increasing/tipus_2/', char(country_list(j)),'.fig'])
      hold off
else
% If it doesn't meet any of the previous
% conditions, it is classified as
% not-clear.
counter_nc=counter_nc+1;
   % NOT CLEAR
   param_ajustos(j,1)=a;
   param_ajustos(j,2)=b;
   param_ajustos(j,3)=c;
   param_ajustos(j,4)=r2;
   param_ajustos(j,5)=1;
   param_ajustos(j,6)=dades_ajustar(initial);

   for m=1:length(derivada)
      mat_derivada(j,m)= derivada(m);
   end

      for n=1:(fin_fit_point-ini_fit_point+1)
```

```
                    ajustos(j,n)= ajust_aux(n);
                end

                fn=fn+1;
                max_dades=max(dades_ajustar(initial:final));
                if max_dades < 1000
                    lim_sup_y=roundn(max_dades+10,1);
                else
                    lim_sup_y=roundn(max_dades+50,2);
                end
                figure
                hold on
                axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                plot(year(ini_fit_point:fin_fit_point),ajust_aux(:), 'b--')
%                   title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                xlabel('Time (Year)', 'FontSize', 22)
                ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                legend('Epi. data', 'Adjustment')
                set(fn,'visible','off')
                saveas(fn,[dir_imatges, 'not_clear/', char(country_list(j)),'.png'])
                saveas(fn,[dir_imatges, 'not_clear/', char(country_list(j)),'.fig'])
                hold off
            end
        end
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%% SECOND ADJUSTMENT %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% It follows the same idea of the first adjustment. The only
% difference is that a subindex (_2) is used, and the starting
% point for adjusting the data is set 5 points before the last
% point of the best first adjustment. If the length avalaible
% to start the screening and adjustment is not long enough, the
% second adjustment does not take place.
ini_2=fin_fit_point-5;
if (final-ini_2+1) >= min_numpoints
    counter_2_adjustments=counter_2_adjustments+1;
    max_numpoints=final-ini_2+1;

    n_fitting=final-ini_2-min_numpoints+2;
    fin_2=ini_2+min_numpoints-1;
    xvalues=year(ini_2:fin_2);
    yvalues=dades_ajustar(ini_2:fin_2);

    [coef_fitting, gof]=fit(xvalues,yvalues,'poly2');
    fitted_parameters=coeffvalues(coef_fitting);

    a_2=fitted_parameters(1);
    b_2=fitted_parameters(2);
    c_2=fitted_parameters(3);
    r2_2=gof.rsquare;
    ini_fit_point_2=ini_2;
    fin_fit_point_2=fin_2;

    for numpoints=min_numpoints:max_numpoints
        ini_2=fin_fit_point-5;
```

```
    fin_2=ini_2+numpoints-1;

  if fin_2 <= final

      xvalues=year(ini_2:fin_2);
      yvalues=dades_ajustar(ini_2:fin_2);

      [coef_fitting, gof]=fit(xvalues,yvalues,'poly2');
      fitted_parameters=coeffvalues(coef_fitting);

      a_aux=fitted_parameters(1);
      b_aux=fitted_parameters(2);
      c_aux=fitted_parameters(3);
      r2_aux=gof.rsquare;

    if r2_aux > r2_2
        a_2=fitted_parameters(1);
        b_2=fitted_parameters(2);
        c_2=fitted_parameters(3);
        r2_2=gof.rsquare;
        ini_fit_point_2=ini_2;
        fin_fit_point_2=fin_2;
    end

    for k=2:n_fitting
        ini_2=ini_2+1;
        fin_2=ini_2+numpoints-1;

      if fin_2 <= final
          xvalues=year(ini_2:fin_2);
          yvalues=dades_ajustar(ini_2:fin_2);

