國立台灣科技大學

工業管理系

碩士學位論文

學號：E10301001

支氣管炎與癌症轉移之風險與經濟性評估：半馬可夫鏈分析

Risk and Economic Assessment of bronchitis and cancer metastasis: Semi-Markov chain analysis

研 究 生：David Ferreon Mestre

指導教授：王孔政 博士

中華民國一零四年二月

ABSTRACT

The current situation in Taiwan regarding the Department of Public Health expenses requires study. The gratuitousness of the system promotes an abusive consumption by the population and it is not clear if the amount of money that the government is expending in Health Care is worthy and focuses on the right issues. The aims of this study are two: to find the relationship between common chronic diseases, such as cold, influenza and asthma, respiratory diseases, chronic bronchitis, and lung cancer and its most common metastasis; and to make an estimation of the costs of the medical treatment for these diseases. The study is based on data obtained from the NHI data bank.

To do such study, the first step was gathering the data from the National Health Insurance Database (NHIDB), where patient's medical data is recorded and diseases are encoded following the ICD9 codes. Thereafter, transition times for each transition and patient were calculated by deducting the first visit date of both diseases of the transition. Owing to the absence of data regarding health and death states, some hypotheses were necessary to be assumed in order to set the transitions to such states. One state was removed from the model due to the lack of data and the final sample of population was set.

Once the data was gathered, the semi-Markov model was built. It consisted of 10 states including health and death and all the possible connections between them. Using the transition times calculated from the NHIDB time dependent transition probabilities were obtained. Furthermore, with the real treatment costs from the data base a distribution was calculated for each cost. All this was defined as parameters for the simulation.

At the same time, the model tree was built on the TreeAgePro software in order to simulate proceed with the simulation. Two different simulations were run: the first one consisted in a transient state study and the second one consisted in a steady state. Afterwards the results of both simulations were obtained and analysed meticulously.

To conclude based on the results obtained, this model can be better used to simulate and predict virtual patients for short term simulations (one year). Long term simulation

cannot be run using time dependent transition probabilities in such a big model (ten states) because simulated patients do not move to so many states.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

This chapter presents the overview of current research. It includes the problem statement, objectives of the study, the background study and the limitation constrained in this research.

1.1. Problem statement

Malignant neoplasm, most common known as cancer is one of the deadliest diseases nowadays. Worldwide, lung cancer is the most common form of cancer in terms of incidence and mortality causing about 1,180,000 deaths every year. Lung cancer is a group of diseases resulting from malignant growth of cells of the respiratory tract, particularly the lung tissue. It usually arises from epithelial cells, and may result in metastasis and infiltration to other body tissues. Neoplasms that metastasize to the lung tumours from other body parts are excluded from lung cancer definition.

The most common symptoms are breathlessness, coughing, including bloody cough, and weight loss, as well as chest pain, hoarseness, and swelling in the neck and face. Early diagnosis of lung cancer is the main factor in successful treatment. In early stages, lung cancer can be treated by surgical resection successful healing, in about 20% of cases. However, due to its virulence and the difficulty of early detection, in most cases where diagnostic and metastasis occurs, lung cancer has the worst predictions.

Lung cancer is one of the most serious diseases and one of the cancers with the highest incidence in humans, responsible for the highest rates of cancer mortality worldwide. It is the leading cause of cancer death in men and the third, after the colon and breast cancer in women, causing more than a million deaths each year worldwide. In the US, lung cancer is the most common cause of cancer deaths in both sexes, and mortality rates in women have risen 500% since 1950. In the European Union countries, although age-standardized mortality rates have decreased for most cancer sites, lung cancer mortality rates have significantly risen in women. Regarding Taiwan, a rising death rate from lung cancer has also been observed. Between 1971 and 2001, age-standardized lung cancer mortality rates per 100000 per year in Taiwan have increased sharply, from 12.66 to 32.93 among men and from 7.83 to 14.94 among women. Today, in Taiwan, lung cancer is the leading cause of cancer death in women and the second leading cause in men.

In the following table we can see the 10 leading cause of death in Taiwan and its percentage of total deaths in 2012, according to Taiwan's Ministry of the Interior's preliminary statistics. Table 1:

Table 1 Worldwide causes of death

Cause of death	**Percentage**
1.- Malignant neoplasms	28.4%
2.- Diseases of heart (except hypertensive diseases)	11.1%
3.- Cerebrovascular diseases	7.2%
4.- Pneumonia	6.1%
5.- Diabetes mellitus	6%
6.- Accidents and adverse effects	4.5%
7.- Chronic lower respiratory diseases	4.1%
8.- Hypertensive diseases	3.2%
9.- Chronic liver disease and cirrhosis	3.2%
10.- Nephritis, nephrotic syndrome and nephrosis	2.8%

As it can be seen in Table 1, malignant neoplasm is the most deadly cause in Taiwan. Furthermore, when it comes to malignant neoplasms, lung and liver cancer were still the two leading causes of cancer death in 2012. The number of deaths caused by cancer was 43,665, accounting for 28.4% of total deaths, giving a 131.3 deaths per 100,000 standard population, which was 0.7% lower than the rate of 2011 and 8.2% lower than the rate of 2001. Below is detailed the mortality rates in number of deaths out of 100,000 population depending on the type of cancer (Table 2).

Table 2 Standardized mortality rates for leading causes of death

Unit: Rate per 100,000 population, %

	2012 (A)			2001 (B)			Percentage Change		
	Both sexes	Males	Females	Both sexes	Males	Females	Both sexes	Males	Females
Malignant neoplasms	**131.3**	**170.4**	**95.1**	**143.1**	**180.7**	**104.2**	**-8.2**	**-5.7**	**-8.7**
Cancers of trachea, bronchus and lung	25.4	34.7	17.0	28.3	38.6	17.5	-10.4	-10.2	-3.0
Cancers of liver and intrahepatic bile ducts	24.7	35.8	14.4	28.0	40.9	14.9	-11.6	-12.4	-3.7
Cancers of colon, rectum and anus	14.9	18.1	12.1	15.0	17.2	12.7	-0.4	5.2	-4.9
Cancer of breast (Female)	11.6	...	11.6	10.7	...	10.7	8.1	...	8.1
Cancer of oral cavity	8.1	15.3	1.2	6.7	12.2	1.1	20.7	25.1	10.3
Cancer of stomach	6.9	9.0	5.0	10.5	13.9	7.0	-34.7	-35.5	-29.2
Cancer of prostate	6.7	6.7	...	5.9	5.9	...	13.1	13.1	...
Cancer of pancreas	4.9	5.8	4.0	4.3	5.4	3.2	13.0	8.6	24.1
Cancer of esophagus	4.9	9.4	0.6	4.4	7.9	0.7	11.6	18.5	-18.2
Cancers of cervix uteri and uterus, part unspecified	3.9	...	3.9	8.2	...	8.2	-52.6	...	-52.6

Lung cancer accounts for over 90% of lung tumours, of this 90%, 93% corresponds to primary lung cancer. The main causes of lung cancer and cancer in general, are the cigarette smoke, ionizing radiation and viral infections. Epidemiological studies have shown that cigarette smoking is the major cause of lung cancer in both sexes. However, smoking habits do not seem to be the main explanation of the epidemiological characteristics of female lung cancer mortality in Asian countries, where the prevalence of smoking is relatively low but lung cancer mortality rates are relatively high. Factors other than smoking habits might contribute to the variability in lung cancer mortality. That is one of the main reasons why this study is focused on finding relationships between lung cancer and other diseases and search for a pattern to explain possible causes of lung cancer. Moreover, this study is focused to ascertain whether lung cancer and metastatic cancer is related with another major disease and with different common chronic diseases.

On the other hand, the treatment of a disease such as lung cancer can be long and enormously expensive. That is why a cost study is also needed, in order to explain how much this kind of treatments costs to the state. And in the future this study can be a guide to study a possible adjustment or improvement for the budget set for such treatments.

1.2. Objectives

The objectives of this study were (1) to understand the relationship between lung cancer and its possible metastasis as well as with other major diseases, by compiling the

patient's clinical records from National Health Insurance research database (NHIDB) in Taiwan, (2) to construct the semi-Markov and simulation model which represent the evolution of patients diagnosed with any of these diseases and (3) to conduct a cost study involving treatment of patients diagnosed with diseases that are listed in this thesis.

1.3. Background Information

The following sections contain a basic introduction about the most remarkable diseases studied in this thesis, as well as some statistics about the Taiwanese population diseased by them.

1.3.1. Bronchitis

Bronchitis is an inflammation of the mucous membranes of the bronchi (the larger and medium-sized airways that carry airflow from the trachea into the more distal parts of the lung parenchyma). Bronchitis can be divided into two categories: acute and chronic.

Acute bronchitis is characterized by the development of a cough or small sensation in the back of the throat, with or without the production of sputum (mucus that is expectorated, or "coughed up", from the respiratory tract). Acute bronchitis often occurs during the course of an acute viral illness such as the common cold or influenza. Viruses cause about 90% of acute bronchitis cases, whereas bacteria account for about 10%.

