國立臺灣大學公共衛生學院流行病學與預防醫學研究所

碩士論文

Graduate Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Master Thesis

佇列閾值及寇斯多相統計模型探討與大腸直腸癌早期發現和住院之相關時間分布

Queue, Hurdle, and Coxian Phase-type Model for Time Distributions Related to Early Detection and Hospitalization of Colorectal Cancer

任小萱

Grace Hsiao-Hsuan Jen

指導教授：陳秀熙　博士

Advisor：Tony Hsiu-Hsi Chen, Ph.D.

中華民國 104 年 5 月

May, 2015

誌謝

在研究所以前，我並未接觸過公共衛生及流行病學領域，很感謝流預所的老師們讓我有機會進入生物統計組來學習這方面的知識。在研究所學習的這兩年，在老師們細心且專業地指導下，讓我慢慢奠定生物統計的基礎以及培養研究的樂趣，此外我也漸漸體會到公共衛生對於國家社會的重要性，讓我覺得踏上這條路非常的值得。

在這兩年，非常感謝我的指導教授－陳秀熙老師，在老師身上總是可以學到很多，不管是專業上的知識或是生活上的經驗，老師都會毫無保留的與我們分享，讓我在這兩年過得非常充實、受益良多。這篇論文的完成，最感謝的就是陳老師，在論文研究及寫作的路上，老師寶貴的指導和建議，幫助我培養自己的思考能力，能與老師一起做研究是一件非常幸福的事。此外還要感謝嚴明芳老師、陳立昇老師、范靜媛老師以及邱月暇老師，老師們總是在我遇到困難時，給予及時的幫助；感謝邱瀚模醫師及李宜家醫師讓我從臨床的觀點學習到許多醫學上的知識；感謝我的口試委員張淑惠教授、林明薇教授以及丘政民教授給予我論文寶貴的意見。

在兩年的研究生活，特別感謝研究團隊的許辰陽博士、莊書琳博士、曹慧嫻博士以及蘇秋文學姐，教導我程式技術及學術知識；謝謝思敏幫了我許多忙，我們一起奮鬥、一起努力，有妳在就不孤單；謝謝芸婕總是給我力量，彼此打氣，有妳的陪伴生活充滿趣味；謝謝琮閔雖然我們都很忙碌，但還是可以互相分享生活點滴；謝謝同學和朋友們，偶爾一起放鬆心情，開心大笑後再繼續往前進。

最後，感謝我的爸媽及家人給予我的支持及鼓勵，提供我最好的環境來學習，聽我分享生活點滴並回饋給我寶貴的生活經驗，使我越來越強壯，有你們無私的奉獻與愛才有現在的我，我愛你們！

任小萱　謹誌于

臺灣大學流行病學與預防醫學研究所

中華民國 104 年 7 月

中文摘要

背景

台灣大腸直腸癌的發生率逐年增加，對於大腸直腸癌的早期發現可以先透過糞便潛血檢查再進一步地接受大腸鏡來進行確診，在確診為大腸直腸癌病人後，後續的住院治療這些都是不容忽視的問題。然而為了考慮民眾的篩檢到達率、未接受大腸鏡確診者的特性、等待接受大腸鏡確診的時間以及大腸直腸癌病人接受後續住院治療的住院天數，傳統的佇列模型是無法實行的。

研究目的

本論文的研究目的是將佇列過程、閾值模型以及寇斯多相模型整合為一個統計方法，並將此方法應用在分析台灣全國大腸直腸癌篩檢之陽性個案所需進行大腸鏡確診的等待時間以及大腸直腸癌病人的住院天數。

研究方法

閾值模型由邏輯斯迴歸以及截尾卜瓦松迴歸模型所組成，邏輯斯迴歸用來研究未接受大腸鏡確診者的特性，截尾卜瓦松模型則用來分析等待接受大腸鏡確診的時間分布。而寇斯多相模型可以探討等待時間的最佳隱藏階段，處理接受轉介民眾之間的異質性。為了可以更進一步地考慮民眾的篩檢到達率，我們結合了卜瓦松過程、閾值模型以及寇斯多相模型進而發展出一個佇列閾值寇斯多相模型。在住院治療方面，我們利用寇斯多相模型對 178 位大腸直腸癌病人的住院天數進行分析，探討其最佳的隱藏階段個數。

結果

第一部份：在篩檢前期（2004-2009 年），閾值模型的結果顯示女性、年齡較高者、居住在東部、離島或非都會區民眾、在醫院進行篩檢的民眾或是盛行篩檢個案（首

次參與篩檢）有較高的機率不接受後續轉介，而居住在中部或大都會地區、在衛生所或健康服務中心接受篩檢的民眾或是非首篩個案其所需等待接受大腸鏡確診的時間較短。

第二部份：在佇列閾值二階段寇斯多相階段模型中，等待大腸鏡確診的時間可被分類為等待時間較短階段以及等待時間較長階段，其結果顯示民眾的篩檢到達率每人天為 0.00021，不接受後續確診的機率為 0.26，一年大約有 15% 的民眾對於後續大腸鏡的確診會猶豫不決而陷入等待時間較長的階段。在等待時間較短階段的平均等待時間為 32 天而在等待時間較長階段的平均等待時間為 169 天。當我們將危險分數考慮到模型中進行分析時，佇列閾值二階段寇斯多相階段模型顯示低分群在等待時間較短階段的平均等待時間為 36 天而高分群為 30 天，在等待時間較長階段，兩群的平均等待時間皆為 167 天。

第三部份：在住院治療方面，我們利用三階段寇斯多相模型對 178 位大腸直腸癌病人的住院天數進行分析，住院天數可被分為短期停留階段、中期停留階段及長期停留階段。在短期停留階段中，平均住院天數為 10 天，而中期停留階段及長期停留階段的平均住院天數均為 49 天。當我們將性別放入模型中考量時，可利用二階段寇斯多相模型對住院天數進行分析，住院天數可被分為短期停留階段及長期停留階段。二階段寇斯多相模型的結果顯示男性會比女性較早出院或死亡。若將年齡放入模型中考量時，年長者相對於年輕的病人較早出院或死亡。

結論

這是一個新的佇列閾值寇斯多相模型，它被用來解決佇列過程、陽性個案不接受後續轉介的閾值問題以及針對等待接受大腸鏡轉介的時間和大腸直腸癌病人接受後續住院治療的住院天數來探討其最佳的隱藏階段數。

關鍵字：寇斯多相模型、閾值模型、等候時間、大腸直腸癌、大腸直腸癌篩檢

Abstracts

Background

As the incidence rate of colorectal cancer (CRC) has been increasing in Taiwan, early detection of CRC through fecal immunochemical test (FIT) screening first and then colonoscopy examination and hospitalization of CRCs cannot be overemphasized. However, the arrival rate of screenees, the non-compliers of undergoing colonoscopy, the waiting time (WT) for undergoing colonoscopy, and the length of stay (LOS) for CRCs has rendered the conventional queue model infeasible.

Aims

The objective was to integrate the queue process, hurdle model, and Coxian phase-type model into a unifying framework that was applied to two empirical datasets, one relating to the WT of undergoing colonoscopy from Taiwanese nationwide screening program, and the other pertaining to the LOS on hospitalized CRCs enrolled from one medical centre.

Methods

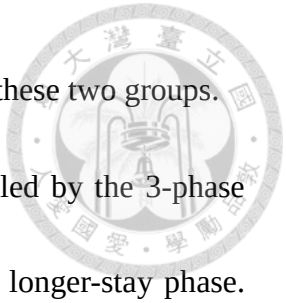
The hurdle model was developed in combination with a mixture of the logistic regression model that dealt with the non-compliance part and the truncated Poisson regression model pertaining to the WT distribution. The Coxian phase-type was further developed to identify the optimal hidden phase of WT. To further consider the arrival

rate of screenees, we developed the queue hurdle Coxian phase-type model which is the combination of the Poisson process, hurdle model and Coxian phase-type model. Data on the LOS of 178 CRCs were modelled by the Coxian phase-type model to identify the optimal number of hidden phases.

Results

Part I : From 2004 to 2009, the results of the hurdle model indicate the factors associated with non-compliance for colonoscopy included female, older age group, eastern Taiwan or offshore islands area, rural area, hospital screening unit and prevalent screening rounds, and the factors associated with shorter WT for colonoscopy included middle Taiwan area, main urban area, public health centers screening unit and subsequent screening rounds.

Part II : The queue hurdle 2-phase Coxian phase-type model was classified as short- and long waiting phase. The arrival rate was 0.00021 per person-days and the probability of non-compliance with colonoscopy was 0.26. Annually, around 15% subjects were so hesitant to be referred to undergo colonoscopy that they were trapped in long waiting phase. The mean WT of short waiting phase and long waiting phase were 32 days and 169 days, respectively. Further to consider the effect of risk score on the model, the queue hurdle 2-phase Coxian phase-type model indicates the mean WT in short waiting phase were 36 days and 30 days for the low score group and the high



score group, separately and 167 days in longer waiting phase among these two groups.

Part III : For hospitalization, the LOS with 178 CRCs was modelled by the 3-phase Coxian phase-type model classified as short-stay, medium-stay and longer-stay phase. In the short-stay phase, the expected LOS was 10 days whereas both the medium- and longer-stay phases were 49 days. When gender was taken into account, the LOS was modelled as a 2-phase Coxian phase-type model, short- and long-stay care. It shows that male would discharge or die earlier than female. Regarding age, it shows the elderly would discharge or die earlier than the young.

Conclusions

A new queue hurdle Coxian phase-type model was developed to solve the queue process, the hurdle issue in relation to the problem of non-compliance with the referral of positive results of screenees to have confirmatory diagnosis, and to identify hidden phases during the WT for undergoing colonoscopy among the referrals and LOS in hospitalization for the treated CRCs.

Keywords : Coxian phase-type, the hurdle model, waiting time, colorectal cancer, colorectal cancer screening

Contents

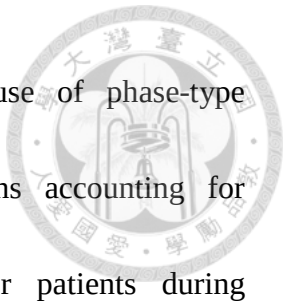
Figure

Table

Chapter 1 Introduction

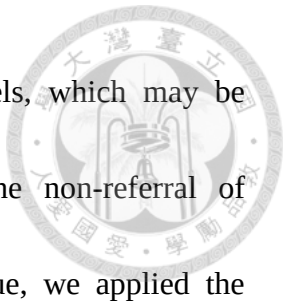
While population-based service screening for colorectal cancer (CRC) with fecal immunological test (FIT) has been demonstrated in reducing mortality in several previous studies[1][2], a concern is raised as to whether the clinical capacity of colonoscopic examination is sufficient enough to meet enormous burden of referrals with positive result of FIT. The waiting time (WT) for undergoing confirmatory diagnosis would be longer if the capacity is not sufficient and vice versa. Although an organized service screening program has been scheduled by the pre-determined referral date, the WT to undergo colonoscopy is still subject to how the referral system with colonoscopy after screening can be offered. It is therefore interesting to get a better understanding of the distribution of WT for undergoing colonoscopy for each organized service screening. The previous study in Canada has shown the average total WT was around 7.5 months[3]. However, few studies have been conducted to address whether relevant postulated factors such as demographic features, type of institution, geographic areas, calendar period, and prevalent screen or subsequent screen affect the WT.

After early detection of CRC, it is also very interesting to note that the length of stay (LOS) for hospitalization among patients diagnosed as CRC would become heterogeneous with LOS for early detected and late detected CRCs.



Motivated by the empirical data mentioned above, the use of phase-type distribution may be justified. The phase-type time distributions accounting for multi-phase transitions such as short LOS to long LOS for patients during hospitalization have been used to get a better understanding of the underlying dynamic hidden phases. These thoughts have been executed by the use of Coxain phase-type model (Marshall et al[4]) to estimate LOS for patients hospitalization. It is well acknowledged that the application of Coxian phase-type is very flexible. For example, the Coxian phase-type distribution may be used to other scenario such as WT for undergoing colonoscopy while a mass screening for CRC is conducted.

Although the Coxian phase-type model has been used in the queue process, its application to population-based screening may need to be modified on the ground of two major reasons. First, the queue process applied to clinical setting is based on individual-based rather than population-based process. How to connect the arrival process among those who have the uptake of population-based screening for CRC with the WT distribution for colonoscopy is the first consideration. In the queue process, the Coxian phase-type model is a specialized case of hyper-exponential queue model. It is natural to consider whether it can be used for hypo-exponential as the referral of participants with positive test may suffer from the problem of non-compliance. From the methodological viewpoint, how to amend the Coxian phase-type model to



accommodate both hyper-exponential and hypo-exponential models, which may be adequate for modelling data on WT making allowance for the non-referral of undergoing colonoscopy, is a challenging task. To solve this issue, we applied the concept of hurdle model with the Poisson regression model to solve the problem of non-compliance while the WT distribution is considered simultaneously. We therefore integrate the queue process, the hurdle model, and the Coxian phase-type model as a unifying model for modelling the queue for colonoscopy and hospitalization of CRC.

As the WT is regarded as time to event, the first part of purpose of my thesis was to develop by combining the hurdle model, the queue process, the Coxian phase-type model as a unified framework to estimate the median and the percentile of WT and further to assess whether the relevant factors are associated with the WT for undergoing colonoscopy.

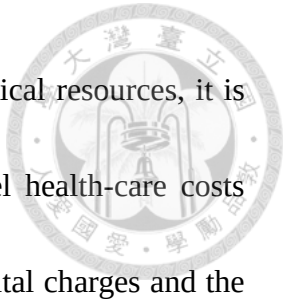
The second part of my thesis was to the application of the Coxian phase-type model to elucidate the hidden phase of LOS for hospitalization among patients diagnosed as CRC.

Chapter 2 Literature Review

2.1 Evolution of Coxian phase-type distribution

Over the past few decades, Coxian phase-type distributions have been gradually used to model the skewed survival data. The typical example was to apply the Coxian phase-type distribution to modelling hospital LOS of patients and the patient WT in Accident and Emergency Departments[5] because the proportion of the elderly population tremendously increased recently, leading to enormous medical expenditures attributed to hospital treatment. Therefore, Marshall[4] developed the Coxian Phase-type Cost Model (CP-CM) in 2007 to evaluate how to allocate the limited medical resources and costs.

A patient's LOS is considered a reliable indicator for measuring the quantity of resources and has a direct impact on the medical expenditures. In this paper, they introduced some previous methods analyzing patient' LOS: Mean LOS. LOS data are positively skewed, indicating that the majority had a short stay in the hospital whereas few patients had a long stay. If we use mean LOS to estimate LOS, it is less reliable and inaccurate. To tackle this property, the compartment models with 2 or 3 stages or with the Coxian phase-type distribution were used to consider the positive skewness and heterogeneity of LOS.