          [coef_fitting, gof]=fit(xvalues,yvalues,'poly2');
          fitted_parameters=coeffvalues(coef_fitting);

          a_aux=fitted_parameters(1);
          b_aux=fitted_parameters(2);
          c_aux=fitted_parameters(3);
          r2_aux=gof.rsquare;

        if r2_aux > r2_2
            a_2=fitted_parameters(1);
            b_2=fitted_parameters(2);
            c_2=fitted_parameters(3);
            r2_2=gof.rsquare;
            ini_fit_point_2=ini_2;
            fin_fit_point_2=fin_2;

        end
      end
    end
  end
end

      derivada_2=zeros((fin_fit_point_2-ini_fit_point_2+1),1);
      l=1;
      for kk=ini_fit_point_2:fin_fit_point_2
        derivada_2(l)=2*a_2*year(kk)+b_2;
        l=l+1;
```

```
end
segona_derivada=2*a_2;
comptador_decre_der=sum(derivada_2 < 0);
comptador_cre_der=sum(derivada_2 >= 0);

    jj=1;
    for ii=ini_fit_point_2:fin_fit_point_2
        ajust_aux_2(jj)=a_2*year(ii)^2+b_2*year(ii)+c_2;
        jj=jj+1;
    end

con_nc_decre=0;
for r=1:(length(nc_to_decre))
  if j == nc_to_decre(r)
    con_nc_decre=1;
  end
end

con_nc_cre=0;
for r=1:(length(nc_to_cre))
  if j == nc_to_cre(r)
    con_nc_cre=1;
  end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%% CLASSIFICATION OF THE 2ND ADJUSTMENT %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
criteri_class=(fin_fit_point_2-ini_fit_point_2+1);

if comptador_decre_der == criteri_class && segona_derivada > 0
  counter_decre_1_2=counter_decre_1_2+1;
  % DECREASING TYPE-1
    param_ajustos_decre_2(j,1)=a_2;
    param_ajustos_decre_2(j,2)=b_2;
    param_ajustos_decre_2(j,3)=c_2;
    param_ajustos_decre_2(j,4)=r2_2;
    param_ajustos_decre_2(j,5)=1;
    param_ajustos_decre_2(j,6)=dades_ajustar(initial);

    for m=1:length(derivada_2)
      mat_derivada_decre_2(j,m)= derivada_2(m);
    end

      for n=1:(fin_fit_point_2-ini_fit_point_2+1)
        ajustos_decre_2(j,n)= ajust_aux_2(n);
      end

      fn=fn+1;
      max_dades=max(dades_ajustar(initial:final));
      if max_dades < 1000
        lim_sup_y=roundn(max_dades+10,1);
      else
        lim_sup_y=roundn(max_dades+50,2);
      end
      figure
      hold on
      axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
      plot(year(initial:final),dades_ajustar(initial:final), 'ok')
```

```
                plot(year(ini_fit_point_2:fin_fit_point_2),ajust_aux_2(:), 'r--')
%                  title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                xlabel('Time (Year)', 'FontSize', 22)
                ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                legend('Epi. data', 'Adjustment')
                set(fn,'visible','off')
                saveas(fn,[dir_imatges, 'decreasing/tipus_1/', char(country_list(j)),'_2.png'])
                saveas(fn,[dir_imatges, 'decreasing/tipus_1/', char(country_list(j)),'_2.fig'])
                hold off
            else
              if comptador_decre_der == criteri_class && segona_derivada > 0
                counter_decre_2_2=counter_decre_2_2+1;
              % DECREASING TYPE-2
                param_ajustos_decre_2(j,1)=a_2;
                param_ajustos_decre_2(j,2)=b_2;
                param_ajustos_decre_2(j,3)=c_2;
                param_ajustos_decre_2(j,4)=r2_2;
                param_ajustos_decre_2(j,5)=2;
                param_ajustos_decre_2(j,6)=dades_ajustar(initial);

                for m=1:length(derivada_2)
                  mat_derivada_decre_2(j,m)= derivada_2(m);
                end

                  for n=1:(fin_fit_point_2-ini_fit_point_2+1)
                    ajustos_decre_2(j,n)= ajust_aux_2(n);
                  end