Chronic bronchitis, a type of chronic obstructive pulmonary disease (COPD), is characterized by the presence of a productive cough that lasts for three months or more per year for at least two years. Chronic bronchitis usually develops due to recurrent injury to the airways caused by inhaled irritants. Cigarette smoking is the most common cause, followed by exposure to air pollutants such as sulphur dioxide or nitrogen dioxide, produced by vehicle exhausts and occupational exposure to respiratory irritants. Individuals exposed to cigarette smoke, chemical lung irritants, or who are immunocompromised have an increased risk of developing bronchitis. Other respiratory diseases, such as cold and influenza often cause chronic bronchitis. Individuals diagnosed with asthma have a higher chance of eventually develop chronic bronchitis. Symptoms of chronic bronchitis may include wheezing and shortness of breath, especially upon exertion and low oxygen saturations. The cough is often worse

soon after awakening and the sputum produced may have a yellow or green colour and may be streaked with specks of blood.

Individuals with an obstructive pulmonary disorder such as bronchitis may present with a decreased FEV1 and FEV1/FVC ratio on pulmonary function tests. Unlike other common obstructive disorders such as asthma or emphysema, bronchitis rarely causes a high residual volume (the volume of air remaining in the lungs after a maximal exhalation effort).

Evidence suggests that the decline in lung function observed in chronic bronchitis may be slowed with smoking cessation. Chronic bronchitis is treated symptomatically and may be treated in a nonpharmacologic manner or with pharmacologic therapeutic agents. Typical nonpharmacologic approaches to the management of COPD including bronchitis may include: pulmonary rehabilitation, lung volume reduction surgery, and lung transplantation. Inflammation and edema of the respiratory epithelium may be reduced with inhaled corticosteroids. Wheezing and shortness of breath can be treated by reducing bronchospasm (reversible narrowing of smaller bronchi due to constriction of the smooth muscle) with bronchodilators such as inhaled long acting β2-adrenergic receptor agonists (e.g., salmeterol) and inhaled anticholinergics such as ipratropium bromide or tiotropium bromide. Mucolytics may have a small therapeutic effect on acute exacerbations of chronic bronchitis. Supplemental oxygen is used to treathypoxemia (too little oxygen in the blood) and has been shown to reduce mortality in chronic bronchitis patients. Oxygen supplementation can result in decreased respiratory drive, leading to increased blood levels of carbon dioxide (hypercapnea) and subsequent respiratory acidosis.

1.3.2. Primary Lung Cancer

Primary lung cancer (carcinoma of the lung or pulmonary carcinoma) is a malignant lung tumour characterized by uncontrolled cell growth in tissues of the lung. It can spread beyond the lung by process of metastasis into nearby tissue or other parts of the body. The main primary types are small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). The most common symptoms are coughing (including coughing up blood), weight loss, shortness of breath, and chest pains.

The majority of cases of lung cancer (80–90%) are due to long-term exposure to tobacco smoke. About 10–15% of cases occur in people who have never

smoked. These cases are often caused by a combination of genetic factors and exposure to radon gas, asbestos or other forms of air pollution. Lung cancer is seen by chest radiographs and computed tomography scans.

Treatment and long-term outcomes depend on the type of cancer, the stage (degree of spread), and the person's overall health, measured by performance status. Common treatments include surgery, chemotherapy, and radiotherapy. NSCLC is sometimes treated with surgery, whereas SCLC usually responds better to chemotherapy and radiotherapy.

As shown on the previous section, lung cancer is the one which contributes the most on the number of annual deaths. When combining this fact with the long history of lung diseases in the Taiwanese population, it seems worth studying lung cancer and other lung diseases' connections.

1.3.3. Cancer Metastasis

Lung cancer can lead to metastasis and infiltration to other tissues of the body. That is why this study includes cancer metastasis from primary lung cancer. The most common places where the metastatic cancers coming from the lung develop are:

- Brain metastatic cancer
- Bone metastatic cancer
- Liver metastatic cancer
- Adrenal glands metastatic cancer

Nonetheless, data for the patients suffering from a metastatic cancer original from lung primary cancer are scarce, due to the high deadly rate of primary lung cancer.

1.3.4. Semi-Markov Model

When the interest on health care and medical studies, including economic evaluation on that field, started to gain popularity, new statistical methods for analyzing the data emerged constantly. At the beginning regression models were used in order to estimate patient's survivability and evaluate the medical costs incurred, until Castelli argued that such methods were not satisfactorily fitted to survival data and could not take the clinical progression of the disease into account. To solve this, several analysts favored to use multi-state model, such as decision tree or ordinary Markov chain to analyze

healthcare issues, giving preference to Markov model. Even so, Markov model is considered to have a far too restrictive constraint to address clinical problems, since it has a strong memoryless assumption and constant transition probability. In return, an extension of an ordinary Markov model which relaxed the underlying Markovian assumption in more flexible manner and allowed the sojourn times to be fitted into various distribution, defined as Semi-Markov model, was designed to handle and generalize those assumptions since it involves the distribution of waiting time as model parameters, so that an accurate representation of real clinical settings can be achieved. In conclusion, the Semi-Markov model is more suitable for this study since it can represent with more precision the reality of a disease in which the development continuously changes over time.

Moreover, since the waiting time was explicitly distributed, three nested-possible waiting time distribution were proposed to be used in semi-Markov modeling: (i) basic exponential distribution, (ii) Weibull distribution and (iii) generalized Weibull distribution. It has been demonstrated that the Weibull distribution is more suitable for clinical studies.

Furthermore, the Semi-Markov model divides de evolution of the process in several independent states. All the important events are defined as transitions from one state to another and the over time is divided in intervals called cycles , which are defined to represent a significant interval of time that in the progression of the process. In every cycle, each subject can either stay in the same state or change to another, except when they are in an absorbing state.

Basing on the properties and capacities described above, the Semi-Markov model is considered to be the most appropriate for this study. Finally, this model will be evaluated using the Monte Carlo simulation, which will be explained in the next section.

1.3.5. Monte Carlo Simulation

Monte Carlo simulation, it is a type of computer simulation that has been regarded as a fundamental numerical experimentation technique in probabilistic analysis. The main capability of Monte Carlo simulation is to translate uncertainties (statistics) sampled from model input variables in each of experiment trials into a set of statistically computed output variables. In other words, Monte Carlo simulation performs multiple

deterministic analyses of the system for different sets of input variables value which is sampled in a basis of a series of sampling distribution before calculates the required statistics of the possible outcomes. Moreover, Monte Carlo simulation provides a method for evaluating the long term effect of random variability of the input parameters on the outcomes. In decision analysis, specifically, the goal of such a simulation is generally to calculate an expected value for each strategy being compared on the basis of the average of the random walks elapsed.

Since its introduction, this simulation has been applied in an enormous number of study in various research fields, including in healthcare and biomedicine study. One of the usefulness of Monte Carlo simulation is its ability to assess the variability of the experimental outcome which can be occurred due to the individual-level random walks (1^{st}-order uncertainty).

In the other hand, Monte Carlo simulation also can be used to calculate an expected value over parameter uncertainty (2^{nd}-order uncertainty) for particular strategy. In fact, Monte Carlo simulation is powerful to calculate cost-effectiveness ratios and evaluate the uncertainties involved in the model (what-if analysis). Indeed, a number of researchers have applied Monte Carlo simulation into their study in order to perform probabilistic sensitivity analysis (PSA) for the Markov model, specifically in medical decision analysis. In general, the Monte Carlo PSA is used to sample the parameter from its probability distribution and assesse their affection to the model outcomes, presented as incremental cost, incremental benefit and the incremental value of their ratio. By using such approach, the most cost-effective strategy can be identified.

In addition, another advantage of using a Monte Carlo simulation is the fact that Monte Carlo simulation analysis has an ability to capture the history of each patient's clinical progression without a need to expand the Markov state space. By that way, a complex Markov structure (e.g. by using tunneling state) can be avoided. This feature allowed the dynamic transition probability to be implemented while keeping the model as simple as possible.

1.4. Restrictions and limitations

A nationwide patient from NHI research database in Taiwan was used in this research. However, the database only recorded registration files and original claim data for

reimbursement, administratively. It did not include any clinical record information of the patients (e.g. stage of the cancer on the diagnosis, disease severity progression). As the result, the definition and the evolution of the health condition of the patient can only be illustrated in a very limited way. Thus, several hypotheses had to be made in order to adjust the data for the study. Therefore, the results depend on these hypotheses and may not be 100% realistic.

Moreover, one of the states initially designed in the model had to be removed due to the shortage of the sample that represented the population diagnosed with that disease.

CHAPTER 2 MATERIALS AND MODEL DEVELOPMENT

This chapter consists of three major parts. First section covers the explanation of the Markov modelling. The second section explain the data structure used in this study. It includes the description of the data source and its inclusion criterion as well as the pre-processing sequences from NHI database. Subsequently, the third part of this chapter describes the simulation structure developed for the analysis in this study.