To discuss the problem of medical costs under the limited medical resources, it is still worthwhile to review several previous methods used to model health-care costs including (1) two regression models that was to estimate mean hospital charges and the other was to estimate the ratio of the average charge per day; (2) Poisson mixture distributions that considered the heterogeneity of patient populations; (3) queuing theory that used the queueing theory to model patient LOS, and then determined the optimal allocation of hospital resources and costs; (4) survival analysis to estimate the medical care costs by using survival analysis, where patient cost data were recognized as survival times with the Kaplan-Meier estimation and the Cox regression model could be considered;(5) 2-state compartment model that represents the acute care and long-stay care separately. Patients in the same compartment had similar characteristics, but they had dissimilar characteristics and different costs in the distinct compartment.

The recently proposed Coxian phase-type distribution could be interpreted as distinct clinical stages of patients in hospital interpreted by the clinical experts. It is natural to extend the Coxian phase-type distribution to the CP-CM enabling the expected expenditures to be estimated. This new model can solve many problems encountered in the previous methods. The following is pertaining to why the CP-CM is thought of as an appropriate model. Firstly, if we use regression analysis, we need normality and equal variance assumption. However, as the LOS data have the skewed

property and heterogeneity, albeit we can take logarithms of LOS to follow normality assumption, regression analysis still cannot be applicable for coping with the heterogeneity. Secondly, if we use the Poisson mixture distribution, we cannot estimate the transition rate from multi-state phase-type transitions. It will be also subject to over-dispersion. Thirdly, although survival analysis is appropriate for censored data and used for a variety of distributional forms, cost estimates may be biased if survival exceeds the maximal censoring time. Therefore, it is justified to use the CP-CM to overcome these situations.

2.2 Model structure of the Coxian phase-type distribution

The Coxian phase-type distribution describes the time to absorption of a finite Markov chain in continuous time over the phases {1,2,…,k,k+1}. This Markov chain has one absorbing phase (k+1th) and k transient phases (1,…,k). The process only starts in the first transient phase. If the transition is within transient phase, the parameter of its transition rate is denoted as λ_i. If the process is from transient phase to absorbing phase, the parameter of its transition rate is symbolized as μ_i. Therefore, given in the ith phase at time t, the probability of patient in the i+1th phase after a short time (Δt) can be expressed as

$$P(X(t+\Delta t)=i+1 \mid X(t)=i)=\lambda_i \Delta t+o(\Delta t), \text{for } i=1,\ldots,k-1$$

Given in the ith phase at time t, the probability of patient in the absorbing phase after a short time (Δt) can be expressed as

$$P(X(t+\Delta t)=k+1 \mid X(t)=i)=\mu_i \Delta t+o(\Delta t), \text{for } i=1,\ldots,k$$

The probability density function (pdf) of the random time variable T, representing the time until absorption, is given by

$$f(t)=\mathbf{p}\exp(\mathbf{Q}t)\mathbf{q} \tag{2-1}$$

$$\mathbf{p}=(1\ 0\ 0\ \cdots\ 0)$$

$$\mathbf{Q}=\begin{bmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -\mu_k \end{bmatrix}$$

$$\mathbf{q}=(\mu_1\ \mu_2\ \cdots\ \mu_k)^T$$

It comprises the probability defining initial transient phases (**p**), transition rates restricted to the transient phases (**Q**) and transition rates from transient phases to the absorbing phase (**q**).

As mentioned above, due to the different costs of distinct treatment and stages of health-care, the CP-CM was developed to model patient's medical costs and derive the expected total cost from the moment generating function (MGF). It is assumed that the system has been running long enough to reach steady state and that each phase of the system is operating at maximum capacity. Some random variables are defined. It could be divided into three categories:

(1) Time. The random variable T_i is defined the length of time spent in phase i, and follows exponential distribution with $\lambda_i + \mu_i$. Then the MGF of T_i is

$$M_{T_i}(\theta) = \int_0^{\infty} \exp(\theta t) f_i(t) dt = \frac{\lambda_i + \mu_i}{\lambda_i + \mu_i - \theta} \quad (2\text{-}2)$$

(2) Cost. The cost rates are time homogeneous but phase dependent. So if assuming the cost per subject per time unit z in phase i is c_{iz}, it becomes c_i. Then, it defined D_{ij} as the total cost per subject that leaves phase j of the system, given it stayed in phase i. It can easily know that D_{ij} has a linear relationship with T, so the MGF of D_{ij} is

$$D_{ij} = \sum_{\psi=i}^{j} c_{\psi} T_{\psi}$$

$$M_{D_{ij}}(x) = \prod_{\psi=i}^{j} \frac{\lambda_\psi + \mu_\psi}{\lambda_\psi + \mu_\psi - x c_\psi} \quad (2\text{-}3)$$

(3) The number of subjects. They defined Z_{ij} as the number of subjects who leave the system from phase j, given that they started in phase i. It could be disassembled into the number of subjects initially in phase i multiplied by the probability of leaving the system from phase j, given they started in phase i. Therefore, it follows multinomial distribution, and the MGF is

$$Z_{ij} = n_i \times p_{ij}$$

$$M_{z_{ij}, z_{ij+1}, \ldots, z_{ik}}(t_{ij}, t_{ij+1}, \ldots, t_{ik}) = \prod_{i=1}^{k} \left(\sum_{j=i}^{k} p_{ij} \exp(t_{ij}) \right)^{n_i} \quad (2\text{-}4)$$

$$p_{ij} = \frac{\left(\prod_{\gamma=i}^{j-1} \lambda_\gamma \right) \mu_j}{\left(\prod_{\gamma=i}^{j} \lambda_\gamma + \mu_\gamma \right)} \text{ if } j \neq k$$

$$p_{ij} = \frac{\left(\prod_{\gamma=i}^{j-1} \lambda_\gamma\right)}{\left(\prod_{\gamma=i}^{j-1} \lambda_\gamma + \mu_\gamma\right)} \text{ if } j = k$$

Finally, if we want to get the total cost defined as T_N for all patients while in the system, we can find that it has a linear relationship with Z_{ij} and D_{ij}. Therefore, we can figure out its MGF and get the expected future cost for all subjects in the system.

$$T_N = \sum_{i=1}^{k} \sum_{j=i}^{k} Z_{ij} D_{ij}$$

$$M_{T_N}(x) = \prod_{i=1}^{k} \left(\sum_{j=i}^{k} p_{ij} M_{D_{ij}}(x)\right)^{n_i} \qquad (2\text{-}5)$$

Marshall used the CP-CM model to model patients' costs and calculated the expected cost of patients in hospital. They found that it is an appropriate model and can provide some useful information to clinicians or hospital managers as their future decisions.

2.3 Semi- and Hidden Markov Process

Continuous-time multistate models are widely used in the natural history of chronic diseases. But if we only can observe the process at discrete time points, we have no information about the times or types of events between observation times. The inference becomes difficult. To overcome this issue, the Markov assumption has been made to imply that the sojourn time in these disease states follows exponential distribution which possess the memoryless property, so that it can limit the transition

rates between these states no longer depend on time since entry into the current state. However, actually the transition intensities of the process often depend on time since entry into a state that calls semi-Markov process. Therefore, the study conducted by Titman[6] provides an alternative to alleviate this problem by developing an approach that used the phase-type sojourn time distribution to fit semi-Markov models with panel-observed data. In addition, the approach was extended to data where the observed states were subject to measurement errors.

Panel-observed data are that the observation time periods of each measurement are identical for the same patient. Therefore, given the certain observation time, we can observe types of disease states. It no longer needs Markov assumption. Therefore, the panel-observed data can make the inference become easier.

There were several previous studies which also proposed different ways to fit semi-Markov process: (1) If the observation scheme is sufficiently frequent, the likelihood for a semi-Markov model can be expressed easier. All transitions can be observed, although transition times are interval censored. If the process is a panel data, multiple transitions may occur between observations and we need to use multidimensional integral to obtain the likelihood, which becomes very complicated. (2) When it comes to multiple transitions mentioned in the previous study, the likelihood function would become complex. If it is a progressive model that means

there is only one possible path of transitions and cannot reverse, computation of the integral may be feasible as the model has a small number of states such as 3 or 4. (3) Nonparametric estimation is possible via self-consistent estimators in progressive model. (4) Progressive model can be fitted semiparametrically with penalized likelihood. (5) Taking two-state recurrent model for example, as it allows reverse transitions that means it can return to the original state, computation of the likelihood will become more intractable. Regarding evenly spaced observation, a minimum chi-square estimation approach can be used to overcome the problem for this model. (6) Stopping-time resampling has been proposed as a simulation based method of computation. (7) If at least one state in the model has the Markov property, the inference for the panel observed semi-Markov models will be much easier. Because of Markov property, the likelihood for an individual can be factorized into sojourn times of departure from the Markov state. (8) In a two state recurrent disease process with panel observed data, they assumed the existence of latent process was a time homogeneous birth death process and its state space was {0,1,2,…}. If a subject was in state 0, he/she would be considered to be disease free. Other stated were considered ill. Therefore, sojourns in the observable illness state are not exponential and the observable process was a semi-Markov process. However, the computation might become straightforward, if the latent Markov structure of the model allowed the likelihood to be expressed as a hidden

Markov model (HMM).

In many clinical studies, the x_i may be regarded as the measurements of a biomarker or screening test. These measurements may have measurement error so that there is a nonzero probability that the state is misclassified. Instead of observing the x_i directly, we observe $o_1, \ldots, o_n$. The misclassification probabilities are defined as

$$P\{O(t) = s \mid X(t) = r\} = e_{rs}. \quad (2\text{-}6)$$

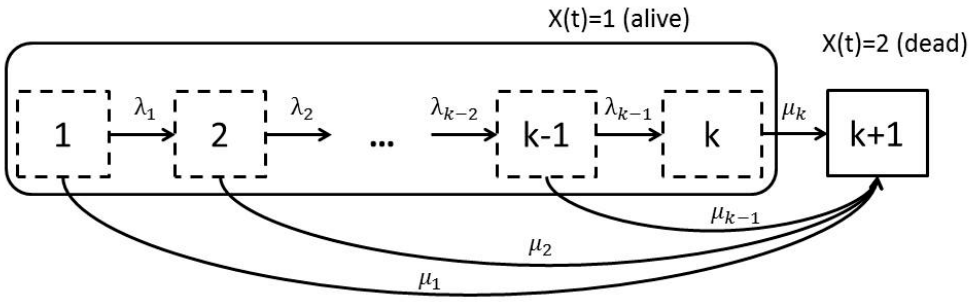
That means at time t, it is exactly in state r, but we observe it is in state s. Based on the misclassification probabilities, e_{rs} remains constant through time and X(t) is a Markov process, so we know that conditional on the true underlying states, the observed states are independent and the o_i can be modeled by a HMM. To present the likelihood contribution of misclassification for an individual, each transition depends on the complete history of the process. So for each individual, the matrices were constructed as $M_1, \ldots, M_n$, and M_i is an $R \times R$ matrix with (r,s) entry $p_{rs}(t_{i-1}, t_i) \times e_{s,o_i}$ with $t_0 = 0$. It presents the misclassification probability that a subject is in state r at time i-1 and actually reaches in state s at time i, but is misclassified in state o_i. Then, the likelihood contribution for an individual can be written as

$$L = \pi \mathbf{M_1 M_2} \ldots \mathbf{M_n 1} \quad (2\text{-}7)$$

where π presents the vector of initial state probabilities and $\mathbf{1}$ presents a vector of ones of length R. Covariates affecting the transition rates can be modeled by

$\mu_{rs}(t;\mathbf{y}) = \mu_{rs}(t)\exp(\beta_{rs}^{T}\mathbf{y})$, where $\mathbf{y}$ is a vector of explanatory variables. Covariate effects may also be incorporated into the matrix of misclassification probability by assuming linearity on a logit scale $\text{logit}(e_{rs}) = \alpha_{rs}^{T}\mathbf{y}$.

To describe a Coxian phase-type distribution, they gave a simple two state (alive,dead) survival model for example, demonstrating how a Coxian phase-type distribution could be applied to the sojourn time distribution of each transient state of a general, multistate, semi-Markov model. Consider a two state survival model X(t) with state {1=alive,2=dead}, for which the transition intensity from alive to dead is time inhomogeneous. For a Coxian phase-type model, the sojourn time in the transient state is assumed to be governed by a latent Markov process X*(t) with k transient phases and one absorbing phase k+1 (=dead). The latent process is progressive, so the movement from transient phase $j \in \{1, \ldots, k\}$ is either to the adjacent phase j+1 or to the absorbing state k+1 as below.



The solid line frame presents the observed state X(t) that we can only observe a subject is either alive or dead. The dashed line frame means the latent state X*(t) that

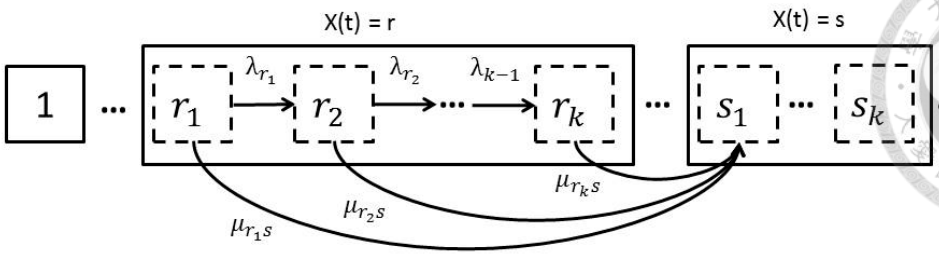
we cannot observe. At time zero, the process is in phase 1. There are two types of parameters. One is $(\lambda_1, \dots, \lambda_{k-1})$, the transition intensities between transient phases and the other is $(\mu_1, \dots, \mu_k)$, the transition intensities from the transient phase to the absorbing state. These parameters are constant with time, but intensities are different between phases. It induces time inhomogeneity in the movement between the observable states (from alive to dead).

Consider a semi-Markov process X(t) with state space S={1,…,R}, where R is an absorbing state, and t represents time from entry into the initial state. For each of the observable states $r \in S$ we assume there exists a latent process X*(t) with states $r_1, \dots, r_k$ but we observe only that the subject is in state r. The state space **S*** of latent process X*(t) are

$$\mathbf{S}^* = \{1_1, 1_2, \dots, 1_k\} \cup \{2_1, 2_2, \dots, 2_k\} \cup \cdots \cup \{(R-1)_1, (R-1)_2, \dots, (R-1)_k\} \cup R \ ,$$

its dimension is {k(R-1)+1}. In each observable state, it is not necessary to have the same number of latent states.

The sojourn distribution of each nonabsorbing state r of X(t) is assumed to be a k-phase Coxian phase-type distribution, with parameters $\lambda_{r_1}, \dots, \lambda_{r_{k-1}}$, the rates for movement between phases of state r and $\mu_{r_1 s}, \dots, \mu_{r_{k-1} s}$, the rates for movement out of state r to state s as follows.