                  fn=fn+1;
                  max_dades=max(dades_ajustar(initial:final));
                  if max_dades < 1000
                    lim_sup_y=roundn(max_dades+10,1);
                  else
                    lim_sup_y=roundn(max_dades+50,2);
                  end
                  figure
                  hold on
                  axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                  plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                  plot(year(ini_fit_point_2:fin_fit_point_2),ajust_aux_2(:), 'r--')
%                    title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                  xlabel('Time (Year)', 'FontSize', 22)
                  ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                  legend('Epi. data', 'Adjustment')
                  set(fn,'visible','off')
                  saveas(fn,[dir_imatges, 'decreasing/tipus_1/', char(country_list(j)),'_2.png'])
                  saveas(fn,[dir_imatges, 'decreasing/tipus_2/', char(country_list(j)),'_2.fig'])
                  hold off
              else
               if comptador_cre_der == criteri_class && segona_derivada > 0
                 counter_cre_1_2=counter_cre_1_2+1;
                % INCREASING/GROWING TYPE-1
                param_ajustos_cre_2(j,1)=a_2;
                param_ajustos_cre_2(j,2)=b_2;
                param_ajustos_cre_2(j,3)=c_2;
                param_ajustos_cre_2(j,4)=r2_2;
                param_ajustos_cre_2(j,5)=1;
                param_ajustos_cre_2(j,6)=dades_ajustar(initial);
```

```
                for m=1:length(derivada_2)
                    mat_derivada_cre_2(j,m)= derivada_2(m);
                end

                    for n=1:(fin_fit_point_2-ini_fit_point_2+1)
                        ajustos_cre_2(j,n)= ajust_aux_2(n);
                    end

                    fn=fn+1;
                    max_dades=max(dades_ajustar(initial:final));
                    if max_dades < 1000
                        lim_sup_y=roundn(max_dades+10,1);
                    else
                            lim_sup_y=roundn(max_dades+50,2);
                    end
                    figure
                    hold on
                    axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                    plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                    plot(year(ini_fit_point_2:fin_fit_point_2),ajust_aux_2(:), 'r--')
%                       title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                    xlabel('Time (Year)', 'FontSize', 22)
                    ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                    legend('Epi. data', 'Adjustment')
                    set(fn,'visible','off')
                    saveas(fn,[dir_imatges,'increasing/tipus_1/', char(country_list(j)),'_2.png'])
                    saveas(fn,[dir_imatges,'increasing/tipus_1/', char(country_list(j)),'_2.fig'])
                    hold off
                else
                 if comptador_cre_der == criteri_class && segona_derivada < 0
                    counter_cre_2_2=counter_decre_2_2+1;
                   % INCREASING/GROWING TYPE-2
                   param_ajustos_cre2_2(j,1)=a_2;
                   param_ajustos_cre2_2(j,2)=b_2;
                   param_ajustos_cre2_2(j,3)=c_2;
                   param_ajustos_cre2_2(j,4)=r2_2;
                   param_ajustos_cre2_2(j,5)=2;
                   param_ajustos_cre2_2(j,6)=dades_ajustar(initial);

                   for m=1:length(derivada_2)
                       mat_derivada_cre2_2(j,m)= derivada_2(m);
                   end

                       for n=1:(fin_fit_point_2-ini_fit_point_2+1)
                           ajustos_cre2_2(j,n)= ajust_aux_2(n);
                       end

                       fn=fn+1;
                       max_dades=max(dades_ajustar(initial:final));
                       if max_dades < 1000
                           lim_sup_y=roundn(max_dades+10,1);
                       else
                               lim_sup_y=roundn(max_dades+50,2);
                       end
                       figure
                       hold on
                       axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                       plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                       plot(year(ini_fit_point_2:fin_fit_point_2),ajust_aux_2(:), 'r--')
```

```
%                            title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                            xlabel('Time (Year)', 'FontSize', 22)
                            ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                            legend('Epi. data', 'Adjustment')
                            set(fn,'visible','off')
                            saveas(fn,[dir_imatges,'increasing/tipus_2/', char(country_list(j)),'_2.png'])
                            saveas(fn,[dir_imatges,'increasing/tipus_2/', char(country_list(j)),'_2.fig'])
                            hold off
                        else
                         counter_nc_2=counter_nc_2+1;
                         % NOT CLEAR
                         param_ajustos_2(j,1)=a_2;
                         param_ajustos_2(j,2)=b_2;
                         param_ajustos_2(j,3)=c_2;
                         param_ajustos_2(j,4)=r2_2;
                         param_ajustos_2(j,5)=0;
                         param_ajustos_2(j,6)=dades_ajustar(initial);