2.1. Multi-State Semi-Markov modelling

Before the data used in the study is descripted, an explanation of the multi-state Semi-Markov model used in this study is needed. The semi-Markov process is a particular case of ordinary Markov model in which the restrictive homogeneous transition probabilities of Markov property assumption are relaxed. Moreover, in Semi-Markov model the transition probabilities depend on the time that a variable has been on the current state, that is why fits better in medical models than traditional Markov models.

2.1.1. Model definition

The main diseases studied in this thesis are bronchitis and lung cancer and a focus of this study is also the relationship among them. In addition, on one hand, this thesis pursue to study the causes of these diseases by prospecting for relationships with common diseases as cold, influenza or asthma with bronchitis; on the other hand it also aim to obtain relationships between lung cancer and the most common metastasis induced by it, such as brain, bone, liver and adrenal gland metastasis. In order to create a model capable to interpret all these relations we designed a Semi-Markov model disintegrated into 11 different states, 9 states have been assigned to the 9 different diseases that have been studied and the other two states correspond to the absorbing states of health and death. The connections between the different states have been decided following two criteria, only the states that can be directly connected in reality and the ones that have connections which fit in this study are connected in the model. For example, normally nobody dies because of having a cold; consequently there is no connection between the cold state and the death state. Following with the example about the cold disease, as long as we are not interested in the direct relationship between having a cold and then developing a lung cancer, in the model there is no direct connection between the cold state and the lung cancer state.

After obtaining and studying the data used to define the transition time between states, one state was detected to have too few data. This entails that in this specific sample there wasn't enough data to represent the population. Thus, the state was removed from the model leaving the model with 10 states. The removed state was the Adrenal Gland Metastasis state. A detailed explanation of the different states will be done in the next section.

2.1.2. Model structure

As it is said previously in this chapter, the model is characterized by a disaggregation of the problem into 10 different states in order to study the phenomenon among state transition.

The first three states correspond to cold, influenza and asthma, which are the most common causes that can drive into bronchitis. That is why these three states are connected with the bronchitis state, the next state of the model. The bronchitis state leads to the lung cancer state, as one of the main objectives of this study is to understand the relationship existent between these two diseases. The next three states are the four kinds of metastasis mentioned above in this chapter minus the one that has been removed due to a lack of data; it is assumed that these are the only possible metastasis that can occur if a patient has lung cancer. It is important to notice that these four states only refer to patient that had the respective metastasis after having lung cancer, they doesn't refer to a primary cancer of any kind.

The last two states are the absorbing states that define if a patient is dead or he/she has healed. The death and health state were more complicated to define, since there is not clear information about them in the data. Two assumptions had been done in order to define these states. It is assumed that a patient is death when there is record of his death and he/she stops coming to the hospital after having a treatment longer than 1 year, if the patients had a treatment shorter than one year and then there is not more record of him/her this patient is assumed to be healthy. The reason of these hypothesis relapse on information from previous studies which say that a patient of lung cancer that has the disease more than one year has few possibilities of surviving.

The following table make a classification of the Markov states with a brief description and the number assigned to each one. In addition, a diagram of the model is incorporate for the purpose of help the comprehension of the model.

Table 3 Markov states classification and numbering

State	Description	Number
Cold	Patients diagnosed with cold who are diagnosed with bronchitis later on	1
Influenza	Patients diagnosed with influenza who are diagnosed with bronchitis later on	2
Asthma	Patients diagnosed with asthma who are diagnosed with bronchitis later on	3
Bronchitis	Patients diagnosed with bronchitis	4
Primary lung cancer	Patients diagnosed with primary lung cancer	5
Brain metastasis	Patients diagnosed primary lung cancer with a metastasis on the brain	6
Bone metastasis	Patients diagnosed primary lung cancer with a metastasis on the bones	7
Liver metastasis	Patients diagnosed primary lung cancer with a metastasis on the liver	8
Health	Patients assumed to be healthy	9
Death	Patients assumed to be death	10

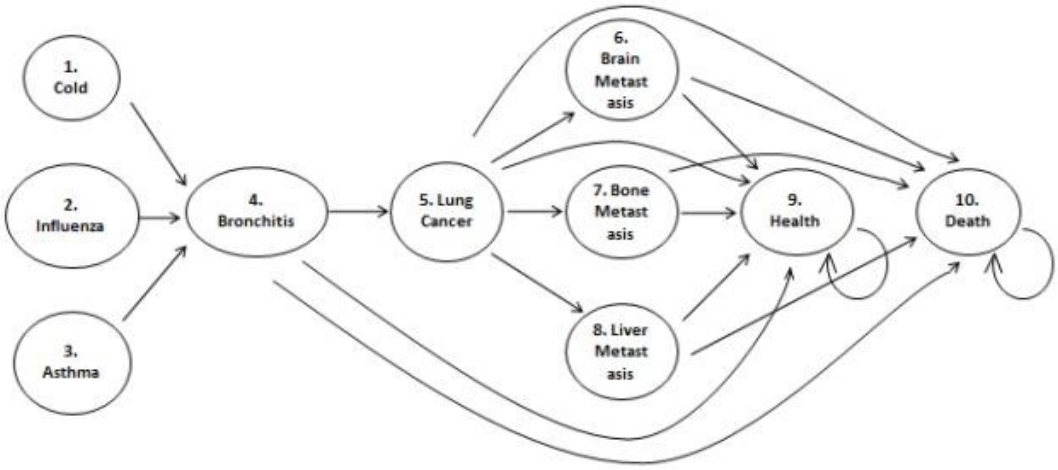

Figure 1 Semi-Markov states diagram

As it is explained in the previous chapter, patients diagnosed with Cold, Influenza and Asthma have a higher chance of eventually develop chronic bronchitis. Furthermore, it is known that having previous bronchitis pathology, such as chronic bronchitis increases the chances of eventually develop lung cancer. Moreover, as described in the previous chapter too, bone, brain and liver metastasis are three of the most common ones in lung cancer patients. Here two different literatures that support the links of this model can be fund:

- María Concepción Beuses Salcedo (2004). Primary malignant neoplasms of lung correlation anatomo clinic: study of twenty years of autopsy general hospital of south.
- Michael A. Beckles, MB, BS; Stephen G. Spiro, MD; Gene L. Colice, MD, FCCP; Robin M. Rudd, MD (2003). Initial Evaluation of the Patient With Lung Cancer: Symptoms, Signs, Laboratory Tests, and Paraneoplastic Syndromes

2.2. Data Description

Once the modeling procedure is clear, an explanation of the data structure used in the study can be given without create confusion. This section describes the sources of the data along with the definition used to extract the desirable data.

2.2.1 Data Source

The data used in the project is taken from the National Health Insurance Data Bank (NHIDB). The study population, which include patients with Cold, Influenza, Asthma,

Bronchitis, Lung Cancer and Brain, Bone, Liver or Adrenal Gland Metastasis were identified from 1997-2011 Inpatient Admission (DD) and Outpatient Ambulatory Care Visit (CD), Inpatient Admissions have information about hospitalizations that lasted more than one day and Ambulatory Care Visits regard singular visits. Those types of diseases were selected on the basis that lung cancer is the deadliest cancer in Taiwan and the most common metastasis occurred in a lung cancer patient are bone, brain, liver and adrenal gland. Moreover, the objectives of this study include the search for a relation between lung cancer and an important respiratory disease such as bronchitis and other common diseases that can cause it. They were then investigated and distinguished by their ICD-9-CM diagnostic code. The ambulatory visit files provide up to three ICD-9-CM diagnostic codes while the inpatient admission files provide up to five diagnostic codes. Those codes allow identifying the diagnosed diseases for each visit. The information is stored in columns and is really detailed; there are lots of specific data of every visit. Disease code, ID number, Date of birth, Physician name, visit date, medical department, diagnostic procedure, drug and treatment cost, treatment procedure and payment information are some of the most remarkable categories. As an introduction, this text will not show a copy of each category of the database.

As it will be specified in the second part of this chapter, in addition to the diseases mentioned above the model has two other states which correspond to the absorbing states that define when a patient is healed or dead. In order to define if a patient is dead; the information was taken from the inpatient admissions files. And if a patient didn't have a record of his death and he didn't have any visit after a treatment longer than one year he/she was also considered dead, if otherwise the treatment had been shorter than one year he or she was considered to have healed.

One of the strengths about using the NHIDB for the research is the sample size. It covers over 99% of Taiwan's 23 million population. Furthermore, it is so detailed that provide us information about medical expenditures, including the diagnosis procedures, the costs of each treatment or drug prescriptions. Moreover, we can find various personal information of each patient that may be useful for the study.

On the other hand, it is important to notice that the data was not recorded for a statistical study like this; instead it was designed as administrative data and there is some information that is missing or it has to be assumed in order to complete the model.

Furthermore, the data format has been changed in order to have it all together with the same format to make it suitable for study.