The likelihood can be expressed as (2-7), where for an individual the matrix $\mathbf{M_i}$ become {k(R−1) + 1} × {k(R−1) + 1} with (r, s) entry $e_{s,x_i} p_{rs}(t_{(i-1)}, t_i)$, for s ∈ S*. If s is a phase of the observed state x_i, then $e_{s,x_i} = P\{X(t) = x_i | X^*(t) = s\}$ takes the value 1 and 0 otherwise.

To incorporate misclassification error, the process is extended to the hidden semi-Markov model (HSMM). The details of the framework refers to Titman et al[6].

Suppose the misclassification probability matrix is **e** as (2-6) and each state in X(t) is phase-type distribution. If the latent process $X^*(t) \in \{r_1, \dots, r_k\}$ then X(t)=r for r=1,…,R. So the misclassification probability

$$e^*_{r_j s} = P\big(O(t) = s \big| X^*(t) = r_j\big) = P(O(t) = s | X(t) = r) = e_{rs} \ , \qquad (2\text{-}8)$$

for r,s=1,…,R and j=1,…,k. We can find that they are independent of j. Therefore, the latent Markov process, X*(t), defines X(t) deterministically and O(t) | X(t) is multinomial.

The likelihood contribution from an individual can be calculated as

$$L = \pi \mathbf{M_1^*} \mathbf{M_2^*} \dots \mathbf{M_n^*} \mathbf{1}, \qquad (2\text{-}9)$$

where M_i^* is a {k(R−1) + 1} × {k(R−1) + 1} matrix with (r*, s*) entry

$e^{*}_{s^{*},o_i} p_{r^{*}s^{*}}(t_{(i-1)}, t_i)$, for r*, s* ∈ S*. The difference between the HSMM and semi-Markov model is that the e_{s,x_i} in the semi-Markov case is either 0 or 1, but in the hidden semi-Markov case, the e_{rs} may lie between 0 and 1 and can be treated as unknown parameters.

To explore the development of bronchiolitis obliterans syndrome (BOS) in post-lung-transplantation patients, they used the HMM and the HSMM to fit the data to identify which model was better. It shows the HMM might be the lack of time homogeneity, so the HSMM could provide a better fit to the data using the phase-type methodology. Through these methods they were able to better characterize the natural history of lung function decline after thoracic transplantation.

Chapter 3 Data

I. Compliance with colonoscopy from positive FIT of Taiwan nationwide colorectal cancer screening program

The Taiwan Nationwide CRC screening by FIT is offered for subjects aged 50 to 69 years. The main purpose was to reduce mortality from CRC through early detection. Those who had fecal hemoglobin concentration (f-Hb) higher than the cutoff of 20 μg Hb/g of feces were considered as positive and were referred for confirmatory tests by colonoscopy.

All the subjects who had ever attended this nationwide screening program during the period from 2004 to 2013 with positive FIT were enrolled in this study. Those who had f-Hb concentration less than 20 μg Hb/g of feces but with family history were also considered as positive test in this study. Those who underwent screening at unknown place, receiving unknown brand to evaluate test characteristics, or having missing f-Hb value were excluded from the following analysis. A total of 4,978,350 subjects attended CRC screening and 316,864 of them had positive test.

Study variables and definition

Baseline characteristics include gender, age at screening, geographic areas, type of screening units, urbanization levels, calendar periods, and round of screening. Subjects

who were detected as positive case for first-time screening were defined as ‘prevalent screen’ and those who were detected for later screening rounds were defined as ‘subsequent screen’. Besides, calendar periods were divided into two periods. In the inaugural 5 years (2004-2009) of the screening program, screening service was mainly offered at the public health center. From 2010 onward, hospitals and local clinics actively invited people for screening. It was divided into two periods, inaugural period and rolling out period, respectively.

Positive Rate, Referral Rate and Median Waiting Time for Colonoscopy

Total positive rate was 6.36% and positive rates of the corresponding characteristics are shown in Table 3-1 and Figure 3-1. Female and those aged 60 years and younger or screening hold at public health centers and during the period of 2004-2009 had a lower positive rate. The difference of positive rates among geographic areas, urbanization levels and rounds of screening were pretty small. The trend of number of attendees and positive rates are presented in Figure 3-2. A soaring trend of attendees and positive rates were noted at year 2009 and 2010. Chronological changes of positive FIT rate, referral rate, and colonoscopy WT are shown in Table 3-2.

Here, we only considered subjects who underwent colonoscopy within 6 months after being detected as positive case and they would be regarded as ‘successful’ referral.

The referral rate and median WT for colonoscopy among subjects with positive FIT are listed in Table 3-3. Overall referral rate and median WT were 72.78% and 28 days during the inaugural period and those for the rolling out period was 59.42% and 46 days. During the inaugural period, referral rate within age groups and urbanization levels shows a small difference. Male had slightly higher rate than female. Besides, attendees lived in middle Taiwan, underwent screening at public health centers, or was detected at subsequent screen had higher referral rate, and those lived in eastern Taiwan or offshore islands, underwent screening at hospitals, or was detected at prevalent screen had lower referral rate. In rolling out period, within gender groups, the difference in referral rates became small. Those who aged 65-69, underwent at local clinics, or lived at rural area had lower referral rates and northern Taiwanese had higher rate. The rounds of screening had the same drift in these two periods. Both illustrate subsequent screen had higher referral rate. Subjects in rolling out period had lower referral rate and were needed to wait longer time for colonoscopy especially those undergoing screening at local clinics with the WT of 92 days. However, a special discovery was that screening at hospital during rolling out period would reduce WT for colonoscopy from 51 days to 42days. The time trend of referral rate, medium and third quantile of WT are given in Figure 3-3. A decrease in referral rate and increase in WT was observed during the year from 2009 to 2010.

II. Hospitalization of colorectal cancer patients in Shin Kong Wu Ho-su Memorial Hospital

Based on the computerized information system of Shin Kong Wu Ho-su Memorial Hospital (SKH) between 1999 and 2013, patients who had received hospital treatment and whose International Classification of Diseases (ICD) was recorded as 153 or 154 were enrolled as our study population. There were 178 CRC patients who had ever been hospitalized in SKH between 1999 and 2013.

Study variables and definition

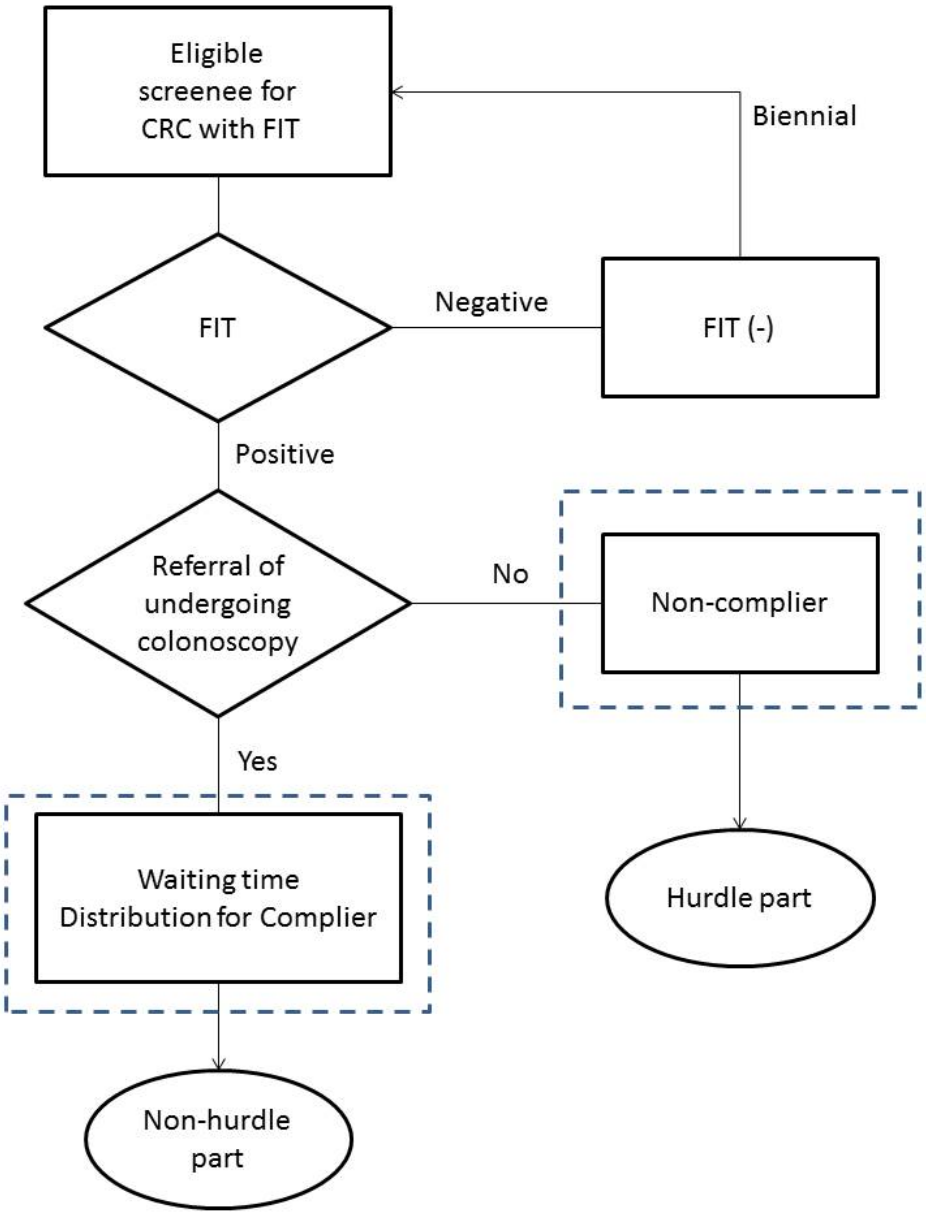
The variables of interest included the patients' LOS (recorded in days), measured from the day of admission of a patient until they have been discharged. There are six discharge types and their distributions are shown in Table 3-4. The LOS ranges from 1 to 215 days, with a mean of 13.8 days and a median of 7 days.

Baseline characteristics included gender and age at hospitalization. We divided age into those aged below 60, aged between 60 and 74, and aged above 75.

Chapter 4 Methodology

4.1 The hurdle model

Analytical framework of WT for colonoscopy among positive-FIT screenees with the statistical hurdle model is delineated in the following Figure.



To accommodate the non-referral of undergoing colonoscopy (non-complier) and also WT for undergoing colonoscopy among the compliers (see Figure above), we proposed the hurdle model to consider both the non-complier and the WT distribution for the complier. The hurdle part is the application of logistic regression model to identify factors affecting non-compliance with colonoscopy and the progressive part is modelled by the truncated Poisson regression model given the count greater than one to identify factors affecting WT for undergoing colonoscopy.

In the hurdle model assuming there are G subsets determined by relevant covariates (such as age, gender, and so on), $y_{ij} = 0$ representing the j-th screenee of subset i did not undergo colonoscopy and $y_{ij} = 1$ represents the j-th screenee of subset i had underwent colonoscopy for j=1,…,N_i, therefore $y_i = \sum_j y_{ij}$ is the number of screenees required for undergoing colonoscopy in subset i and the total number of screenee in subset i is N_i for i=1,…,G. t_{ij} is time to undergo colonoscopy of the j-th screenee in subset i, therefore $t_i = \sum_j t_{ij}$ represents total WT in subset i. p_i is the probability of refusing to undergo colonoscopy (non-complier) estimated with the logistic regression model, and λ_i is the mean arrival rate of receiving colonoscopy estimated with the truncated Poisson regression model which is conditional on at least one screenee undergoing colonoscopy. The hurdle model distribution can be expressed as

$$f(y_i|\mathbf{x}_i) = \begin{cases} p_i^{N_i - y_i} & , \text{ nonreferral} \\ (1-p_i)^{y_i} \frac{e^{-\lambda_i t_i}(\lambda_i t_i)^{y_i}}{y_i!(1-\exp(-\lambda_i t_i))} & , \text{referral} \end{cases} \quad (4\text{-}1)$$

where $0 \le p_i \le 1$; $\lambda_i, t_i > 0$; $N_i, y_i \ge 0$, x_i represents relevant covariates.

The hurdle regression model

The effect of relevant covariates on the non-complier was modelled by using the following logistic regression model

$$\log\left(\frac{p_i}{1-p_i}\right) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \cdots + \gamma_k x_{ik} \quad (4\text{-}2)$$

where $\gamma_j's$ are regression coefficients corresponding to covariates $x_{ij}'s$, and k is the number of covariates considered in each of the model.

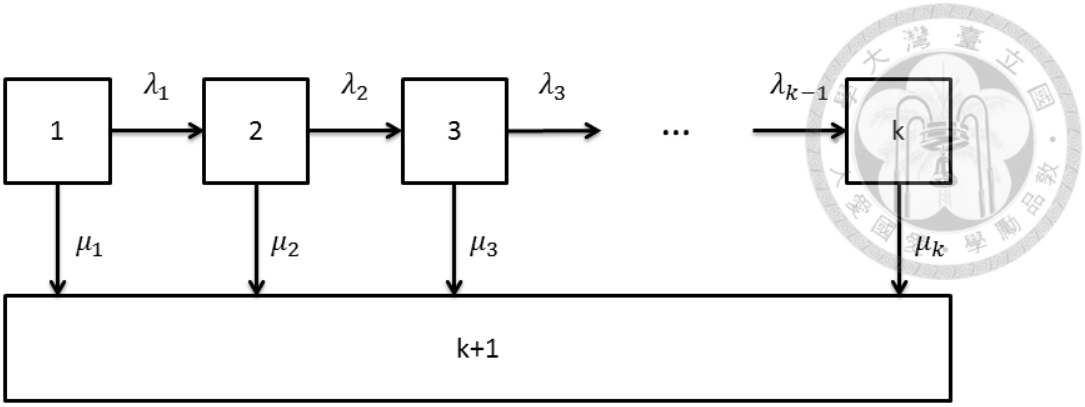
The effect of relevant covariates on the WT of the complier was modelled by using the truncated Poisson regression model

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \quad (4\text{-}3)$$

where $\beta_j's$ are regression coefficients corresponding to covariates $x_{ij}'s$, and k is the number of covariates considered in each of the model.

4.2 Coxian phase-type distributions

The Coxian phase-type distribution describes the time to absorption of a finite Markov chain in continuous time. This Markov chain has one absorbing phase and k transient phases as follows.