                         for m=1:length(derivada_2)
                            mat_derivada_2(j,m)= derivada_2(m);
                         end

                            for n=1:(fin_fit_point_2-ini_fit_point_2+1)
                               ajustos_2(j,n)= ajust_aux_2(n);
                            end

                            fn=fn+1;
                            max_dades=max(dades_ajustar(initial:final));
                            if max_dades < 1000
                               lim_sup_y=roundn(max_dades+10,1);
                            else
                               lim_sup_y=roundn(max_dades+50,2);
                            end
                            figure
                            hold on
                            axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                            plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                            plot(year(ini_fit_point_2:fin_fit_point_2),ajust_aux_2(:), 'r--')
%                            title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                            xlabel('Time (Year)', 'FontSize', 22)
                            ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                            legend('Epi. data', 'Adjustment')
                            set(fn,'visible','off')
                            saveas(fn,[dir_imatges, 'not_clear/', char(country_list(j)),'_2.png'])
                            saveas(fn,[dir_imatges, 'not_clear/', char(country_list(j)),'_2.fig'])
                            hold off
                         end
                        end
                       end
                     end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  %%%%%%%%%%%%%% GRAPHIC OF BOTH ADJUSTMENTS %%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  % The program only plots the two adjustments and stores the image if
  % the second adjustment can be completed.
                            fn=fn+1;
```

```
                    figure
                    hold on
                    axis([lim_inf_x lim_sup_x lim_inf_y lim_sup_y])
                    plot(year(initial:final),dades_ajustar(initial:final), 'ok')
                    plot(year(ini_fit_point:fin_fit_point),ajust_aux(:), 'b--')
                    plot(year(ini_fit_point_2:fin_fit_point_2),ajust_aux_2(:), 'r--')
%                       title(['TB ', str_eix_y, ' at ', char(country_list(j))])
                    xlabel('Time (Year)', 'FontSize', 22)
                    ylabel(['TB ', str_eix_y, ' (\itper\rm 100,000 individuals)'], 'FontSize', 22)
                    legend('Epi. data', 'First adjustment', 'Second adjustment')
                    set(fn,'visible','off')
                    saveas(fn,[dir_imatges, 'dos_ajustos/', char(country_list(j)),'.png'])
                    saveas(fn,[dir_imatges, 'dos_ajustos/', char(country_list(j)),'.fig'])
                    hold off
        end
    end
        initial=initial+country_data(j); % The initial value of the
        % data index is updated.
end



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  %%%%%%%%%%%%%%%%%%%% DATA OUTPUT %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

  % In order to export a more undestandable data and reducing the export
  % time to the various Excel Worksheets, the creation and use of filters
  % is required. The filters will allow to clear the rows from every
  % matrix that does not contain any relevant data. Thus, all countries
  % that had not been through any adjustment will be erased.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  %%%%%%%%%%%%%% FILTERS FOR THE FIRST ADJUSTMENT %%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  % Different country, ISO-3 and WHO Region lists are created to store data.
      country_list_decre=country_list;
      country_list_decre2=country_list;
      country_list_cre=country_list;
      country_list_cre2=country_list;
      country_list_nc=country_list;
      iso_list_decre=iso3_list;
      iso_list_decre2=iso3_list;
      iso_list_cre=iso3_list;
      iso_list_cre2=iso3_list;
      iso_list_nc=iso3_list;
      region_list_decre=region_list;
      region_list_decre2=region_list;
      region_list_cre=region_list;
      region_list_cre2=region_list;
      region_list_nc=region_list;

  % The existance of empty data is checked upon the second column of the
  % adjustment data matrices. If it is empty (0), no adjustment has taken
  % place. The filter is applied on all matrices and vectors.
      F_decre=ajustos_decre(:,2) == 0;
      ajustos_decre(F_decre,:)=[];
```