Finally, it must be remarked that to comply with Taiwan's Personal Information Protection Act, every patient, medical provider and medical staff is codified with an encoded personal identifier. Therefore, all data in this study contain the real tracing of the different diseases patients at their clinical visits and they were analysed anonymously.

2.2.2. Data Definition and Criterion

As it is explained before, this study is focused on the following diseases and their relationship: cold, influenza, asthma, bronchitis, lung cancer, brain, bone, liver and adrenal gland metastasis. It is important to keep in mind that the types of cancer studied that are not lung cancer are only metastasis and the data that has been used comes from patients who have been diagnosed of lung cancer and then one of these metastasis. The ICD-9-CM diagnostic code is a coding system that has a unique code number for each disease.

In ICD-9-CM coding system, respiratory diseases are coded in the range of 460 for acute respiratory infections (460-466), other diseases of the upper respiratory tract (470-478), pneumonia and influenza (480-488), chronic obstructive pulmonary disease and allied conditions (490-496), pneumoconiosis and other lung diseases due to external agents (500-508), up to other diseases of respiratory system (510-519). Furthermore, the cancers and neoplasms are coded between 140 and 239, including malignant neoplasm of respiratory and intrathoracic organs (160-165), secondary malignant neoplasm of respiratory and digestive systems (197) and secondary malignant neoplasm of other specified sites (198).

As for the objected patients of this study, we defined the corresponding ICD-9-CM code for cold as 460; ICD-9-CM code of 487.0 - 487.8 for influenza; ICD-9-CM code of 493.0 – 493.9 for asthma; ICD-9-CM code of 490 for bronchitis; ICD-9-CM code of 162.2 – 162.9 for primary lung cancer; ICD-9-CM code of 198.3 for brain metastasis; ICD-9-CM code of 198.5 for bone metastasis; ICD-9-CM code of 197.7 for liver metastasis; ICD-9-CM code of 198.7 for adrenal gland metastasis. It is important to notice that every disease condition has its own ICD-9-CM code, including metastasis cases. Furthermore, one of the focuses of this study is to obtain a relation model

between primary lung cancer and its more common metastasis, so it is critical to have distinct definition of it and not confuse metastatic cancer with a possible second primary cancer. To summarize, Table 4 shows the list of code used to differentiate each disease.

Table 4 ICD-9-CM definition of the objected diseases

Disease	**ICD-9-CM**
Cold	460
Influenza	487.0 - 487.8
Asthma	493.0 – 493.9
Bronchitis	490
Lung Cancer	162.2 – 162.9
Brain Metastasis	198.3
Bone Metastasis	198.5
Liver Metastasis	197.7
Adrenal gland Metastasis	198.7

As it is said before in this chapter, the Adrenal gland Metastasis had to be removed due to the lack of data, so at the end the data regarding this metastasis was not used.

2.2.3. Population study

The selection procedure of the study population in this study is chronologically described as follows. First, we used ICD-9-CM code to identify patients diagnosed with two of the diseases described at Table 4, these two diseases needed to be consecutive in the model and the patient at issue needed to be diagnosed with the two diseases at right order. This procedure was done for every pair that was defined to be related in the model. Secondly, we obtained the transition time between the two states for each pair by calculating the difference between the dates of the first visit for the first disease and the first visit for the second disease, every time for the corresponding patient and for each pair of diseases. Once done that, the next step was to put together all the transitions. After all this, we needed to identify and add to the study population the possible transitions to the health or death states for every patient. The procedure to

acquire it was as follow. First of all we identified which patients had a transition to the state Health and which ones had it to the state Death. The data prevenient from the NHI data base has a column that indicates whether a patient has dead or not, but since the proportion of deaths was to low it was clear that some information about the deaths was missing. Thus, some hypotheses were required. If a patient was going to be considered dead after the treatment he/she needed to fulfill one of the two following conditions, whether there record of his/her death in the data base or his treatment had been longer than one year. Whether his/her treatment had been shorter than one year and there wasn't record of his/her death, the patient was considered to heal from the disease. This hypothesis was made after an exhaustive research in which was found that after one year of treatment for the diseases of study the probability of survival was too small.

After that, the transition time for these two states was calculated. Again an assumption was needed here, it was considered that the transition time for these two states was the difference between the date of the first visit of the corresponding disease and the last visit of the same disease. It is true that there is not an exact date of the death or the healing of the patients, but since the matter of study is the duration and the cost of the treatment, it is considered that the death or health takes place when the treatment is finished.

Moreover, as it is said before in this chapter, the Adrenal gland Metastasis had to be removed due to the lack of data, so at the end the data regarding this metastasis was not used.

2.3. Simulation Procedure

2.3.1. Identification of Transition Probabilities

In order to be able to start the simulation, the next step was to determine the transition matrix. And to do so the first thing to know is that one fundamental property of the Semi-Markov process is that the transition probabilities depend on the duration of time spent in the current state. This information was identified by analysing the transitions experienced by each patient in the previously selected population. And the transition matrix was calculated using the package for R software called "semiMarkov" package (Listwoń-Krol & Saint-Pierre, 2014). The calculation of the transition probabilities among states was described as follows.

Let us consider a model with k state belonging to finite state space $E = \{1, 2,\ldots, k\}$. Then, consider $X_0, X_1, X_2,\ldots, X_n \in E$ be the successive states in N visits by a random process, in which $0 = T_0 < T_1 < \ldots < T_n$ are the consecutive entrance times into each of these states. In the form of homogeneous Markov chain, the probabilities to move to state j from state i can be written as:

$$P_{ij} = P(X_{n+1} = j \mid X_n = i) \qquad (2.1)$$

If state i is transient (not an absorbing state), then $P_{ij} \geq 0$ for $i \neq j$ and $P_{ij} = 0$ for $i = j$ while $P_{ij} = 0$ for $i \neq j$ and $P_{ij} = 1$ for $i = j$ if otherwise. However, it is obvious that the process does not deal with time issues. Therefore, it is considered as semi-Markov process if the random process regards the transition sojourn time $(T_{n+1} - T_n)$ and its distribution satisfies:

$$Q_{ij}(t) = P(X_{n+1} = j,\ T_{n+1} - T_n \leq t \mid X_n = i) \qquad (2.2)$$

The probability density function and the probability distribution function of the waiting time in state i before passing to state j is given by:

$$f_{ij}(t) = \lim_{\Delta t \to 0} \frac{P(t < T_{n+1} - T_n < t + \Delta t \mid X_{n+1} = j,\ X_n = i)}{\Delta t} \qquad (2.3)$$

Subsequently, we deduced from (3.3) the cumulative probability function, $F_{ij}(t)$ and the corresponding survival function of waiting time in state i $S_{i.}(t)$ as defined by:

$$F_{ij}(t) = P(T_{n+1} - T_n \leq t \mid X_{n+1} = j, X_n = i) = \frac{Q_{ij}(t)}{P_{ij}} \qquad (2.4)$$

$$S_{i.}(t) = 1 - P(T_{n+1} - T_n \leq t \mid X_n = i) = \sum_{j \in E} P_{ij}\,(1 - F_{ij}(t)) \qquad (2.5)$$

Moreover, we considered Weibull distributions to be used as the waiting time distribution. It generalizes exponential distribution by incorporating two parameters $W(\sigma_{ij}, \nu_{ij})$, which is more flexible and well adapted to various shapes, especially when it was used in survival and medical studies (Castelli et al., 2007). Therefore, following the application of Weibull distribution, instead of exponential distribution, the semi-Markov model is then generalized. The hazard function is defined by:

$$\alpha_{ij}(t) = v_{ij}\left(\frac{1}{\sigma_{ij}}\right)^{v_{ij}} t^{v_{ij}-1}, \forall\, t \geq 0, \forall \sigma_{ij} \geq 0, \forall v_{ij} \geq 0 \tag{2.6}$$

Eventually, the hazard function of the semi-Markov process, which represents the probability of transition towards state *j* between time *t* and *t* + Δ*t*, given that the process is in state *i* for a duration *t* can be drawn. It follows:

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P(X_{n+1} = j,\, t < T_{n+1} - T_n < t + \Delta t \mid X_n = i,\, T_{n+1} - T_n > \mathrm{t})}{\Delta t}$$

$$= \frac{P_{ij}\, f_{ij}(t)}{S_{i.}(t)} = \frac{P_{ij}\, S_{ij}(t)\, \alpha_{ij}(t)}{S_{i.}(t)}, \qquad \begin{matrix} i, j \in E \\ i \neq j \end{matrix} \tag{2.7}$$

$$\lambda_{ii}(t) = -\sum_{i \neq j} \lambda_{ij}(t)$$

The result from (3.7), can be interpreted as the subject's risk of progressing from state *i* to state *j* after stayed in state *i* for *t* duration (Listwon-Krol & Saint-Pierre, 2014). Later on, those values were derivatively inherited to the simulation procedure as the input for the transition probability among Markov states.