The process only starts in the first transient phase. We know $(\mathbf{p}, p_{k+1}) = (1,0,0,\cdots,0,0)$. While LOS data are analyzed, transient phases can represent the severity of illness and absorbing phase can represent discharge or death, and while WT data are analyzed, transient phases can represent the hidden transition and absorbing phase can represent referral for colonoscopy. When entering the system from the first phase, the subject may move to the second transient phase or the absorbing phase. It is a progressive model and does not allow reverse transitions. As indicated earlier, if the process is from transient phase to transient phase, the parameter of its transition rate is λ_i. If the process is from transient phase to absorbing phase, the parameter of its transition rate is μ_i. Therefore, given in the ith phase at time t, the probability of patient in the i+1th phase after a short time (Δt) is $\lambda_i \Delta t + o(\Delta t)$ for i=1,…,k-1.

$$P(X(t+\Delta t) = i+1 \mid X(t) = i) = \lambda_i \Delta t + o(\Delta t) \quad (4\text{-}4)$$

Given in the ith phase at time t, the probability of patient in the absorbing phase after a short time (Δt) is $\mu_i \Delta t + o(\Delta t)$ for i=1,…,k.

$$P(X(t+\Delta t) = k+1 \mid X(t) = i) = \mu_i \Delta t + o(\Delta t) \quad (4\text{-}5)$$

Phases {1,…,k} are transient and phase k+1 is absorbing.

The random variable T that is defined as the time to absorption is said to have a Coxian phase-type distribution. The infinitesimal generator for the Markov chain can be written in block-matrix form as

$$G = \begin{pmatrix} \mathbf{Q} & \mathbf{q} \\ \mathbf{0} & 0 \end{pmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -\mu_k \end{bmatrix}$$

$$\mathbf{q} = (\mu_1 \ \mu_2 \ \cdots \ \mu_k)^T$$

To ensure absorption in a finite time with probability one, it requires that every non-absorbing state is transient, so they block the matrix **G** and let the matrix **Q** do not consider the absorbing state. Due to the absorption in a finite time with probability one, the process with **Q** is an honest Markov process. Therefore, when we want to get solution of the differential equations, we can consider the use of forward and backward Kolmogorov equations. Both sets of equations have the same unique solution to an honest Markov process.

Suppose that initially state i is occupied by X(0)=i, and let

$$p_{ij}(t) = P(X(t) = j \mid X(0) = i).$$

The forward equations are obtained by the following argument. For $\Delta t>0$,

$$p_{ik}(t + \Delta t) = p_{ik}(t)[1 + g_{kk}\Delta t] + \sum_{j \neq k} p_{ij}(t) g_{jk} \Delta t + o(\Delta t),$$

leading to

$$p'_{ik}(t) = \sum_j p_{ij}(t) g_{jk}.$$

If we define a matrix **P(t)**, having $p_{ij}(t)$ as its (i,j)th element, then

$$\mathbf{P}'(\mathbf{t}) = \mathbf{P}(\mathbf{t})\mathbf{G}. \tag{4-6}$$

Now consider the backward equations, it can be obtained by the following argument. For $\Delta t>0$,

$$p_{ij}(t + \Delta t) = p_{ij}(t)[1 + g_{ii}\Delta t] + \sum_{k\neq i} p_{kj}(t) g_{ik}\Delta t + o(\Delta t),$$

leading to

$$p'_{ij}(t) = \sum_k g_{ik} p_{kj}(t).$$

If we define a matrix **P(t)**, having $p_{ij}(t)$ as its (i,j)th element, then

$$\mathbf{P}'(\mathbf{t}) = \mathbf{G}\mathbf{P}(\mathbf{t}). \tag{4-7}$$

The (4-6) and (4-7) with initial condition $\mathbf{P}(\mathbf{0}) = \mathbf{I}$ have the formal solution

$$\mathbf{P}(\mathbf{t}) = \exp(\mathbf{G}t) \tag{4-8}$$

$$= \sum_{n=0}^{\infty} \mathbf{G}^n \frac{t^n}{n!} \tag{4-9}$$

When **G** is a finite matrix, that is when the number of states of the process is finite, the series (4-9) is convergent and (4-8) is the unique solution of both forward and backward equations.

If we assume the process has 3 transient phases and one absorbing phase (4^{th} phase). The transition matrix is expressed as

$$G = \begin{bmatrix} -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \mu_1 \\ 0 & -(\lambda_2 + \mu_2) & \lambda_2 & \mu_2 \\ 0 & 0 & -\mu_3 & \mu_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We can obtain its transition probability by (4-8) or using the stochastic integral as follows.

$$P_{12}(t) = \frac{\lambda_1[e^{-(\lambda_1+\mu_1)t} - e^{-(\lambda_2+\mu_2)t}]}{(\lambda_2 + \mu_2) - (\lambda_1 + \mu_1)}$$

$$P_{13}(t) = \frac{\lambda_1\lambda_2}{\mu_3 - (\lambda_2 + \mu_2)}\left[\frac{e^{-(\lambda_1+\mu_1)t} - e^{-(\lambda_2+\mu_2)t}}{(\lambda_2 + \mu_2) - (\lambda_1 + \mu_1)} - \frac{e^{-(\lambda_1+\mu_1)t} - e^{-\mu_3 t}}{\mu_3 - (\lambda_1 + \mu_1)}\right]$$

$$P_{23}(t) = \frac{\lambda_2[e^{-(\lambda_2+\mu_2)t} - e^{-\mu_3 t}]}{\mu_3 - (\lambda_2 + \mu_2)}$$

$$P_{24}(t) = 1 - e^{-(\lambda_2+\mu_2)t} - \frac{e^{-(\lambda_2+\mu_2)t} - e^{-\mu_3 t}}{\mu_3 - (\lambda_2 + \mu_2)}$$

$$P_{34}(t) = 1 - e^{-\mu_3 t}$$

$$P_{14}(t) = 1 - P_{11}(t) - P_{12}(t) - P_{13}(t) = 1 - e^{-(\lambda_1+\mu_1)t} - P_{12}(t) - P_{13}(t) \quad (4\text{-}10)$$

If the random variable T_i represents their LOS/WT in phase i, where $T_i \sim \exp(\lambda_i + \mu_i)$, the MGF for the length of time a patient spends in phase i is given by

$$M_{T_i}(\theta) = \frac{\lambda_i + \mu_i}{\lambda_i + \mu_i - \theta}$$

Therefore, the expected LOS/WT in phase i, determined by

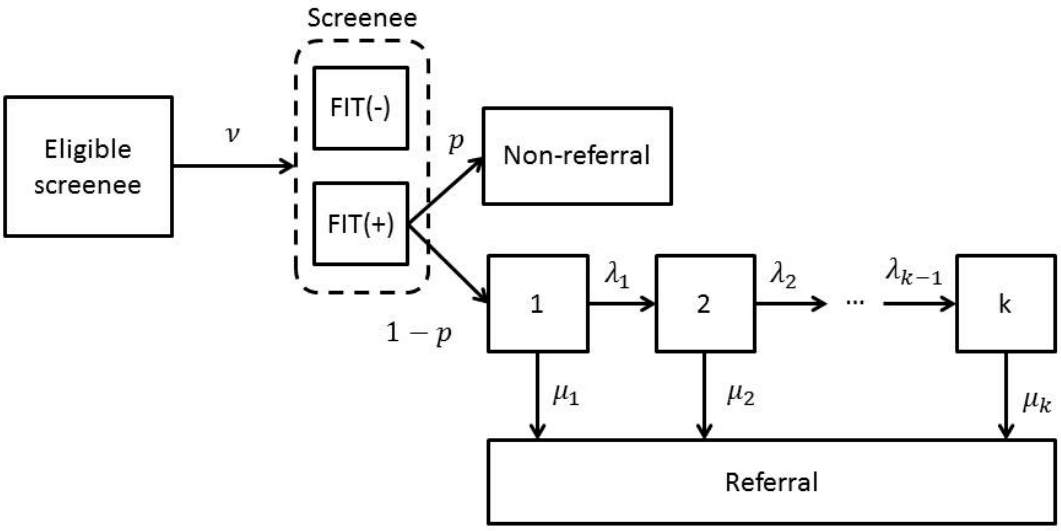
$$E[T_i] = \frac{1}{\lambda_i + \mu_i}$$

And the marginal mean LOS/WT in the system can be obtained by

$$E[T] = \int_0^{\infty} t \, dP_{14}(t)$$

4.3 Queue Hurdle Coxian Phase-type model

In order to take into account the arrival rate of eligible screenees, non-compliance with colonoscopy and the WT for undergoing colonoscopy simultaneously, we used the Poisson distribution to model arrival rate and apply Coxian Phase-type distribution to non-hurdle part of hurdle model. As a result, we developed the Queue Hurdle Coxian Phase-type model as follows.



In the Queue Hurdle Coxian Phase-type model, it has three components: (1) Poisson Queue process, ν is the arrival rate of eligible screenees per person-days, $y_i = 1$ is positive FIT and $y_i = 0$ is negative FIT, the Poisson distribution can be displayed as

$$f(y_i) = \frac{\exp(-\nu t_i)(\nu t_i)^{y_i}}{y_i!},\ \nu > 0; i = 1, \ldots, n\ ,$$

(2) Probability of non-compliance with colonoscopy, say p, (3) Coxian Phase-type distribution, its pdf is the derivative of the transition probabilities derived from (4-8), therefore the Queue Hurdle Coxian Phase-type distribution can be expressed as

$$f(t_1, t_2) = \begin{cases} e^{-\nu t_1} & , \text{FIT } (-) \\ e^{-\nu t_1}(\nu t_1) \times p & , \text{FIT } (+) \text{ but nonrefer} \\ e^{-\nu t_1}(\nu t_1) \times (1-p) \times f_C(t_2) & , \text{FIT } (+) \text{ and referral} \end{cases} \quad (4\text{-}11)$$

where t_1 is the arrival time from invitation date to screening date, t_2 is the WT for undergoing colonoscopic exam, and $f_C(t_2)$ is the p.d.f of Coxian Phase-type distribution based on the derivative of the transition probabilities derived from (4-8).

Chapter 5 Results

I. Computer Simulation for Estimating Parameters

To test if the Coxian phase-type model can be simulated by directly using a mixture of Poisson process, we fit the continuous positively skewed data on which the research conducted by Marshall[4] was based after simulating their tabular data on LOS of the geriatric patients. As the most adequate model was fitted by a 3-phase Coxian phase-type distribution we simulated the data by a 3-state mixture Poisson process with the probability density function expressed as follows.

$$f(t) = \pi_1 \times \theta_1 e^{-\theta_1 t} + \pi_2 \times \theta_2 e^{-\theta_2 t} + \pi_3 \times \theta_3 e^{-\theta_3 t}, (\pi_1 + \pi_2 + \pi_3 = 1)$$

We set $\pi_1 = 0.46$, $\pi_2 = 0.40$, $\pi_3 = 0.14$ and $\theta_1 = 0.07$, $\theta_2 = 0.05$, $\theta_3 = 0.02$. The data set in Marshall's study indicated the LOS ranged from 0 to 350 days, with a mean of 23 days and a median of 12 days. The simulated data shows the LOS ranged from 0 to 358 days, with 23 days and a median of 14 days (Figure 5-1), which was very close to their original empirical findings.

The Coxian phase-type distribution was fitted to the simulated data by using SAS software. SAS implements an optimization function with the method of maximum likelihood estimation (MLE) given the formulation of the log-likelihood function for different kinds of phase-type distribution. We used the Newton-Raphson algorithm and the minimum Bayesian information criterion (BIC) to decide the most parsimonious

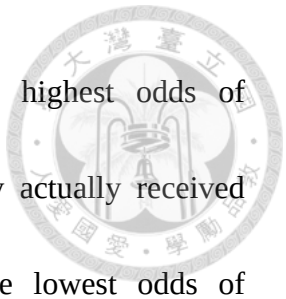
model. In Table 5-1, the estimated parameters in the one or two phase Coxian phase-type model were close between the original results and our simulated data. However, when the number of phase increased there was a larger discrepancy. The results of simulation suggest while the hidden phases increased the heterogeneity across different phases could not be captured by a mixture of Poisson process.

II. Compliance with colonoscopy from positive FIT of Taiwan nationwide colorectal cancer screening program

Univariate Analyses and Multivariate Analyses for the Hurdle model

In order to identify factors associated the non-compliance for colonoscopy and those affecting WT for undergoing colonoscopy, we used the hurdle model to deal with these two problems simultaneously.

The hurdle part is to identify which factors might influence subject not to take colonoscopic exam and the non-hurdle part is to identify which factors would affect WT for colonoscopy among attendees complying with colonoscopy. As shown in Table 5-2, the effects of gender on both parts of model were lacking of statistical significance. Compared with the age group of 50-54, the older age groups had higher odds of refusing to receive colonoscopy whereas the complier after they underwent colonoscopy exam, the effect of age on WT became not statistically significant. In geographic area,



those residing in Eastern Taiwan or offshore islands had the highest odds of non-compliance and had the longest WT for colonoscopy if they actually received colonoscopy exam; those dwelling in Northern Taiwan had the lowest odds of non-compliance and had the shorter WT for colonoscopy than those dwelling in Southern or Eastern Taiwan or offshore islands. Those who attended screening at public health centers had the lowest odds of non-compliance and had the short WT for colonoscopy; those who attended screening at local clinic had the highest odds of non-compliance and had the longest WT for colonoscopy. Attendees living in secondary urban or undergoing screening at inaugural period or being detected at subsequent screening had the lowest odds of non-compliance and had the shortest WT for colonoscopy.

Before fitting the multivariate model, the model selection was done and shown in Table 5-3. Because the change in the structure of screening program during the year from 2009 to 2010 might results in heterogeneity between the inaugural period and rolling out period, we evaluated the interaction between factors and periods of screening program. The results of model selection reveal that the hurdle part included seven baseline characteristics and interaction of periods of screening program between geographic areas and type of screening units, and the non-hurdle part contained six baseline characteristics (excluding gender effect) and interaction of periods of screening

program between geographic areas, type of screening units and urbanization levels. As presented in Table 5-4, multivariate analysis in hurdle part found female, older people, those who lived less urbanized area and those were detected at prevalent screen had higher chance of not complying with colonoscopy. During inaugural period, attendees of eastern Taiwanese or offshore islands had a highest odds of not complying with colonoscopy (OR: 1.51, 95% CI: 1.36-1.66) compared with the northern attendee, but in rolling out period, middle Taiwanese had the highest odds of not complying with colonoscopy (OR: 1.08 95% CI: 1.06-1.10). Although screening at public health centers had the lowest odds in both periods, screening at hospital had the highest odds (OR: 2.54, 95% CI: 2.39-2.69) in inaugural period but decreased during the rolling out period (OR: 1.08, 95% CI 1.06-1.10) and screening at local clinic (OR: 1.79, 95% CI: 1.74-1.84) had the highest odds in rolling out period. When taking the non-hurdle part into account, the results presented in Table 5-5 show attendees who aged between 65 and 69 years had the longest WT for colonoscopy if they actually complied with colonoscopy, but three other age groups had not much difference. Those detected at subsequent screen had shorter WT for colonoscopy than prevalent screen. During inaugural period, attendees living in middle Taiwan (RR: 1.14, 95% CI: 1.07-1.20) or main urban or undergoing screening at public health centers (RR: 1.22, 95% CI: 0.99-1.46) had the shortest WT for colonoscopy. During the rolling out period, those

who lived in middle Taiwan (RR: 1.13, 95% CI: 1.09-1.16) or secondary urban area (RR: 1.07, 95% CI: 1.06-1.09) or undergoing screening at hospital (RR: 1.14, 95% CI: 1.12-1.15) had the shortest WT. It indicates the similar trend in geographic areas in both periods, and the estimates of RR of northern people increased from 1.03 to 1.12.