```
param_ajustos_decre(F_decre,:)=[];
mat_derivada_decre(F_decre,:)=[];
country_list_decre(F_decre,:)=[];
iso_list_decre(F_decre,:)=[];
region_list_decre(F_decre,:)=[];
F_decre2=ajustos_decre2(:,2) == 0;
ajustos_decre2(F_decre2,:)=[];
param_ajustos_decre2(F_decre2,:)=[];
mat_derivada_decre2(F_decre2,:)=[];
country_list_decre2(F_decre2,:)=[];
iso_list_decre2(F_decre2,:)=[];
region_list_decre2(F_decre2,:)=[];
F_cre=ajustos_cre(:,2) == 0;
ajustos_cre(F_cre,:)=[];
param_ajustos_cre(F_cre,:)=[];
mat_derivada_cre(F_cre,:)=[];
country_list_cre(F_cre,:)=[];
iso_list_cre(F_cre,:)=[];
region_list_cre(F_cre,:)=[];
F_cre2=ajustos_cre2(:,2) == 0;
ajustos_cre2(F_cre2,:)=[];
param_ajustos_cre2(F_cre2,:)=[];
mat_derivada_cre2(F_cre2,:)=[];
country_list_cre2(F_cre2,:)=[];
iso_list_cre2(F_cre2,:)=[];
region_list_cre2(F_cre2,:)=[];
F_nc=ajustos(:,2) == 0;
ajustos(F_nc,:)=[];
param_ajustos(F_nc,:)=[];
mat_derivada(F_nc,:)=[];
country_list_nc(F_nc,:)=[];
iso_list_nc(F_nc,:)=[];
region_list_nc(F_nc,:)=[];


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%% FILTERS FOR THE SECOND ADJUSTMENT %%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
country_list_decre_2=country_list;
country_list_decre2_2=country_list;
country_list_cre_2=country_list;
country_list_cre2_2=country_list;
country_list_nc_2=country_list;
iso_list_decre_2=iso3_list;
iso_list_decre2_2=iso3_list;
iso_list_cre_2=iso3_list;
iso_list_cre2_2=iso3_list;
iso_list_nc_2=iso3_list;
region_list_decre_2=region_list;
region_list_decre2_2=region_list;
region_list_cre_2=region_list;
region_list_cre2_2=region_list;
region_list_nc_2=region_list;


F_decre_2=ajustos_decre_2(:,2) == 0;
ajustos_decre_2(F_decre_2,:)=[];
param_ajustos_decre_2(F_decre_2,:)=[];
```

```
        mat_derivada_decre_2(F_decre_2,:)=[];
        country_list_decre_2(F_decre_2,:)=[];
        iso_list_decre_2(F_decre_2,:)=[];
        region_list_decre_2(F_decre_2,:)=[];
        F_decre2_2=ajustos_decre2_2(:,2) == 0;
        ajustos_decre2_2(F_decre2_2,:)=[];
        param_ajustos_decre2_2(F_decre2_2,:)=[];
        mat_derivada_decre2_2(F_decre2_2,:)=[];
        country_list_decre2_2(F_decre2_2,:)=[];
        iso_list_decre2_2(F_decre2_2,:)=[];
        region_list_decre2_2(F_decre2_2,:)=[];
        F_cre_2=ajustos_cre_2(:,2) == 0;
        ajustos_cre_2(F_cre_2,:)=[];
        param_ajustos_cre_2(F_cre_2,:)=[];
        mat_derivada_cre_2(F_cre_2,:)=[];
        country_list_cre_2(F_cre_2,:)=[];
        iso_list_cre_2(F_cre_2,:)=[];
        region_list_cre_2(F_cre_2,:)=[];
        F_cre2_2=ajustos_cre2_2(:,2) == 0;
        ajustos_cre2_2(F_cre2_2,:)=[];
        param_ajustos_cre2_2(F_cre2_2,:)=[];
        mat_derivada_cre2_2(F_cre2_2,:)=[];
        country_list_cre2_2(F_cre2_2,:)=[];
        iso_list_cre2_2(F_cre2_2,:)=[];
        region_list_cre2_2(F_cre2_2,:)=[];
        F_nc_2=ajustos_2(:,2) == 0;
        ajustos_2(F_nc_2,:)=[];
        param_ajustos_2(F_nc_2,:)=[];
        mat_derivada_2(F_nc_2,:)=[];
        country_list_nc_2(F_nc_2,:)=[];
        iso_list_nc_2(F_nc_2,:)=[];
        region_list_nc_2(F_nc_2,:)=[];