2.3.2. Simulation Model Structure

The expected risk and survivability of the patient as well as their expected monetary costs have been analysed by following hypothetical patient's progress through the Semi-Markov model. This analysis was carried out by using Monte Carlo simulation procedure in TreeAge Pro 2014 software, which presented the Semi-Markov model in a probability tree structure as shown in the next figure. It should be noted that this simulation has been done with hypothetical patient cohort. The study population from NHIDB only served as the basis and source for input parameter data.

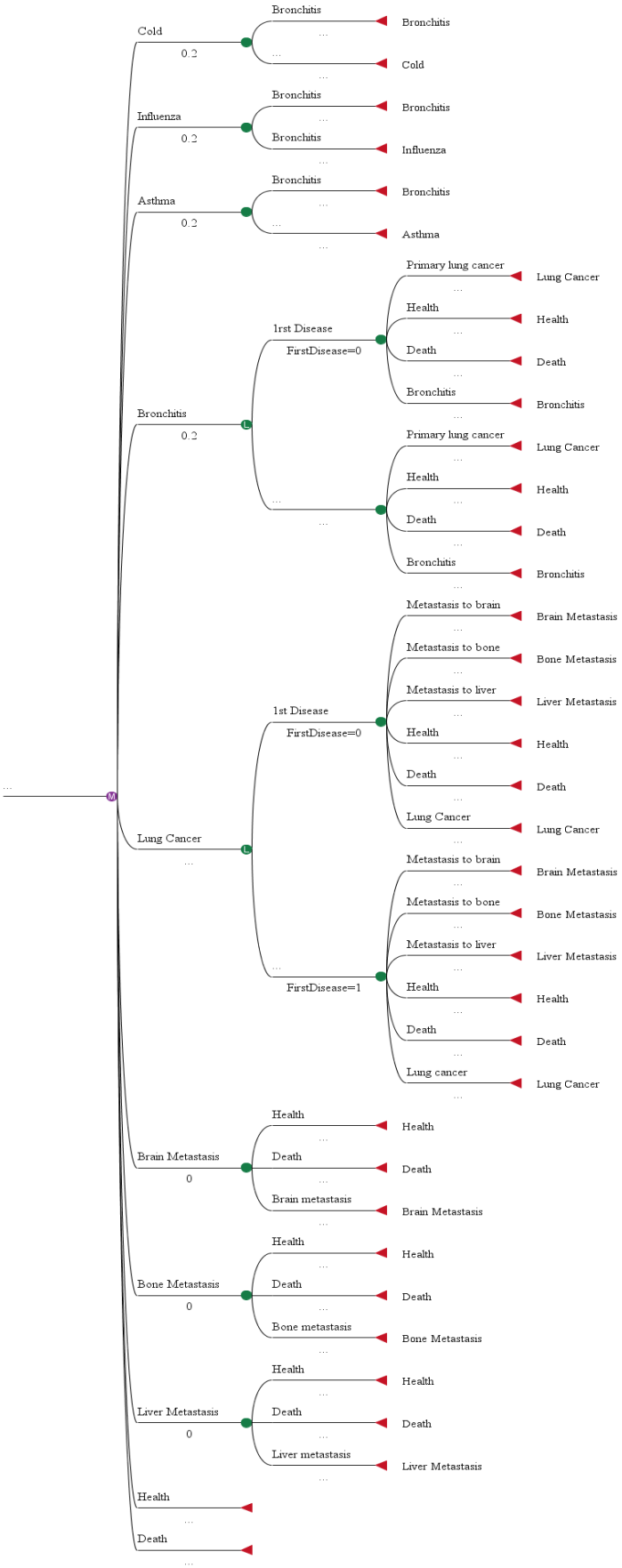

Figure 2 Semi-Markov model simulation structure

The simulation model basic structure can be observed in the figure above. As it is explained in section 3.2, two different simulations were done and the initial provability was different for each simulation. In Figure 2 the initial probabilities of the steady state simulation can be seen, it is all better explained in section 3.2 and 3.3.3. In both simulations a zero probability was set for the metastasis states and the two absorbing states (health and death). Moreover, it is important to notice that while cold, influenza and asthma can only be initial states, bronchitis and lung cancer can be both initial and not initial states. Those different possibilities were marked with a logic node in the tree structure. Since the possible evolution of both diseases doesn't depend on whether it is an initial state or not, the chance nodes that represent possible later events are the same in both cases (initial and not initial state). This distinguish was done in order to record hypothetical patients of each case.

As it is mentioned before in this report, the transition probabilities depend on the time spent in the first state of the transition. That is why, in the simulation, the transition made by each patient, during each cycle, depends on the duration of stay in the current state; and the probability of the chance nodes was set following this concept. These time-dependent probabilities were obtained as explained in the previous section.

Furthermore, a set of Markov rewards were also given to the patient during their stay in a particular state in each cycle. Each non-transient state possesses a different set of rewards. This set includes the survivability (lifetime) reward and cost reward. Both of them will be calculated and reported at the end of the simulation. The cost reward sets are obtained from NHI database. In addition, different tracker variables were also defined to track additional information about events experienced by each patient, such as the duration of one disease before the next was developed or the different costs of each disease. Moreover, it is important to notice the difference between the lifetime reward or time trackers and the cost rewards or cost trackers. Lifetime reward and all the time trackers only need a simple addition of 3 for each cycle that a patient is in the correspondent state, since a cycle has duration of 3 months and the results will give the duration in months. But when it comes to the cost issue it is not that simple, in the study population every patient has a different cost for each treatment so a distribution for each cost was made from the real data using the EasyFit software, then those distributions were input into the TreeagePro. In order to make the simulation calculate the cost of each treatment, a tracker and a variable for each cost was defined, the variable will take

a different value from the distribution every cycle and the tracker will save the addition of these costs for each cycle.

Furthermore, half-cycle correction (Sonnenberg & Beck, 1993) is applied to the simulation model in this study. This process is carried out to comply with the assumption that in reality, however, transitions occur gradually in continuous fashion throughout each cycle. By applying this correction, transitions is assumed to occur, on average, in half-way of a cycle, instead of simultaneously at the end of a cycle which may lead in overestimation of expected results.

CHAPTER 3 EXPERIMENTS AND RESULTS

This chapter describes the experimental procedures carried out during this research. It explains the analysis sequences in both semi-Markov model and Monte Carlo simulation model. Finally, at the last part of this chapter the results from all procedures and experiments performed in this study are presented.

3.1. Semi-Markov modeling

As said above, the purpose of this section is to explain the calculation of the transition probability for each transition of the model. As it is well explained in another chapter of this report, these transition probabilities are time dependent and are obtained using (3.7) in the Semi-Markov calculation from the lead time of the patients in the NHI data base. It is also worth noting that the Weibull distribution is used as the distribution of the waiting time for this study. Defined this, the Semi-Markov model has been generalized.

To generate this time-dependent hazard rate among states the real waiting times for each transition from the NHI database data has been employed. The real data has been used to obtain the transition probabilities from one state to another that will be used in the simulation model. Once the transition probabilities were obtained we could proceed to the simulation process.

3.2. Monte Carlo simulation

In the following section the procedure to realize the simulation is described. Two different simulations have been executed. The first one consists of a transient state simulation, where 50,000 hypothetical patients have been generated and have been simulated through the semi-Markov simulation model for 4 cycles using Monte Carlo method. As each Markov cycle has been defined to be 3 month long, the total simulation time horizon is 12 months, that is to say 1 year. When it comes to the second simulation, again 50,000 hypothetical patients have been generated, but this time they have been simulated through the semi-Markov simulation model for 80 cycles, what means 20 years. Since 20 years is a larger period of time than the average survivability time of the patients diagnosed with lung cancer, this length brings the simulation into a steady state.

The Monte Carlo method allows running the simulation for each hypothetical patient individually. Thus, at the end of the simulation detailed information about each patient can be obtained. The results will be explained in the next section.

Following the simulation structure explained in 2.3.2., for the transient state simulation, at the beginning of the simulation each patient in a hypothetical cohort was imposed to one of the diseases set as initial states with a proportion equivalent to the proportion of real patients that are diagnosed with one of these diseases. This proportion is taken from the study population obtained from the NHI data base. When it comes to the steady state simulation, each patient in a hypothetical cohort was randomly imposed to one of the diseases set as initial states with equal proportion for each disease at the beginning of the simulation. Since the Semi-Markov theory says that a transition only depends on the current state, the initial probabilities will not affect the final results when the simulation arrives at a steady state where all the possible transitions have been experienced.

Moreover, for both cases, the starting state is interpreted as the first time that a patient is diagnosed with one of the diseases. In subsequent cycle, patients were able to either remain in their assigned state, which mean they weren't diagnosed with another disease or didn't develop metastasis during the duration of the cycle, or progress to a new state (different disease or metastasis if they are in the lung cancer state). This means that patients will start the next cycle at the state they have ended up in the previous cycle. If a patient was either in the bronchitis state or the lung cancer state, he/she could also progress to the death state or to the health state. Once a metastasis was developed, patients continued to stay in their current state until they died or they heal. The simulation stops when either the patient dies or the end of simulation period has been reached, whichever comes first. A new simulation for the next patient runs subsequently after the previous simulation has finished.