Queue Hurdle Coxian Phase-type model

As we had already known there was heterogeneity between the inaugural period (2004-2009) and the rolling out period (2010-2013), so we analyzed these two separately. In the current thesis, we only considered the modelling with the Coxian phase-type model using the data on the inaugural period. The continuous data are positively skewed with a long tail, representing those few attendees who had not received colonoscopic exam for an extremely long WT (Figure 5-2) that justifies the WT had better be modelled by the Coxian phase-type distribution. To decide the most appropriate model, we still used the minimum BIC to determine. In Table 5-6, we found the Queue hurdle 2-phase Coxian phase-type model was the most suitable model due to minimum BIC score and could be classified as short waiting (step-by-step) phase and long waiting (shilly-shally) phase. It can be clearly seen that the 3-phase Coxian phase-type model had higher BIC value and also showed the identifiability problem between the referral rate from the moderate waiting phase (μ_2) and the transition rate

from the moderate waiting phase to long waiting phase (λ_2). We observed that regardless of numbers of WT phases in the model, all of them indicate the same arrival rate equal to 0.00021 per person-days and the probability of not complying with colonoscopy was 0.2647. It is very interesting to note that the referral rate was five times greater in the short waiting phase than the long waiting phase. Around 15% subjects were in a dilemma to be referred to undergo colonoscopy so as to be trapped in long waiting phase. In Queue hurdle 2-phase Coxian phase-type model, Table 5-7 shows that the mean WT in short waiting phase and the mean WT in long waiting phase was 32 days and 169 days, respectively. The marginal mean WT was 35 days.

Assuming covariates would affect referral rate and the transition rate, we used the coefficients estimated from the non-hurdle part of the hurdle regression model as a new covariate (score):

$$\begin{aligned}\text{score}_{\text{i}} = &-3.7554 + 0.0217 \times \text{age}_{50-54} + 0.0181 \times \text{age}_{55-59} \\ &+0.0198 \times \text{age}_{60-64} + 0.0321 \times \text{area}_{\text{north}} + 0.1276 \times \text{area}_{\text{mid}} \\ &+0.0770 \times \text{area}_{\text{south}} - 0.2038 \times \text{unit}_{\text{hospital}} \\ &+0.2006 \times \text{unit}_{\text{public}} - 0.0227 \times \text{urban}_{\text{secondary}} \\ &-0.0567 \times \text{urban}_{\text{rural}} + 0.0364 \times \text{subsequent}\end{aligned}$$

All of these significant covariates transformed into a new continuous covariate, so we could reduce parameters to be estimated from 11 to 1. We also made it become a

binary outcome according to the cutoff of median value

$$G = \begin{cases} 1 \text{ , if score}_i > -3.4981 \\ 0 \text{ , if score}_i < -3.4981 \end{cases}$$

Using the proportional hazards regression form to compare referral rate of higher score group and lower score group gave the following expression:

$$\mu_1 = \mu_{01}\exp(\gamma \times G)$$

$$\mu_2 = \mu_{02}\exp(\gamma \times G)$$

$$\lambda_1 = \lambda_{01}\exp(\gamma \times G)$$

We fitted 1- and 2-phase model to determine which model was more suitable to explain data. In Table 5-8, the result of 1-phase model shows the higher score, the faster referral rate for colonoscopy (P < 0.001). The mean WT was 38 days in lower score group and 32 days in higher score group. In the 2-phase models, risk score might have the impact on the transition rate from short waiting phase to undergoing colonoscopy (μ_1), or from long waiting phase to undergoing colonoscopy (μ_2), or from short waiting phase to long waiting phase (λ_1). Table 5-9 indicates the model with score related to μ_1 was the most appropriate model. In addition, the 2-phase model was better than 1-phase model as well. The mean WT in short waiting phase were 36 days and 30 days corresponding to low score group and high score group, separately. In longer waiting phase, the mean WT was 167 days among these two groups. The marginal mean WT was 38 days in low

score group and 32 days in high score group.

According to this model with score related to the referral rate, we could predict its transition probabilities at different times. In Figure 5-3, the probability of staying in short waiting phase, P_{11}, declined over time and those with lower score had longer WT in short waiting phase than higher score given the same probability of staying in short waiting phase. The transition probability from short waiting phase to long waiting phase, P_{12}, was pretty small and no difference between these two groups. The transition probability of undergoing colonoscopy, P_{13}, increased over time, because patients would receive colonoscopic exam eventually. Under the same transition probability to undergo colonoscopy, low score group had longer WT than high score group.

III. Application II : Hospitalization of colorectal cancer patients

There were 178 CRC patients in Shin Kong Wu Ho-su Memorial Hospital (SKH) between 1999 and 2013. The variables of interest include the patients' LOS (recorded in days), measured from the day of admission of a patient until they have been discharged.

The continuous data are positively skewed with a long tail, representing those few patients who have remained in hospital for an extremely long time (Figure 5-4). The LOS ranges from 1 to 215 days, with a mean of 13.8 days and a median of 7 days.

The Coxian phase-type distribution was fitted to the LOS data. Table 5-10 displays

the fitted parameters and the BIC score for each of the model under consideration. It could be found that the most suitable number of phases was 3, because it had the minimum BIC score, which was 1157. A 3-phase model could be classified as short-stay, medium-stay and longer-stay care because if patient had the serious condition, he/she would have poor resistance to infections. Therefore, the severer condition CRC patients had, the longer patients stayed in the hospital (Figure 5-5). The absorbing rate of discharge for short-stay was five times than that of medium-stay and long-stay. The transition from short-stay to medium stay was five times that from medium-stay to long-stay. The expected LOS is displayed in Table 5-11. In short-stay phase, the expected LOS was 10 days whereas both medium-stay and longer-stay phases were 49 days. The marginal expected LOS was 14 days.

As we found that the transition rate from medium stay to discharge or death and the transition rate from longer stay to discharge or death were very close we therefore tested this current 3-phase model against a new model assuming these two transition rates were equal. Table 5-12 shows the original 3-phase model was better because of the smaller AIC. Therefore, we still kept two transition rates distinct although they were close.

To make sure if there still existed a better model, we attempted to fit another model. We could find that discharge type 1, 3 and 5 had shorter LOS and could be regarded as

discharge due to recovery. Discharge type 4, 6 and A had longer LOS and could be regarded as discharge due to severer condition. These patients had aggravated condition or even death. Therefore, we divided discharge types into recovery and death and we then used Coxian phase-type model with two absorbing states to fit the data. At first, we needed to determine how many phases would be appropriate, so we were also based on BIC score to decide. As shown in Table 5-13, 3-phases model was the most suitable case. However, if we only focused on 3-phases Coxian phase-type models, it indicates the model with one absorbing state was still better than that with two absorbing states due to the smaller BIC score. As a result, we reckon the 3-phases Coxian phase-type model with one absorbing state was the most appropriate model.

Coxian phase-type models with covariates

After confirming the 3-phase Coxian phase-type model was the most suitable one to fit the data, we wonder if the transition rate would be influenced by covariates so that the number of phase could be reduced in the model. As a result, we used the 2-phase Coxian phase-type model to explore this issue and applied the proportional hazards form. We assumed that gender or age would affect transition rate in the following five scenarios:

(1) From short-stay to longer-stay

$$\lambda_1 = \lambda_{01} \exp(\beta_1 \times \text{gender}_{\text{Female}})$$

or

$$\lambda_1 = \lambda_{01} \exp(\beta_1 \times \text{age}_{60-74} + \beta_2 \times \text{age}_{\leq 60})$$

(2) From short-stay to absorbing state (Death/Discharge)

$$\mu_1 = \mu_{01} \exp(\beta_1 \times \text{gender}_{\text{Female}})$$

or

$$\mu_1 = \mu_{01} \exp(\beta_1 \times \text{age}_{60-74} + \beta_2 \times \text{age}_{\leq 60})$$

(3) From longer-stay to absorbing state

$$\mu_2 = \mu_{02} \exp(\beta_1 \times \text{gender}_{\text{Female}})$$

or

$$\mu_2 = \mu_{02} \exp(\beta_1 \times \text{age}_{60-74} + \beta_2 \times \text{age}_{\leq 60})$$

(4) From short-stay to absorbing state and from longer-stay to absorbing state

$$\mu_1 = \mu_{01} \exp(\beta_1 \times \text{gender}_{\text{Female}})$$

$$\mu_2 = \mu_{02} \exp(\beta_2 \times \text{gender}_{\text{Female}})$$

or

$$\mu_1 = \mu_{01} \exp(\beta_1 \times \text{age}_{60-74} + \beta_2 \times \text{age}_{\leq 60})$$

$$\mu_2 = \mu_{02} \exp(\beta_3 \times \text{age}_{60-74} + \beta_4 \times \text{age}_{\leq 60})$$

(5) Joint effect on 3 transition paths

$$\mu_1 = \mu_{01}\exp(\beta_1 \times \text{gender}_{\text{Female}})$$

$$\mu_2 = \mu_{02}\exp(\beta_2 \times \text{gender}_{\text{Female}})$$

$$\lambda_1 = \lambda_{01}\exp(\beta_3 \times \text{gender}_{\text{Female}})$$

or

$$\mu_1 = \mu_{01}\exp(\beta_1 \times \text{age}_{60-74} + \beta_2 \times \text{age}_{\leq 60})$$

$$\mu_2 = \mu_{02}\exp(\beta_3 \times \text{age}_{60-74} + \beta_4 \times \text{age}_{\leq 60})$$

$$\lambda_1 = \lambda_{01}\exp(\beta_5 \times \text{age}_{60-74} + \beta_6 \times \text{age}_{\leq 60})$$

Firstly, we took gender into account. Distributions of gender list in Table 5-14.The results of these five situations are shown in Table 5-15. Based on their corresponding BIC scores, it indicates that the model having gender effect on transition rate from short-stay to absorbing state was the most appropriate model. It also shows that male would discharge or die earlier than female. The mean LOS in short-stay was 9 and 12 days corresponding to male and female, separately, and both were 53 days in longer-stay. According to this 2-phase Coxian phase-type model with gender effect on the transition rate from short-stay to absorbing state, we could predict its transition probabilities at different times. In Figure 5-6, the probability of staying in short-stay state, P_{11}, declined over time and female had longer LOS in short-stay than male given the same probability of staying in short-stay state. The transition probability from short-stay to longer-stay,

P_{12}, was pretty small, but female still had higher transition probability than male under the same LOS. The transition probability from short-stay to absorbing state, P_{13}, increased over time, because patients would discharge or die eventually. Under the same transition probability from short-stay to absorbing state, female had longer LOS than male.

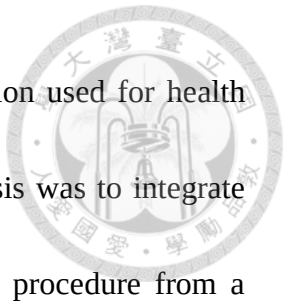
Secondly, we took age into account. Distributions of age list in Table 5-16. The results of these five scenarios are shown in Table 5-17. Based on their corresponding BIC scores, it indicates that the model having age effect on transition rate from longer-stay to absorbing state was the most appropriate model. It also shows that the elderly would discharge or die earlier than the young. All of their mean LOS in short-stay was 9 days and the mean LOS in longer-stay was 18, 19 and 80 days corresponding to those aged above 75, 60-74 and below 60, separately. According to this 2-phase Coxian phase-type model with age effect on transition rate from longer-stay to absorbing state, we could also predict its transition probabilities at different times. In Figure 5-7, there was no age effect on the probability of staying in short-stay state and we also found that those aged at 60-74 or above 74 were not different irrespective of the transition probabilities from short-stay to longer-stay or from short-stay to absorbing state. The transition probability from short-stay to longer-stay in patients aged 60 years or below had higher transition probability than the other two groups given the same

LOS. Under the same transition probability from short-stay to absorbing state, those aged below 60 also had longer LOS.

Chapter 6 Discussion

The innovation of combing queue process, hurdle model and Coxian phase-type distribution

The values of current thesis are not only held from the aspect of applications but also from the aspect the methodology. It is very interesting to put emphasis on the reciprocal feedbacks between both aspects rather than only put emphasis on single aspect. It is very intuitive to ask why we need the Coxian phase-type model here? Is it sufficient for the research people to merely apply the Queue process and the hurdle model with Poisson process? The combination of Queue process, hurdle model, and Coxian phase-type is motivated a very large population-based screening data in Taiwan. We are faced with high demand for around over five million participants eligible for the uptake of CRC screening with FIT, yielding a high demand for the referral of positive FIT to undergo colonoscopy. In contrast to the conventional Queue process that evaluates the arrival rate as opposed to departure rate relating to service time distribution, the non-compliance (non-susceptibility) problem for the referral of positive FIT made the traditional Queue process infeasible and may resort to the use of hurdle model. In addition, those who were willing to consent to undergo colonoscopy may be classified into different types according to WT for colonoscopy. This raised the rationale for using the Coxian phase-type model for detecting whether it can identify hidden



phase during the WT so as to provide a new insight into information used for health promotion for enhancing the referral rate. Although the current thesis was to integrate three types of model, we still analyzed the data with step-by-step procedure from a simple statistical approach to the final new queue hurdle Coxian phase-type model in order to get a better understanding of the contrasts between the proposed models by decomposing each part into analysis.

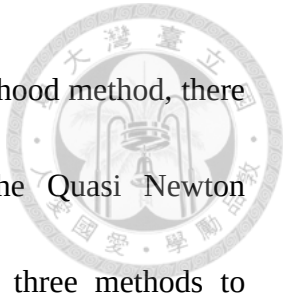
After modelling screening data, health policy-makers are also concerned with the LOS in hospital for CRC patients because different types of LOS may reflect different severity of disease status (including the severity of CRC and co-morbidity) as well as various costs involved in hospitalization and modelling the transition between different hidden phases. The incorporation of relevant covariates is also one of novelties in the current thesis. The idea of this part was identical to those envisaged by the Marshal et al study.