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  %%%%%%%%%%%%%%%%%%%% VECTOR AND MATRIX MERGING %%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  % Since the program works with two different adjustments and stores
  % them in different data structures, those structures must be merged in
  % order that the Excel containing the data from the parameters of the
  % increasing type-1, for instance, contains the data of both best first
  % and second adjustment. On the other hand, if there is a need to have
  % one different file for each type of behaviour (decreasing tyoe-1 or
  % type-2, increasingtype-1 or type-2, or not clear) and adjustment
  % (first or second, the data must be exported before merging the data
  % structures.

  parametres_ajustos_nc=vertcat(param_ajustos,param_ajustos_2);
  parametres_ajustos_decre=vertcat(param_ajustos_decre,param_ajustos_decre_2);
  parametres_ajustos_decre2=vertcat(param_ajustos_decre2,param_ajustos_decre2_2);
  parametres_ajustos_cre=vertcat(param_ajustos_cre,param_ajustos_cre_2);
  parametres_ajustos_cre2=vertcat(param_ajustos_cre2,param_ajustos_cre2_2);
  mat_derivades_nc=vertcat(mat_derivada,mat_derivada_2);
  mat_derivades_decre=vertcat(mat_derivada_decre,mat_derivada_decre_2);
  mat_derivades_decre2=vertcat(mat_derivada_decre2,mat_derivada_decre2_2);
  mat_derivades_cre=vertcat(mat_derivada_cre,mat_derivada_cre_2);
  mat_derivades_cre2=vertcat(mat_derivada_cre2,mat_derivada_cre2_2);
  m_ajustos_nc=vertcat(ajustos,ajustos_2);
```

```
m_ajustos_decre=vertcat(ajustos_decre,ajustos_decre_2);
m_ajustos_decre2=vertcat(ajustos_decre2,ajustos_decre2_2);
m_ajustos_cre=vertcat(ajustos_cre,ajustos_cre_2);
m_ajustos_cre2=vertcat(ajustos_cre2,ajustos_cre2_2);

llista_nc=vertcat(country_list_nc,country_list_nc_2);
llista_decre=vertcat(country_list_decre,country_list_decre_2);
llista_decre2=vertcat(country_list_decre2,country_list_decre2_2);
llista_cre=vertcat(country_list_cre,country_list_cre_2);
llista_cre2=vertcat(country_list_cre2,country_list_cre2_2);
llista_iso_nc=vertcat(iso_list_nc,iso_list_nc_2);
llista_iso_decre=vertcat(iso_list_decre,iso_list_decre_2);
llista_iso_decre2=vertcat(iso_list_decre2,iso_list_decre2_2);
llista_iso_cre=vertcat(iso_list_cre,iso_list_cre_2);
llista_iso_cre2=vertcat(iso_list_cre2,iso_list_cre2_2);
llista_region_nc=vertcat(region_list_nc,region_list_nc_2);
llista_region_decre=vertcat(region_list_decre,region_list_decre_2);
llista_region_decre2=vertcat(region_list_decre2,region_list_decre2_2);
llista_region_cre=vertcat(region_list_cre,region_list_cre_2);
llista_region_cre2=vertcat(region_list_cre2,region_list_cre2_2);



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%% MERGED DATA OUTPUT %%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    Nsortida_nc=length(llista_nc)+1;
    Nsortida_decre=length(llista_decre)+1;
    Nsortida_decre2=length(llista_decre2)+1;
    Nsortida_cre=length(llista_cre)+1;
    Nsortida_cre2=length(llista_cre2)+1;