The aggregate time between the different diseases and until the development of metastasis, the survivability from bronchitis, lung cancer and metastasis condition, along with the lifetime medical costs and the costs for the treatment of each disease or a possible metastasis were estimated and evaluated through Monte Carlo simulation.

3.3. Results

This section describes the results of every analysis done in the study. The baseline data used as the input for some analysis is also explained.

3.3.1. Baseline data

This study included 57,760 patients as the baseline data of the investigation, spread among the different diseases. 67.94% of them are adults aged more than 50 years old and the 51.06% of the total number of patients are female. Taking a close view to the data, we can notice that there are different ranges of ages, going from babies to elder people, and an equal proportion of males and females, with almost half proportion of each. All the patients in the study population have been diagnosed with at least two diseases included in the model, and these two diseases are always consecutive in the model. To make an example, whether there is a patient in the study population that is diagnosed with lung cancer, either he/she is also diagnosed with bronchitis (the state that precedes the lung cancer state) or he she is also diagnosed with one of the three metastasis included in the mode, which are the three possible states where a patient can move from lung cancer. Moreover, all these patients represent the real cases in Taiwan of the diseases studied and the data comprehend 103,895 different transitions. As it can be seen in the following table, around 60% of the transitions are the transitions between the common diseases and bronchitis, with a clear dominance of the cold disease with more than 40% of the transitions. This is explained by the fact that cold, influenza and asthma are really common not only in Taiwan but all around the world, and since it is extremely easy to have access to the healthcare in this country, the number of visits regarding this kind of diseases is very large. In regard to the transition between bronchitis and lung cancer, there are 2,382 transitions, which is more than the 2% of them and a well sized sample for the study. When it comes to the transitions among lung cancer and metastasis, it can be seen that the number of transitions is considerably smaller. This is caused by the fact that there are fewer cases of this kind in the reality, since cancer is far less common in Taiwan compared with the other diseases included in the study. As it is said in a previous chapter of this report, one metastasis was removed of the study because of the small sample obtained from the data regarding it, but the other three metastasis were maintained as they gave a large enough sample.

Table 5 Base profile and proportion of the study population

	Characteristics	No. of patients	Percentage (%)
Gender	Female	29481	51.06%
	Male	28257	48.94%
Age	> 50	18520	32.06%
	≤ 50	39240	67.94%

Initial state	Final state	No. of transitions	Percentage (%)
Cold (state 1)	Bronchitis (4)	45,728	44.014%
Influenza (state 2)	Bronchitis (4)	2,254	2.169%
Asthma (state 3)	Bronchitis (4)	14,403	13.863%
Bronchitis (state 4)	Primary lung cancer (5)	2,382	2.293%
	Health (9)	36,406	35.041%
	Death (10)	48	0.046%
Primary lung cancer (state 5)	Brain metastasis (6)	66	0.064%
	Bone metastasis (7)	95	0.091%
	Liver metastasis (8)	21	0.020%
	Health (9)	1,770	1.704%
	Death (10)	532	0.512%
Brain metastasis (state 6)	Health (9)	43	0.041%
	Death (10)	52	0.050%
Bone metastasis (state 7)	Health (9)	20	0.019%
	Death (10)	46	0.044%
Liver metastasis (state 8)	Health (9)	16	0.015%
	Death (10)	13	0.013%
TOTAL		103,895	100.000%

Once the study population was set, the next step was to calculate the waiting time until the next event for each transition experienced by every patient in the study population. This waiting time, also called transition time, is used as the input parameter for the Semi-Markov model and for the transition probabilities calculation. The following table compiles the average waiting time until every event is experienced.

Table 6 Average waiting time between transition among states (month)

	Cold	Influenza	Asthma	Bronchitis	Lung Cancer	Brain Metastasis	Bone Metastasis	Liver Metastasis	Health	Death
Cold	-	-	-	22.65	-	-	-	-	-	-
Influenza	-	-	-	13.83	-	-	-	-	-	-
Asthma	-	-	-	22.83	-	-	-	-	-	-
Bronchitis	-	-	-	-	10.30	-	-	-	53.42	56.97
Lung Cancer	-	-	-	-	-	6.60	7.10	5.51	52.99	54.68
Brain Metastasis	-	-	-	-	-	-	-	-	48.09	42.48
Bone Metastasis	-	-	-	-	-	-	-	-	69.72	61.59
Liver Metastasis	-	-	-	-	-	-	-	-	62.33	47.88
Health	-	-	-	-	-	-	-	-	-	-
Death	-	-	-	-	-	-	-	-	-	-

After all that, the cost of every treatment was calculated in order to obtain a distribution that explained the cost for each disease treatment. The purpose was to use it as input data for the Monte Carlo simulation and then obtain the different treatment cost for every hypothetical patient that result from the simulation. In the following table it can be seen the average treatment cost for each disease, including the metastasis states. Please be noted the wide difference in cost between the cancer treatments and the other treatments, that phenomenon is easily explained by understanding that a cancer treatment will be always far more expensive as it involves much more expensive processes.

Table 7 Average treatment cost

Disease	Cold	Influenza	Asthma	Bronchitis	Lung Cancer	Brain Metastasis	Bone Metastasis	Liver Metastasis
Average Treatment Cost (NTD)	3,673.04	1,971.58	10,172.21	5,214.24	162,129.74	137,407.08	112,369.50	110,428.28

3.3.2. Semi-Markov simulation results

The calculation of the transition probabilities to be used in simulation was conducted while modelling the semi-Markov process. The time-dependent transition probabilities from one state to another state, which probability values depend on the length of time spent in current state were identified by analysing the transitions experienced by each patient in the previously selected population. Using the baseline model of semi-Markov process the transition probabilities were calculated using (3.7). This computation results on the probability rate for each possible transition over a period of 15 years (duration which NHI database are available).

The exact number of the probabilities rate is not shown in this study, due to huge matrix size. However, the following figures show the evolution of the probabilities over time. It can be observed that all the transition probabilities, except the ones regarding the common diseases and the transition probability between bronchitis and health state, are high in the beginning and decrease along the time with the longer a patient is in a state. This behaviour of the probabilities is matched to real clinical condition in which patients who experienced a recent change in their health state would have a high instability shortly after the transition. But, if they stay in a particular state for longer time, it indicates that the disease is stabilized, reflected by a decrease in probability rates. Moreover, to obtain probability rate when the duration of stay in a state is longer than 15 years during simulation, extrapolation method is used to estimate the value.