Thoughts of Statistical Models

As mentioned above, we had tried a step-by-step approach with various statistical methods to identify factors affected WT for undergoing colonoscopy. At first, we utilized the Cox regression model to elucidate factors affecting WT. It might be inappropriate because we did not take non-complier into account. We attempted to use

logistic regression model to deal with the non-compliance part and the Cox regression model pertaining to WT distribution, and compared with the hurdle model, which is the mixture of the logistic regression model and the truncated Poisson regression model. However, we found that on the non-hurdle part, some characteristics had dissimilar rates to undergo colonoscopic examination between the Cox regression and the truncated Poisson regression. Because in the presence of covariates in the Cox regression model, it cannot have a proportional hazards structure if the covariates are modelled through p via a binomial regression model[7]. The hurdle model provided two sets of results. These results could also be obtained separately by fitting both a logistic regression and Poisson model[8] that we had found the similar results. The main difference between the hurdle model and the separate model of logistic regression and Poisson model is that covariance between each parameters exists in the hurdle model. As a result, we decided to use the hurdle model to deal with WT issues for colonoscopy. Actually, there is the other model which has the similar concept with the hurdle model called zero-inflated model that have been dealt with by COM-Poisson[9] and generalized Poisson model[10]. The zero-inflated model can deal with zero part (non-complier) as well, but it is not appropriate to apply on screen data because that most of invited subjects are willing to undergo colonoscopy.

In the analysis of LOS for hospitalization, parameter estimation for the Coxian



phase-type distribution was nontrivial. Based on the maximum likelihood method, there are lots of algorithms such as the Nelder-Mead algorithm, the Quasi Newton algorithm[11] and the Newton-Raphson algorithm. We used these three methods to estimate parameters and compared their results with BIC score to determine which methods would fit the Coxian phase-type distribution better. In Table 6-1, both of the Newton-Raphson algorithm and Nelder-Mead algorithm show 3-phase model was the most appropriate model due to the minimum BIC. However, the Quasi Newton algorithm shows 2-phase model was better. In addition, we found that if it was a 1- or 2-phase model, all of them would obtain the same estimates, but when we considered 3- or 4-phase model, they became different. Therefore, among the comparison of all algorithms, both 3- and 4-phase model show the Newton-Raphson algorithm was more suitable because it could get smaller BIC. As a result, we thought the Newton-Raphson algorithm might be the most suitable method to estimate parameters.

The Coxian phase-type distribution describes the time to absorption of a finite Markov chain in continuous time and can be adequate for the continuous positively skewed data with a long tail to get a better understanding of the underlying dynamic hidden phases. However, the real scenario of WT distribution also include non-response data (time=0) and queue process that render the conventional Coxian phase-type model inadequate. As mentioned above, to solve these issues, we therefore developed the

hurdle model in combination with the Coxian phase-type. In the queue hurdle Coxian phase-type model, we used the queue process to estimate the arrival rate of eligible screenees, applying the concept of hurdle model to determine if attendees would receive the confirmatory diagnosis or not, and modelled their WT by the Coxian phase-type distribution if they actually complied with colonoscopy. Based on this model, it is more convenient to consider these three scenarios simultaneously.

With the limited clinical resources, the development of the queue hurdle Coxian phase-type distribution not only provides a new insight into the underlying mechanism of WT for early detection and the duration of hospitalization of CRC, but also can help clinicians or hospital managers improve the quality of service and provide some useful information for making decisions. When applying this model to population-based screening program with the problems of queue and non-response to colonoscopy the findings gave a clue to explore the reasons dominating such differences including provider factors such as the implementation of screening program and medical resources and population factors such as the knowledge and attitude toward CRC screening and medical interventions. They also provide more insight on the promotion of the referral of positive FIT identified from the participants with the uptake of screening program.

Limitations

This new model assumed the arrival rate and the probability of non-compliance were independent. However, in fact, the probability of non-compliance would be affected by the arrival rate. To cope the individual correlation between the parameters, we may use the hierarchical model to improve this circumstance, because the complicated processes can be modelled by a sequence of relatively simple models placed in a hierarchy.

In conclusion, we developed a new queue hurdle Coxian phase-type model to solve the compliance with the uptake of screening using the queue process, the problem of non-compliance with the referral of positive results of screenees to have confirmatory diagnosis using the hurdle model in combination with the Coxian phase-type model to identify hidden phases during the WT for undergoing colonoscopy for the referrals. The Coxian phase-type model was also applied to model the LOS in hospitalization for the treated patients diagnosed as CRC.

Reference

1. Chiu HM, Chen LS, Yen MF, Chiu YH, Fann CY, Lee YC, Pan SL, Wu MS, Liao CS, Chen HH, Koong SL, Chiou ST. Effectiveness of Fecal Immunochemical Testing in Reducing Colorectal Cancer Mortality From the One Million Taiwanese Screening Program. *Cancer*. 2015; 10.1002
2. Zorzi M, Fedeli U, Schievano E, Bovo E, Guzzinati S, Baracco S, Fedato C, Saugo M, Dei Tos AP. Impact on colorectal cancer mortality of screening programmes based on the faecal immunochemical test. *GI cancer.* 2014; 10.1136
3. Yu D, Hopman WM, Paterson WG. Wait time for endoscopic evaluation at a Canadian tertiary care centre: Comparison with Canadian Association of Gastroenterology targets. *Can J Gastroenterol.* 2008; **22**(7):621-6.
4. Marshall AH, Shaw B, McClean SI. Estimating the costs for a group of geriatric patients using the Coxian phase-type distribution. *Statistics In Medicine.* 2007; **26**:2716-2729.
5. Marshall AH, McCrink L. Discrete Conditional Phase-Type Model (DC_Ph) for patient waiting time with a logistic regression component to predict patient admission to hospital. *Computer-Based Medical Systems.* 2009; 553-556
6. Titman AC, Sharples LD. Semi-Markov Models with Phase-Type Sojourn Distributions. *Biometrics.* 2010; **66**: 742-752

7. Ibrahim JG, Chen MH, Sinha D. *Bayesian Survival Analysis.* 2001, New York: Springer-Verlag.
8. Dwivedi AK, Dwivedi SN, Deo S, Shukla R, Kopras E. Statistical models for predicting number of involved nodes in breast cancer patients. *Health (Irvine Calif).* 2010; **2**(7):641-651.
9. Conway RW, Maxwell WL. A queuing model with state dependent service rates. *Journal of Industrial Engineering*. 1962; **12**: 132–136.
10. Consul PC, Famoye F. Generalized Poisson regression model. *Communications in Statistics, Theory and Methods.*1992; **21**: 89-109.
11. Marshall AH, Zenga M. Recent developments in fitting Coxian phase-type distributions in healthcare. *The XIIIth International Conference "Applied Stochastic Models and Data Analysis"* 2009; 482-485.

Figure 3-1. Demographics of screening participants in Taiwanese national CRC screening program from 2004 to 2013

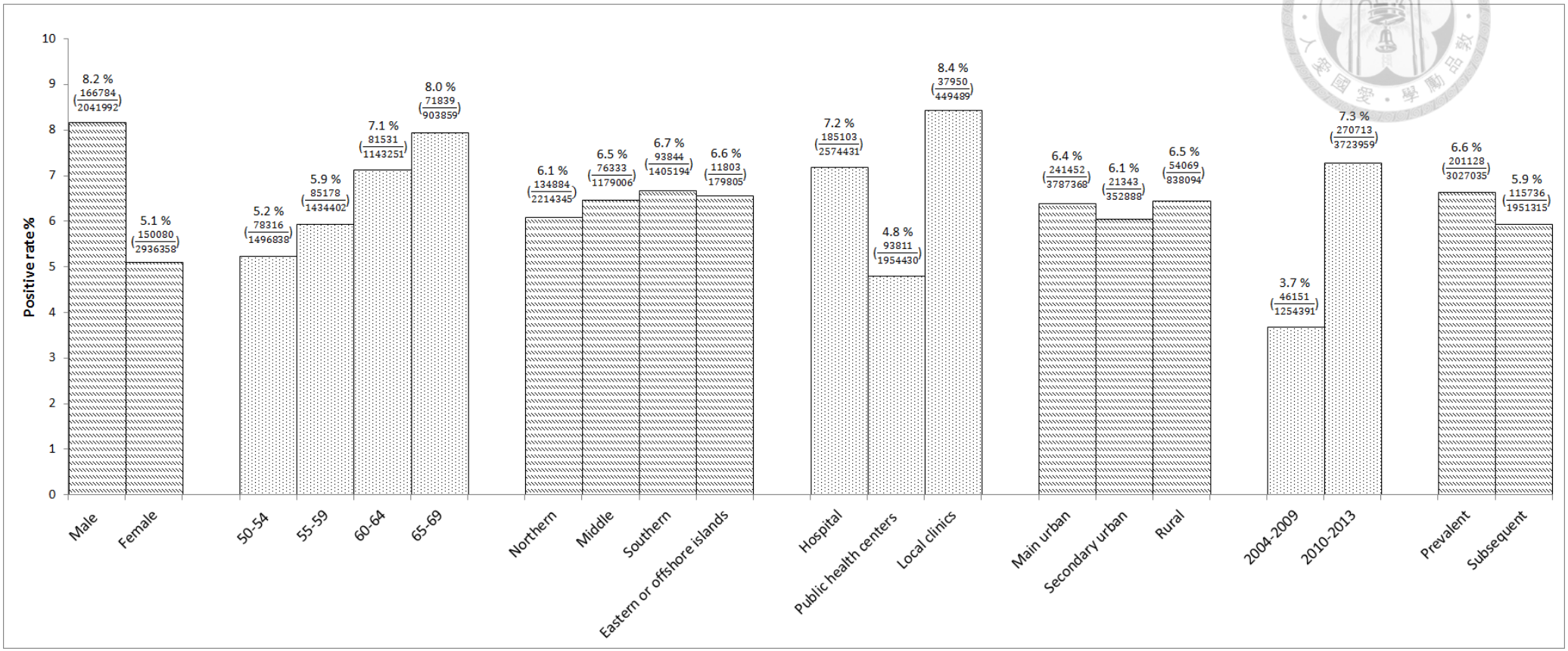


Figure 3-2. Time trend of screening participants number and FIT positive rate in Taiwanese nationwide CRC screening program

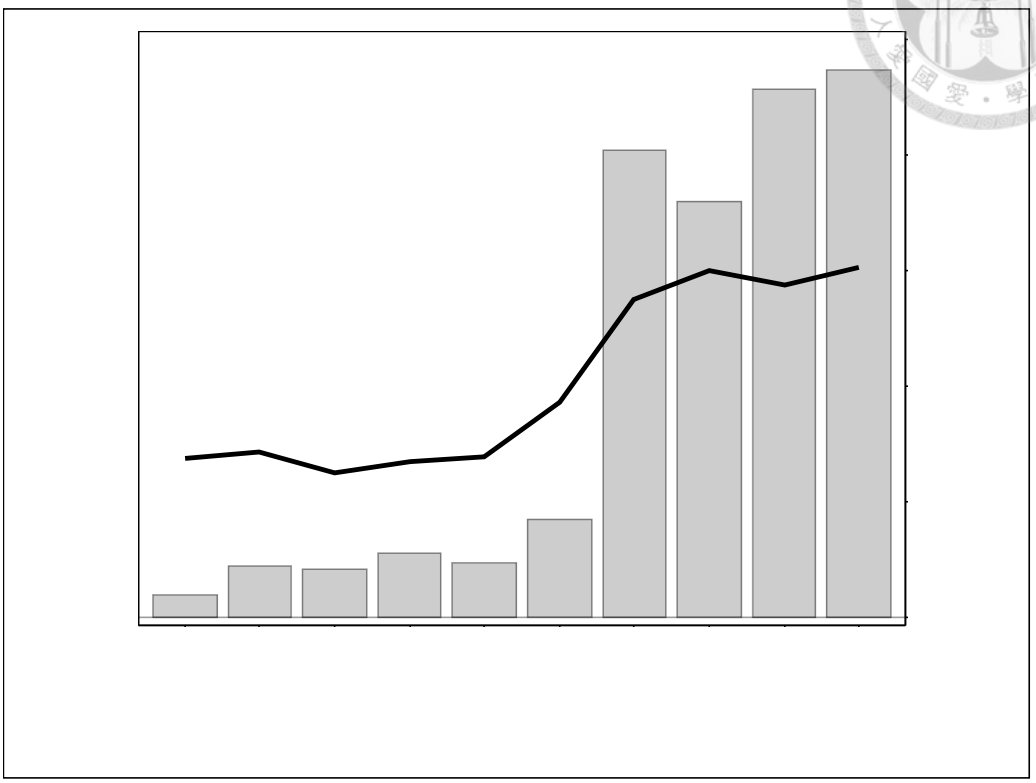


Figure 3-3. Time trend of referral rate and waiting time for colonoscopy in Taiwanese nationwide CRC screening program

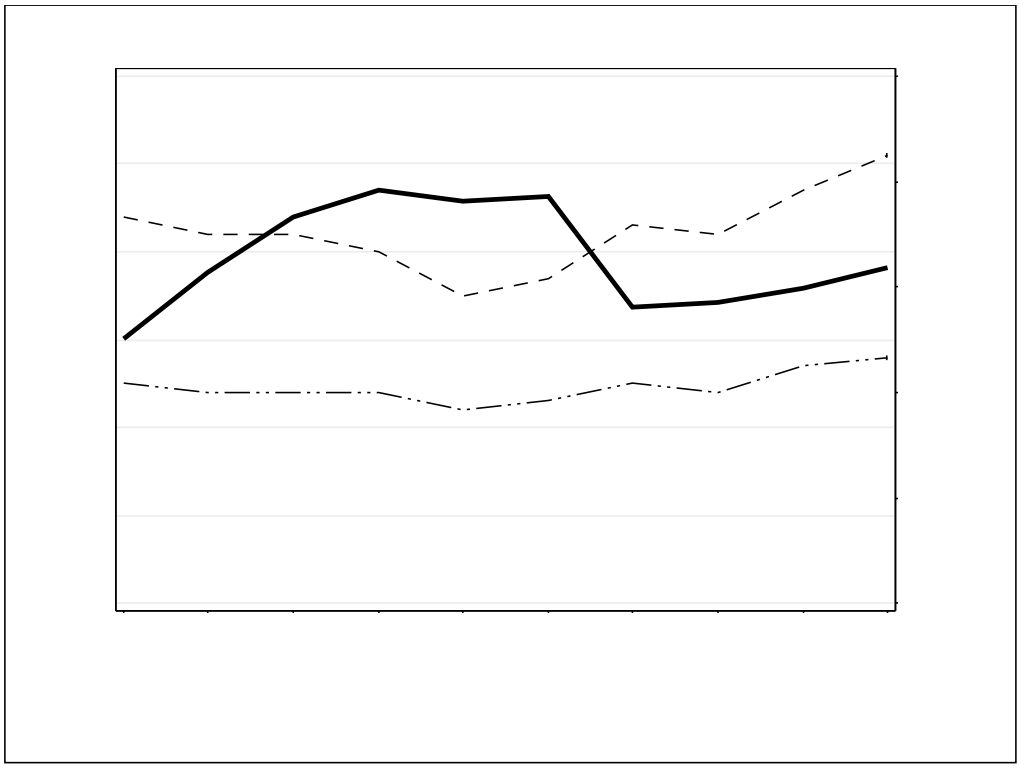
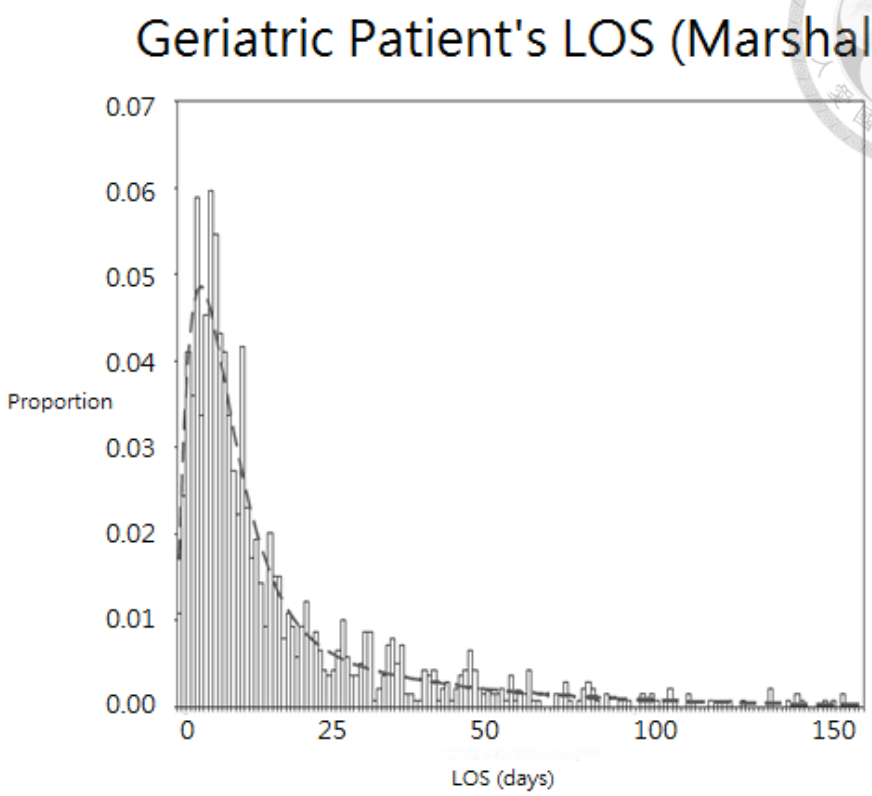