% NOT CLEAR
    xlswrite([dir_ajustos, 'param_ajustos_nc', '.xlsx'],{'Country'},'A1:A1')
    xlswrite([dir_ajustos, 'param_ajustos_nc', '.xlsx'],{'ISO3'},'B1:B1')
    xlswrite([dir_ajustos, 'param_ajustos_nc', '.xlsx'],{'WHO Region'},'C1:C1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],{'a coeff.'},'D1:F1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],{'b coeff.'},'E1:E1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],{'c coeff.'},'F1:F1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],{'r2'},'G1:G1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],{'Tipus ajust'},'H1:H1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],{['Initial ', str_eix_y]},'I1:I1')
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],llista_nc,['A2:A' num2str(Nsortida_nc)])
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],llista_iso_nc,['B2:B' num2str(Nsortida_nc)])
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],llista_region_nc,['C2:C' num2str(Nsortida_nc)])
    xlswrite([dir_ajustos, 'param_ajustos_nc','.xlsx'],parametres_ajustos_nc,['D2:I' num2str(Nsortida_nc)])

    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],{'Country'},'A1:A1')
    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],{'ISO3'},'B1:B1')
    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],{'WHO Region'},'C1:C1')
    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],llista_nc,['A2:A' num2str(Nsortida_nc)])
    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],llista_iso_nc,['B2:B' num2str(Nsortida_nc)])
    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],llista_region_nc,['C2:C' num2str(Nsortida_nc)])
    xlswrite([dir_ajustos, 'mat_der_nc','.xlsx'],mat_derivades_nc,['D2:AB' num2str(Nsortida_nc)])

    xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],{'Country'},'A1:A1')
    xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],{'ISO3'},'B1:B1')
    xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],{'WHO Region'},'C1:C1')
```

```
        xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],llista_nc,['A2:A' num2str(Nsortida_nc)])
        xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],llista_iso_nc,['B2:B' num2str(Nsortida_nc)])
        xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],llista_region_nc,['C1:C' num2str(Nsortida_nc)])
        xlswrite([dir_ajustos, 'ajustos_nc','.xlsx'],m_ajustos_nc,['D2:AB' num2str(Nsortida_nc)])

% DECREASING TYPE-1
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'a coeff.'},'D1:D1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'b coeff.'},'E1:E1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'c coeff.'},'F1:F1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'r2'},'G1:G1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{'Tipus ajust'},'H1:H1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],{['Initial ', str_eix_y]},'I1:I1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],llista_decre,['A2:A' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],llista_iso_decre,['B2:B'
num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],llista_region_decre,['C2:C'
num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_1','.xlsx'],parametres_ajustos_decre,['D2:I'
num2str(Nsortida_decre)])

        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],llista_decre,['A2:A' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],llista_iso_decre,['B2:B' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],llista_region_decre,['C2:C' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_1','.xlsx'],mat_derivades_decre,['D2:AB'
num2str(Nsortida_decre)])

        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],llista_decre,['A2:A' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],llista_iso_decre,['B2:B' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],llista_region_decre,['C2:C' num2str(Nsortida_decre)])
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_1','.xlsx'],m_ajustos_decre,['D2:AB' num2str(Nsortida_decre)])

% DECREASING TYPE-2
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'a coeff.'},'D1:D1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'b coeff.'},'E1:E1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'c coeff.'},'F1:F1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'r2'},'G1:G1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{'Tipus ajust'},'H1:H1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],{['Initial ', str_eix_y]},'I1:I1')
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],llista_decre2,['A2:A' num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],llista_iso_decre2,['B2:B'
num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],llista_region_decre2,['C2:C'
num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'param_ajustos_decre_tipus_2','.xlsx'],parametres_ajustos_decre2,['D2:I'
num2str(Nsortida_decre2)])

        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],{'Country'},'A1:A1')
```

```
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],llista_decre2,['A2:A' num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],llista_iso_decre2,['B2:B' num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],llista_region_decre2,['C2:C'
num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'mat_der_decre_tipus_2','.xlsx'],mat_derivades_decre2,['D2:AB'
num2str(Nsortida_decre2)])