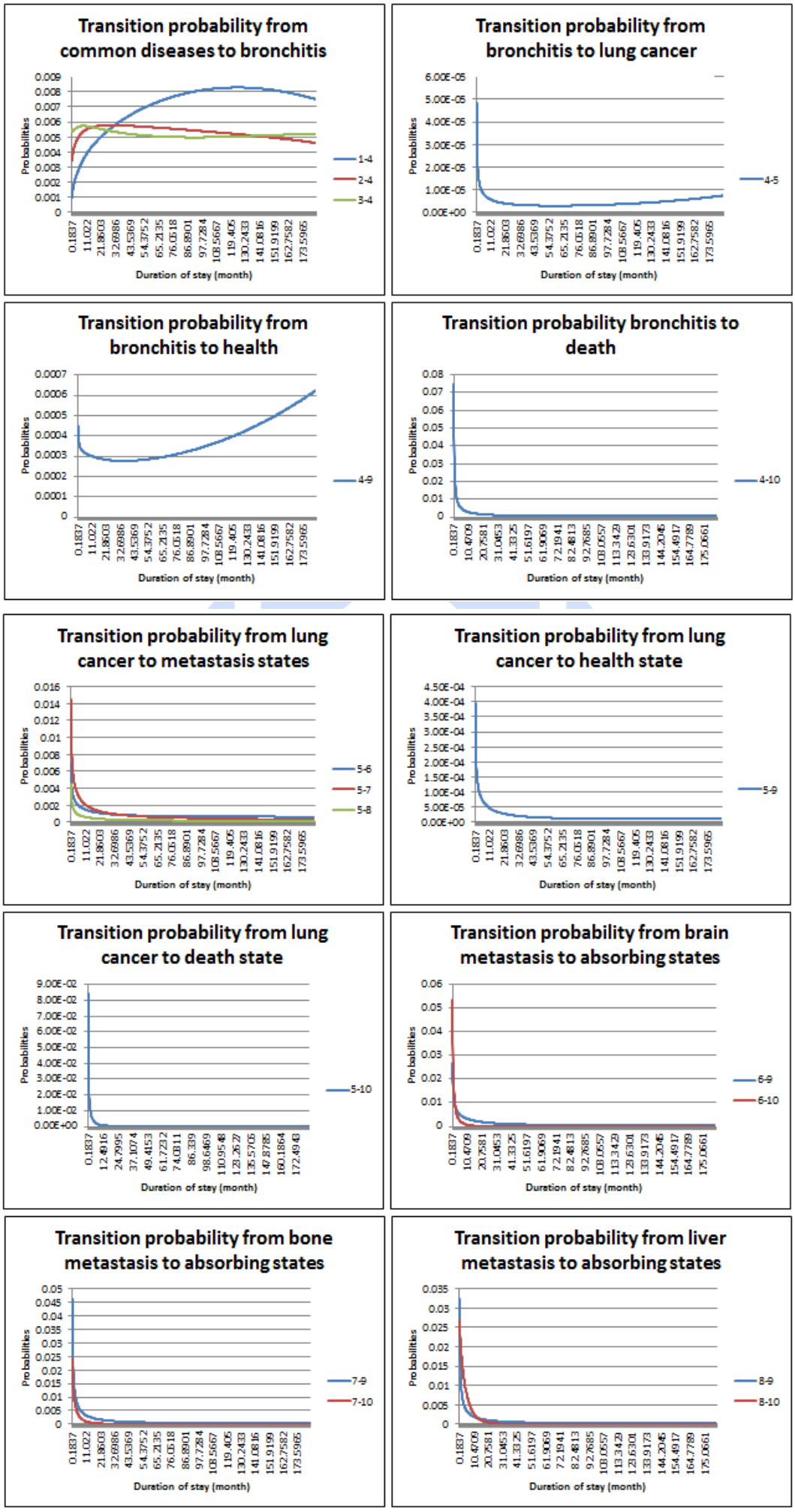

Figure 3 Transition probabilities between states

3.3.3. Semi-Markov Simulation

Two different simulations have been executed following the procedure described in Section 3.2. Meanwhile in this section the results of both simulations are explained in two different subsections.

3.3.3.1. Transient state results

In this simulation, one by one the patients in the hypothetical cohort have been simulated through the Semi-Markov model with an initial probability equivalent to the proportion of real patients that are diagnosed with one of these diseases. The simulation was done this way in order to obtain the results as realistic as possible, since it is a transient state and in most cases only one transition was experimented.

The following table shows the number of transitions among the different states. As it is a transient state and the duration of the simulation is only one year, it can be noted that the number of transitions in the whole simulation is very low.

Table 8 Number of transitions in the transient state

	Cold	Influenza	Asthma	Bronchitis	Lung Cancer	Brain Metastasis	Bone Metastasis	Liver Metastasis	Health	Death
Cold	-	-	-	259	-	-	-	-	-	-
Influenza	-	-	-	28	-	-	-	-	-	-
Asthma	-	-	-	172	-	-	-	-	-	-
Bronchitis	-	-	-	-	0	-	-	-	32	479
Lung Cancer	-	-	-	-	-	6	10	6	0	26
Brain Metastasis	-	-	-	-	-	-	-	-	0	0
Bone Metastasis	-	-	-	-	-	-	-	-	0	0
Liver Metastasis	-	-	-	-	-	-	-	-	0	1
Health	-	-	-	-	-	-	-	-	-	-
Death	-	-	-	-	-	-	-	-	-	-

As it can be seen on Table 8, transitions that contain more cases are the ones that connect the initial common diseases with bronchitis. This is explained by the fact that the initial probability used to run the simulation are taken from the proportion of transitions in the study population and there the proportion of patients diagnosed with these diseases is a lot higher than with the other diseases. Furthermore, since the length of the transition is only one year, the hypothetical patients can not suffer many transitions, that is why the initial diseases have a higher number of transitions. Again just remark that with 50,000 hypothetical patients the results give so few transitions, with transitions that have 0 cases, all this is explained by the reasons given above.

Table 9 shows the number of patients that have passed for each state also giving its percentage. It can be seen that this second table confirms what is pointed in the previous one.

Table 9 Number of patients that experiment each state in the transient state

	Cold	Influenza	Asthma	Bronchitis	Lung Cancer	Brain Metastasis	Bone Metastasis	Liver Metastasis	Health	Death
Number of patients	22016	1110	6932	19194	1207	6	10	6	32	505
Percentage (%)	44.032 %	2.220%	13.864 %	38.388%	2.414%	0.012%	0.020%	0.012%	0.064%	1.010 %

Again Table 9 shows that the initial diseases are the ones that have much more cases; the reasons that explain this phenomenon are the same as explained above in this section. It can be seen that the largely part of hypothetical patients experiment the initial diseases and for what can be seen in the previous table most of them do not pass to the next states. It has to be noted that each percentage showed in the table is the percentage of all patients and not the percentage of transitions, since a patient can experiment more than one transition and the purpose of the table is to show the number of patients in each state.

As it is said above in this report, a number of trackers and rewards were created in order to study the treatment time and cost. The following two tables, Table 10 and Table 11 show the results of those trackers and rewards.

Table 10 Average waiting time and lifetime in the transient state

Disease	Common diseases	Bronchitis	Lung Cancer	Metastasis	Lifetime
Average waiting time (months)	11.93	11.71	11.68	7.88	11.9143

As it can be observed in Table 10, the average time for every disease is almost 12 months excepting the metastasis cases. This second what is said above, as many patients do not change their state they spend the 12 months of the simulation in that state. Moreover, an average lifetime almost equal to 12 months indicates that a very small number of patients ends up in the health or death states before the simulation is finished, and almost all the patients that go to an absorbing state comes from a metastasis states, since those are the only states that have a lower average waiting time.

Table 11 Average treatment cost in the transient state

Disease	Common diseases	Bronchitis	Lung Cancer	Metastasis	Total cost
Average Treatment Cost (NTD)	15,171.66	10,772.95	804,513.69	175,624.98	37,178.3

The previous table (Table 11) present that the most expensive treatment is the lung cancer treatment followed by the metastasis treatment. The low amount observed in the average total cost says that there are much more hypothetical patients diagnosed with a common disease or with bronchitis than patients with lung cancer or metastasis. But since, as it is said above, there are too few transitions we cannot conclude that this is the average cost associated with this kind of patients.

3.3.3.2. Steady state results

In this simulation, one by one the patients in the hypothetical cohort have been simulated through the Semi-Markov randomly being imposed to one of the diseases set as initial states with equal proportion for each disease at the beginning of the simulation.

Since this simulation gives the results of the steady state and the theory of Semi-Markov says that every transition depends only on the current states and not on the previous events, at the end of the simulation, when all events have happened and the patients have arrived to the steady state, the results will not be influenced by the initial probabilities.

The Table 12 shows the number of transitions among the different states. As it is the steady state and the duration of the simulation is 20 years, it can be noted that the number of transitions in the whole simulation has considerably increased compared to the previous simulation.

Table 12 Number of transitions in the steady state

	Cold	Influenza	Asthma	Bronchitis	Lung Cancer	Brain Metastasis	Bone Metastasis	Liver Metastasis	Health	Death
Cold	-	-	-	4431	-	-	-	-	-	-
Influenza	-	-	-	3278	-	-	-	-	-	-
Asthma	-	-	-	3473	-	-	-	-	-	-
Bronchitis	-	-	-	-	0	-	-	-	510	1284
Lung Cancer	-	-	-	-	-	569	498	159	83	257
Brain Metastasis	-	-	-	-	-	-	-	-	29	2
Bone Metastasis	-	-	-	-	-	-	-	-	29	7
Liver Metastasis	-	-	-	-	-	-	-	-	3	7
Health	-	-	-	-	-	-	-	-	-	-
Death	-	-	-	-	-	-	-	-	-	-

As it can be seen on Table 12, transitions that contain more cases are the ones that connect the initial common diseases with bronchitis. This fact matches reality given that those diseases are chronic diseases very common in the population. Another aspect that stands out is the absence of transitions between bronchitis and lung cancer, this can suggest that there is not any relationship between these two diseases and in consequence, there is not relationship between the initial chronic diseases and lung cancer with metastasis. Moreover, it can be noted that regarding the two absorbing states there are also some singular facts. On one hand, coming from the bronchitis state more than two-thirds of patients move to the death state. On the other hand, coming

from the metastasis states there are much more patients that move to the health states than the ones that move to death state. Only lung cancer patients follow what the common sense would say, since more than 75% of them move to the death state. These unusual facts can be explained by two reasons, first of all it have to be understood that the transitions to the absorbing states depend on the hypothesis made at the beginning of the study, since the data doesn't offer accurate information about these two states even if the hypothesis were the more realistic which could be done, is impossible to strictly follow the reality. Moreover, in order to obtain time-dependent transition probabilities an approximation was done and this approximation can cause discrepancies between the simulation and the reality. All this is better analyzed in the next chapter (Chapter 4. Conclusions).

Table 13 shows the number of patients that have passed for each state also giving its percentage. It can be seen that this second table shows how most of the patients stay in the initial diseases, such as cold, influenza, asthma, bronchitis and lung cancer. This phenomenon can be caused by the fact that many hypothetical patients simulated stay all the cycles in the same state or experiment just one transition.