Figure 5-1. Empirical data on Marshall's study (a) and our simulated data (b)



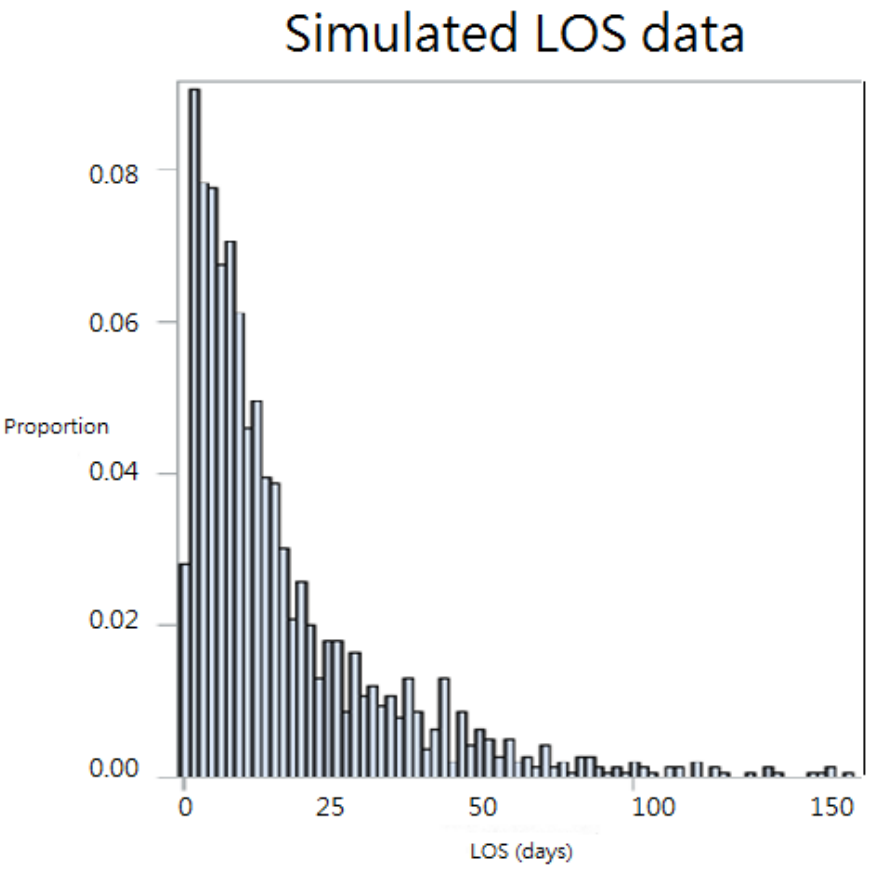


Figure 5-2. Empirical data on waiting time for colonoscopy

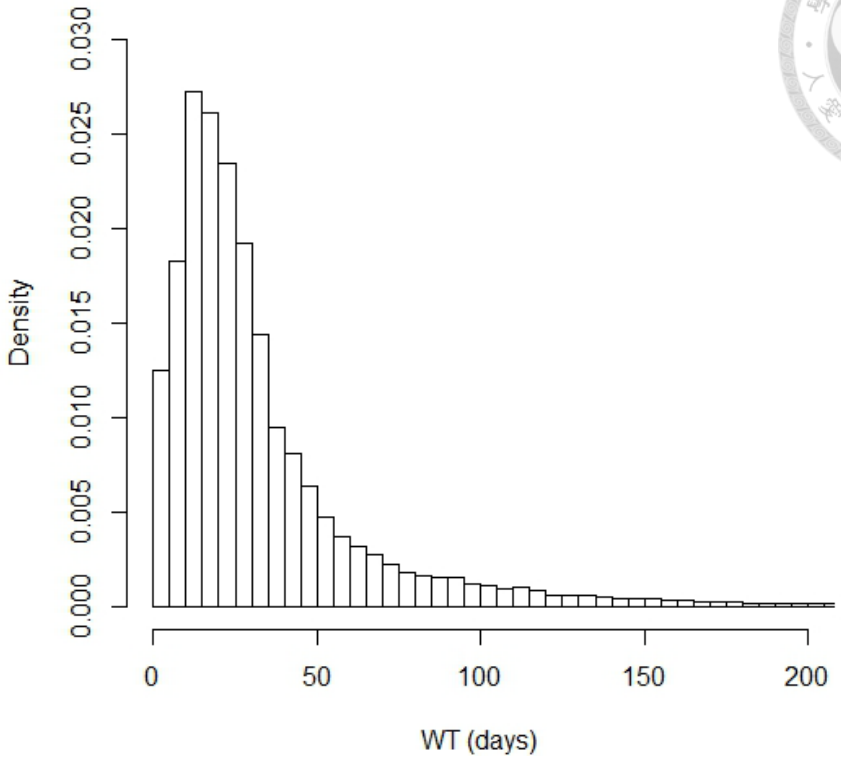


Figure 5-3. Transition probabilities of Coxian two-phase model by risk score

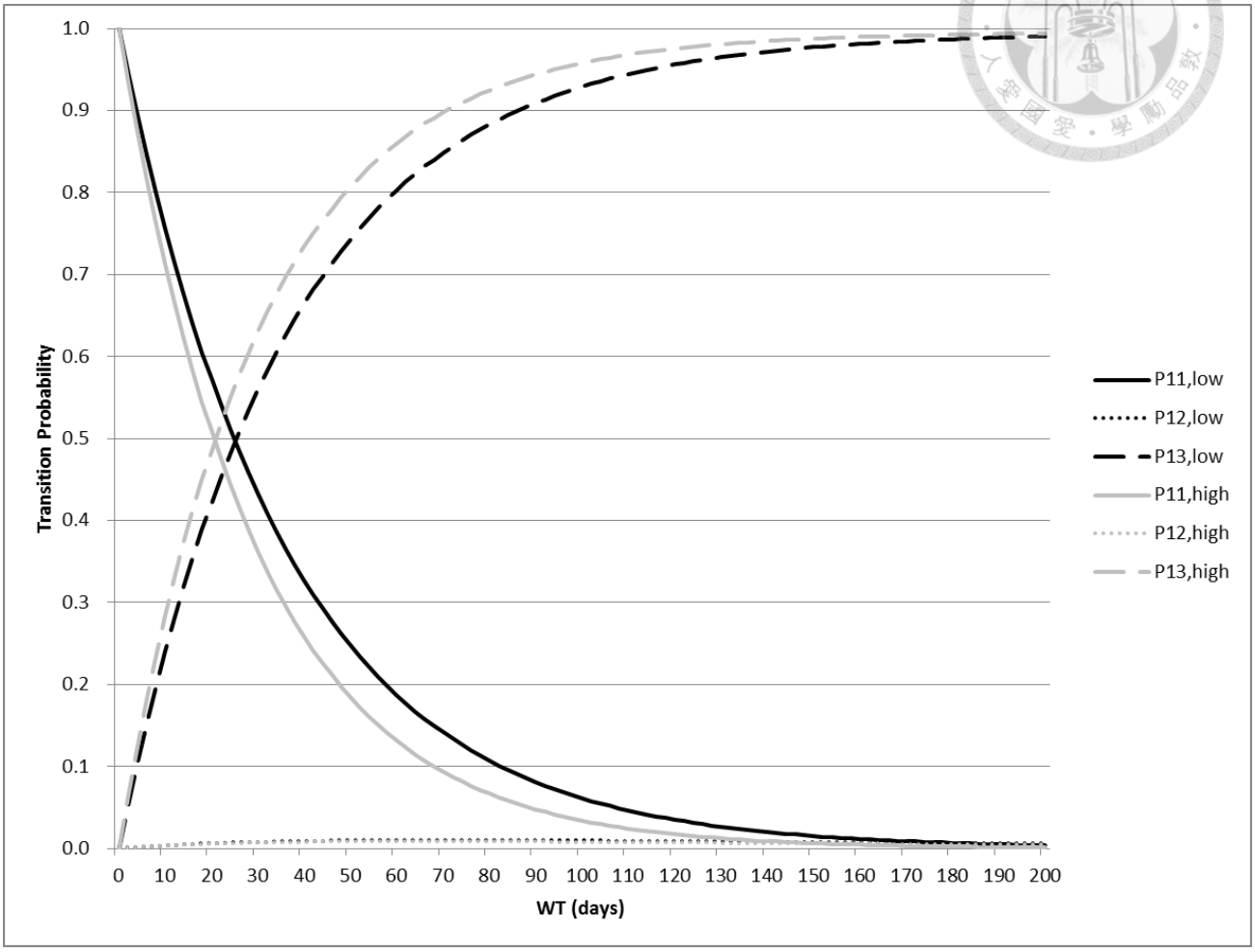


Figure 5-4. Empirical data on LOS

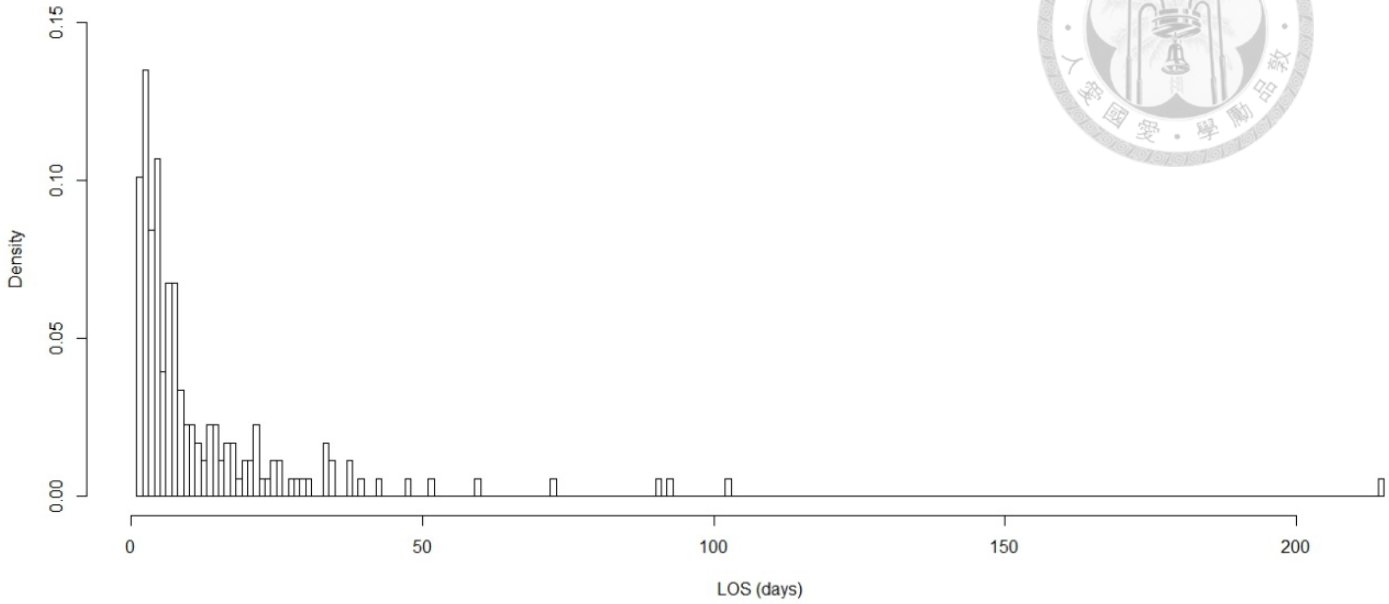


Figure 5-5. Fitted three-phase Coxian phase-type distribution for SKH data set

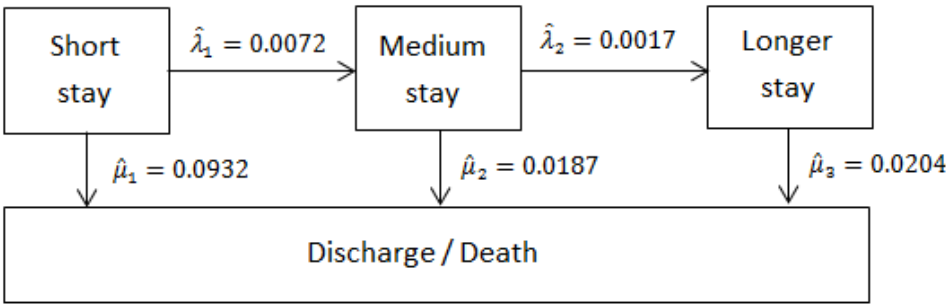


Figure 5-6. Transition probability over time by gender

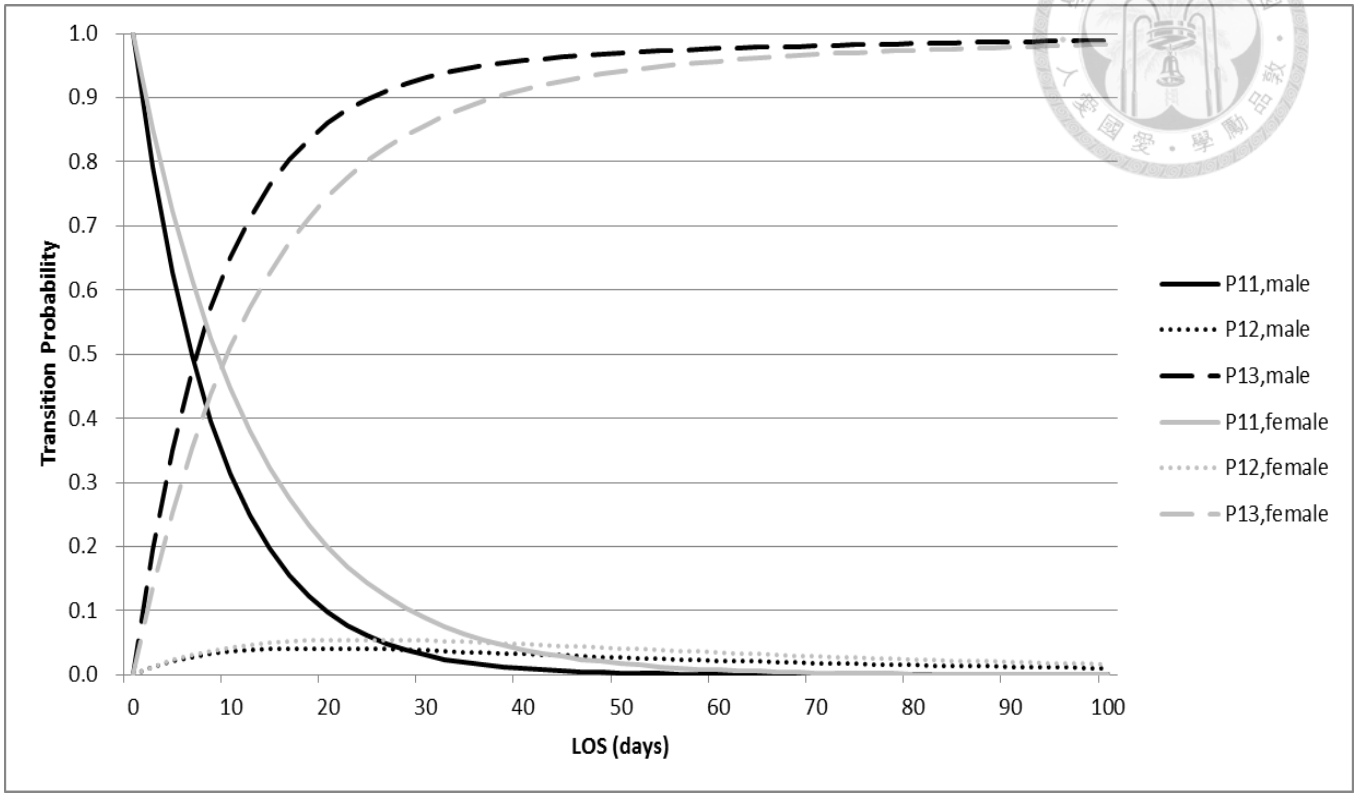


Figure 5-7. Transition probability over time by age

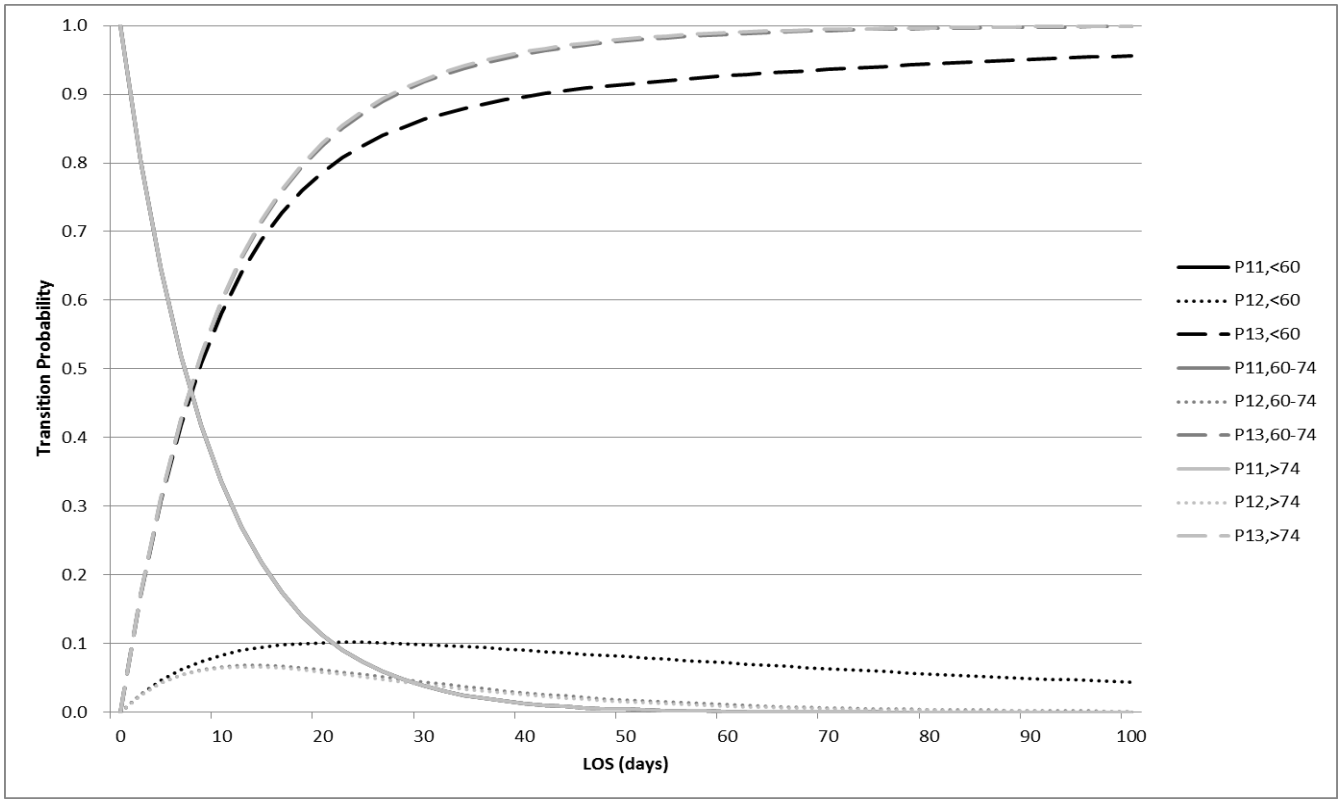


Table 3-1. Demographics of screening participants in Taiwanese national CRC screening program from 2004 to 2013

Characteristics		Number of screening participants	Subjects with positive FIT (%)
Gender	Male	2,041,992	166,784 (8.17)
	Female	2,936,358	150,080 (5.11)
Age (years)	50-54	1,496,838	78,316 (5.23)
	55-59	1,434,402	85,178 (5.94)
	60-64	1,143,251	81,531 (7.13)
	65-69	903,859	71,839 (7.95)
Geographic area	Northern	2,214,345	134,884 (6.09)
	Middle	1,179,006	76,333 (6.47)
	Southern	1,405,194	93,844 (6.68)
	Eastern/offshore islands	179,805	11,803 (6.56)
Type of screening units	Hospital	2,574,431	185,103 (7.19)
	Public health centers	1,954,430	93,811 (4.80)
	Local clinics	449,489	37,950 (8.44)
Urbanization	Main urban	3,787,368	241,452 (6.38)
	Secondary urban	352,888	21,343 (6.05)
	Rural	838,094	54,069 (6.45)
Period	2004-2009	1,254,391	46,151 (3.68)
	2010-2013	3,723,959	270,713 (7.27)
Screening round	Prevalent	3,027,035	201,128 (6.64)
	Subsequent	1,951,315	115,736 (5.93)
Overall		4,978,350	316,864 (6.36)

Table 3-2. Descriptive results of attendees, positive rate, referral rate, the distribution of waiting time (WT)

Year	Number of Attendees	Number of positive attendees	Positive rate	Referral rate		Waiting time			
				Overall	Colonoscopy	Overall		Colonoscopy	
						medium	Q3	medium	Q3
2004	83,756	2,886	3.5	66.7	50.6	26	42	27	43
2005	194,583	6,959	3.6	76.8	60.7	25	44	25	43
2006	210,114	6,576	3.1	82.7	72.8	24	43	24	43
2007	259,450	8,757	3.4	86.4	78.7	24	40	24	40
2008	218,712	7,587	3.5	86.9	77.3	22	35	22	35
2009	287,776	13,386	4.7	84.8	77.3	23	38	23	37
2010	940,241	64,559	6.9	65.7	56.2	25	43	25	43
2011	765,036	57,391	7.5	64.1	57.0	24	42	24	42
2012	1,016,069	72,970	7.2	65.8	59.8	27	47	27	47
2013	1,002,613	75,793	7.6	67.4	63.7	28	51	28	51

Table 3-3. Comparison of referral rate and median WT for colonoscopy in inaugural and rolling out period

Characteristics		Inaugural period (2004-2009)			Rolling out period (2010-2014)		
		No. of subjects referred for colonoscopy (%)	Median WT (days)	IQR (days)	No. of subjects referred for colonoscopy (%)	Median WT (days)	IQR (days)
Gender	Male	15,944 (73.91)	28	55	86,541 (59.59)	46	163
	Female	17,643 (71.78)	28	55	74,328 (59.23)	45	163
Age (years)	50-54	8,296 (72.28)	30	60	40,174 (60.10)	45	163
	55-59	8,901 (73.17)	29	55	43,674 (59.82)	45	163
	60-64	7,587 (72.85)	29	52	42,609 (59.91)	45	163
	65-69	8,803 (72.79)	27	55	34,412 (57.60)	48	162