        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],llista_decre2,['A2:A' num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],llista_iso_decre2,['B2:B' num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],llista_region_decre2,['C2:C' num2str(Nsortida_decre2)])
        xlswrite([dir_ajustos, 'ajustos_decre_tipus_2','.xlsx'],m_ajustos_decre2,['D2:AB' num2str(Nsortida_decre2)])

% INCREASING TYPE-1
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'a coeff.'},'D1:D1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'b coeff.'},'E1:E1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'c coeff.'},'F1:F1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'r2'},'G1:G1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{'Tipus ajust'},'H1:H1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],{['Initial ', str_eix_y]},'I1:I1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],llista_cre,['A2:A' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],llista_iso_cre,['B2:B' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],llista_region_cre,['C2:C' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_1','.xlsx'],parametres_ajustos_cre,['D2:I'
num2str(Nsortida_cre)])

        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],llista_cre,['A2:A' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],llista_iso_cre,['B2:B' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],llista_region_cre,['C2:C' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_1','.xlsx'],mat_derivades_cre,['D2:AB' num2str(Nsortida_cre)])

        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],llista_cre,['A2:A' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],llista_iso_cre,['B2:B' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],llista_region_cre,['C2:C' num2str(Nsortida_cre)])
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_1','.xlsx'],m_ajustos_cre,['D2:AB' num2str(Nsortida_cre)])

% INCREASING TYPE-2
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'a coeff.'},'D1:D1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'b coeff.'},'E1:E1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'c coeff.'},'F1:F1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'r2'},'G1:G1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{'Tipus ajust'},'H1:H1')
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],{['Initial ', str_eix_y]},'I1:I1')
```

```
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],llista_cre2,['A2:A' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],llista_iso_cre2,['B2:B' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],llista_region_cre2,['C2:C' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'param_ajustos_cre_tipus_2','.xlsx'],parametres_ajustos_cre2,['D2:I'
num2str(Nsortida_cre2)])

        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],llista_cre2,['A2:A' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],llista_iso_cre2,['B2:B' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],llista_region_cre2,['C2:C' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'mat_der_cre_tipus_2','.xlsx'],mat_derivades_cre2,['D2:AB' num2str(Nsortida_cre2)])

        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],{'Country'},'A1:A1')
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],{'ISO3'},'B1:B1')
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],{'WHO Region'},'C1:C1')
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],llista_cre2,['A2:A' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],llista_iso_cre2,['B2:B' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],llista_region_cre2,['C2:C' num2str(Nsortida_cre2)])
        xlswrite([dir_ajustos, 'ajustos_cre_tipus_2','.xlsx'],m_ajustos_cre2,['D2:AB' num2str(Nsortida_cre2)])


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
  %%%%%%%%%%% COMPROVATION AND EXECUTION TIME %%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
disp('Process finished')
fprintf('Time calculations finished:  %4.0f%4.0f%4.0f%4.0f%4.0f%4.0f\n',fix(clock))
fprintf('Time calculations began:  %4.0f%4.0f%4.0f%4.0f%4.0f%4.0f\n',C0)
toc
fprintf('Countries analyzed: %d', j)
fprintf('Total of first adjustments: %d', counter_1_adjustments)
fprintf('Total of second adjustments: %d', counter_2_adjustments)
fprintf('Total decre-1 of first adjustments: %d', counter_decre_1)
fprintf('Total decre-1 of second adjustments: %d', counter_decre_1_2)
fprintf('Total decre-2 of first adjustments: %d', counter_decre_2)
fprintf('Total decre-2 of second adjustments: %d', counter_decre_2_2)
fprintf('Total cre-1 of first adjustments: %d', counter_cre_1)
fprintf('Total cre-1 of second adjustments: %d', counter_cre_1_2)
fprintf('Total cre-2 of first adjustments: %d', counter_cre_2)
fprintf('Total cre-2 of second adjustments: %d', counter_cre_2_2)
fprintf('Total nc of first adjustments: %d', counter_nc)
fprintf('Total nc of second adjustments: %d', counter_nc_2)
```