Table 13 Number of patients that experiment each state in the steady state

	Cold	Influenza	Asthma	Bronchitis	Lung Cancer	Brain Metastasis	Bone Metastasis	Liver Metastasis	Health	Death
Number of patients	10135	9875	10027	21070	10075	569	498	159	593	1541
Percentage (%)	20.270%	19.750%	20.054%	42.140%	20.150%	1.138%	0.996%	0.318%	1.186%	3.082%

It has to be noted that each percentage showed in the table is the percentage of all patients and not the percentage of transitions, since a patient can experiment more than one transition and the purpose of the table is to show the number of patients in each state.

Again a number of trackers and rewards were created in order to study the treatment time and cost. The following two tables show the results of those trackers and rewards.

Table 14 Average waiting time and lifetime in the steady state

Disease	Common diseases	Bronchitis	Lung Cancer	Metastasis	Lifetime
Average waiting time (months)	192.04	171.45	214.63	151.39	234.172

As it can be observed, the higher average time is the lung cancer time, which is more than 200 months, what means that most of the patients diagnosed with lung cancer in the simulation stay in the same state for all the simulation. Moreover, an average lifetime almost equal to 240 months indicates that a very small number of patients ends up in the health or death states before the simulation is finished, and almost all the patients that move to an absorbing state comes from a metastasis states, since those are the states that have lower average waiting time.

Table 2 Average treatment cost in the steady state

Disease	Common diseases	Bronchitis	Lung Cancer	Metastasis	Total cost
Average Treatment Cost (NTD)	268,100.50	176,520.71	14,746,987.18	6,099,358.76	3,376,466.557

The previous table present that the most expensive treatment is the lung cancer treatment followed by the metastasis treatment. The lower amount observed in the average total cost says that there are more hypothetical patients diagnosed with a common disease or with bronchitis than patients with lung cancer or metastasis. But since, as it is said above, there are too few transitions we cannot conclude that this is the average cost associated with this kind of patients.

CHAPTER 4 CONCLUSIONS

This chapter drew conclusions from the entire observed phenomenon in the experiments and their results, including all drawbacks and limitations. Furthermore, all the issues and their cause are analysed in order to interpret their reasons and implications.

This study presented a model that relates common chronic diseases with lung cancer, which is one of the leading causes of cancer death in Taiwan, and its possible metastasis; the treatment costs have also been analysed. Multi-state Semi-Markov model has been implemented to model the patient's health state and Taiwan's current healthcare system. Monte Carlo simulation then provides a technique to evaluate the Semi-Markov model and capture the patient's progression as well as the expected monetary costs yielded by patients. The study starts with the collection of the data from the NHIDB and its posterior gathering to form the study population. As it is explained in this report, the real data is used only to calculate the parameters to be introduced into the simulation and the results analysed come from a hypothetical cohort extracted from the Monte Carlo simulation. While in the real data there is a huge difference between the different diseases in terms of number of entries and quality of the data itself, making this kind of simulation allows us to study de model in a larger scale conferring the right number of patients to study the relationships that are given in it and studying it all as a set. Moreover, the cancer database couldn't be used because the model included other kinds of diseases and that made it too difficult to match the two different data banks. This fact caused that the final sample of cancer patients was smaller. Despite this, it was possible to build the model and run a simulation afterwards, proving the robustness of a semi-Markov model combined with a simulation method.

Two different simulations have been run and accordingly two different conclusions have been drawn. When it comes to the first simulation, it was run for a time horizon of 1 year in order to study the transient state. Its results are explained in section 3.3.3.1. Since it is a transient state simulation with a length of 1 year, there is no time for the patients to develop the different diseases of the model. In one year time most of the patients have remain in the initial disease state or have moved to the next state at most. This fact tells us that it takes more time to see if the patients develop the different diseases of the model in order to analyse their relationship. That is why, after this first simulation was done, another simulation that mirrors the steady state was needed.

Referring to the cost study results from the transient simulation, as it can be observed in the section 3.3.3.1, they give a basic insight of the amount spent in each treatment, showing that the most expensive costs are the cancer and metastasis cost, since they involve drugs and procedures much more expensive than what the other treatments entail.

Regarding the steady state simulation, results were far from expected. The most striking of these results is the absence of transitions between bronchitis state and lung cancer state, which suggests that after reach the steady state there is no evidence of a relation between these two diseases. In a previous research of background information about lung cancer a relationship with chronic bronchitis was found, as there are many cases where a lung cancer patient previously had bronchitis. All this can also be explained by the fact that since cigarette smoke often cause chronic bronchitis, a patient diagnosed with a bronchitis caused by cigarette smoke can later develop a lung cancer caused by the cigarette smoke too. Moreover, another issue that can be observed in the results is the small number of patients that move to different diseases and the fact that most of the patients that move to different states only move to the next one.

Furthermore, the results show a really small number of patients going to the death state, indicating that only few cancer patients die after 20 years, which goes against other researches done in the past. These discrepancies can be explained by two reasons. Firstly, since the NHIDB does not have accurate information of the health or death of the patients, a hypothesis was needed in order to set these two states and although the hypothesis was the most accurate that could be done, it can be drifted away from the reality. Secondly, another source of error can be the time dependant method used. Using a time dependant transition probability matrix for the simulation did not follow strictly Markov laws. The proper way to do it should be by a two-step Markov process: firstly, with the fixed probabilities decide which the next state was. Secondly, calculate the time. However, based on previous thesis and the available methods, it was decided to use this adapted method at the beginning of the study because it was considered to be the best tool considering the means which were available.

ACKNOWLEDGMENT

This study is based in part on data from the National Health Insurance Research Database provided by the Bureau of National Health Insurance, Department of Health and managed by National Health Research Institutes, R.O.C. (Taiwan). The interpretation and conclusions contained herein do not represent those of Bureau of National Health Insurance, Department of Health or National Health Research Institutes.

REFERENCES

Barbu, V. S., & Limnios, N. (2008). *Semi-Markov Chains and Hidden Semi-Markov Models towards Application.* New York: Springer Science+Business Media LLC.

Bonate, P. L. (2001). A Brief Introduction to Monte Carlo Simulation. *Clinical Pharmacokinetics*, 40(1): 15-22.

Chen, Y.-C., Yeh, H.-Y., Wu, J.-C., Haschler, I., Chen, T.-J., & Wetter, T. (2011). Taiwan's National Health Insurance Research Database Administrative Health Care Database as Study Object in Bibliometrics. *Scientometrics*, 86: 365-380.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2): 187-220.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., . . . Bray, F. (2013). *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. Retrieved from International Agency for Research on Cancer: http://globocan.iarc.fr, accessed on 6 July 2014.

Hsiao, F.-Y., Yang, C.-L., Huang, Y.-T., & Huang, W.-F. (2007). Using Taiwan's National Health Insurance Research Databases for Pharmacoepidemiology Research. *Journal of Food and Drug Analysis*, 15(2): 99-108.

Katsaliaki, K., & Mustafee, N. (2011). Applications of Simulation Within the Healthcare Context. *Journal of the Operational Research Society*, 62: 1431-1451.

Lipscomb, J., Weinstein, M. C., & Torrance, G. W. (1996). Time Preference. In M. Gold, J. Siegel, L. B. Russel, & M. C. Weinstein, *Cost-Effectiveness in Health and Medicine.* New York: Oxford.

Listwoń, A., & Saint-Pierre, P. (2013). SemiMarkov: An R Package fro Parametric Estimation in Multi-State Semi-Markov Models.

Listwon-Krol, A., & Saint-Pierre, P. (2014). *Package 'SemiMarkov' Reference Manual.* Retrieved from http://cran.r-project.org/web/packages/SemiMarkov/SemiMarkov.pdf

Listwoń-Krol, A., & Saint-Pierre, P. (2014). *SemiMarkov: Multi-States Semi-Markov Models.* R Package version 1.3.1, URL http://cran.r-project.org/web/packages/SemiMarkov/.

Mahadevan, S. (1997). Monte Carlo Simulation. In T. A. Cruse, *Reliability-Based Mechanical Design* (pp. 123-146). New York: CRC Press.

María Concepción Beuses Salcedo (2004). Primary malignant neoplasms of lung correlation anatomo clinic: study of twenty years of autopsy general hospital of south.

Michael A. Beckles, MB, BS; Stephen G. Spiro, MD; Gene L. Colice, MD, FCCP; Robin M. Rudd, MD (2003). Initial Evaluation of the Patient With Lung Cancer: Symptoms, Signs, Laboratory Tests, and Paraneoplastic Syndromes

National Health Insurance Administration. (2012). *2012 National Health Insurance Statistics report.* Taipei: National Health Insurance Administration.

Sheldon M. Ross (2007). *Introduction to Probability Models* (9th ed.). San Diego, CA: Academic Press.

Sonnenberg, F. A., & Beck, J. R. (1993). Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making*, 13: 322-338.

TreeAge Pro 2014 User's Manual

Weiss, G. H., & Zelen, M. (1965). A semi-Markov Model for Clinical Trials. *Journal of Applied Probability*, 2(2): 269-285.