Geographic area	Northern	14,507 (70.32)	33	72	68,880 (60.29)	43	163
	Middle	8,058 (76.16)	24	47	38,502 (58.56)	45	164
	Southern	9,580 (74.83)	25	41	47,791 (58.97)	48	162
	Eastern/offshore islands	1,442 (67.45)	28	58	5,696 (58.93)	56	161
Type of screening units	Hospital	4,940 (58.08)	51	156	107,348 (60.79)	42	163
	Public health centers	28,528 (76.17)	26	42	35,397 (62.81)	44	162
	Local clinics	119 (62.63)	25	43	18,124 (48.00)	92	161
Urbanization	Main urban	24,137 (73.01)	29	56	124,470 (59.73)	45	163
	Secondary urban	2,357 (71.27)	25	51	10,889 (60.36)	42	164
	Rural	7,093 (72.50)	27	56	25,510 (57.60)	49	163
Screening round	Prevalent	24,612 (70.64)	30	66	92,726 (55.76)	51	162
	Subsequent	8,975 (79.34)	26	37	68,143 (65.26)	39	148
Overall		33,587 (72.78)	28	55	160,869 (59.42)	46	163

Table 3-4. Distributions of discharge types

Type of Discharge	N	Mean of LOS (day)	SD	Min	Max
1 : Discharge	2	2.5	0.71	2	3
3 : Discharge with OPD arranged	123	11.21	14.79	1	93
4 : Death	24	29.86	45.44	3	215
5 : AMAD	22	11	12.58	1	40
6 : Transferred	5	13.2	13.03	3	34
A : AMAD under critical condition	2	22	18.38	9	35

*OPD : outpatient department ; AMAD : Against medical advice discharge

Table 5-1. Results for fitting Coxian phase-type distribution to the simulated data on LOS of Marshall study compared with the original findings

No. of phases	Marshall data			Simulated data		
	Parameters	LOS (days)	BIC	Parameters	LOS (days)	BIC
1	$\hat{\mu}_1 = 0.043$	23	11543	$\hat{\mu}_1 = 0.043$	23	11544
2	$\hat{\mu}_1 = 0.056$	15	11456	$\hat{\mu}_1 = 0.052$	16	11500
	$\hat{\mu}_2 = 0.021$	48		$\hat{\mu}_2 = 0.023$	43	
	$\hat{\lambda}_1 = 0.012$			$\hat{\lambda}_1 = 0.010$		
3	$\hat{\mu}_1 = 0.017$	5	11388	$\hat{\mu}_1 = 0.054$	13	11513
	$\hat{\mu}_2 = 0.119$	5		$\hat{\mu}_2 = 0.032$	30	
	$\hat{\mu}_3 = 0.027$	37		$\hat{\mu}_3 = 0.014$	74	
	$\hat{\lambda}_1 = 0.176$			$\hat{\lambda}_1 = 0.022$		
	$\hat{\lambda}_2 = 0.074$			$\hat{\lambda}_2 = 0.002$		
4	$\hat{\mu}_1 = 0.017$	6	11401	$\hat{\mu}_1 = 0.054$	16	11527
	$\hat{\mu}_2 = 0.125$	5		$\hat{\mu}_2 = 4.8 \times 10^{-20}$	21	
	$\hat{\mu}_3 = 0.003$	7		$\hat{\mu}_3 = 0.041$	21	
	$\hat{\mu}_4 = 0.027$	37		$\hat{\mu}_4 = 0.015$	68	
	$\hat{\lambda}_1 = 0.163$			$\hat{\lambda}_1 = 0.009$		
	$\hat{\lambda}_2 = 0.060$			$\hat{\lambda}_2 = 0.048$		
	$\hat{\lambda}_3 = 0.137$			$\hat{\lambda}_3 = 0.007$		

Table 5-2. Univariate analysis of factors affecting the compliance with colonoscopy and WT for undergoing colonoscopy

Characteristics		Hurdle part		Non-hurdle part		P-value
		Coefficient	OR (95% CI)	Coefficient	RR (95% CI)	
Gender	Male	-0.5440 *	1	-3.7709 *	1	0.3679
	Female	0.0119	1.012 (0.991,1.033)	0.0083	1.008 (0.996,1.021)	
Age (years)	50-54	-0.5628 *	1	-3.7703 *	1	< 0.0001
	55-59	0.0076	1.008 (0.984,1.032)	0.0060	1.006 (0.992,1.020)	
	60-64	0.0185	1.019 (0.995,1.043)	0.0050	1.005 (0.991,1.020)	
	65-69	0.0770	1.080 (1.054,1.107)	0.0017	1.002 (0.987,1.017)	
Geographic area	Northern	-0.5507 *	1	-3.7521 *	1	< 0.0001
	Middle	0.0334	1.034 (1.012,1.057)	0.0370	1.038 (1.024,1.051)	
	Southern	0.0080	1.008 (0.988,1.029)	-0.0680	0.934 (0.923,0.946)	
	Eastern/offshore islands	0.0492	1.050 (1.004,1.100)	-0.0735	0.929 (0.903,0.956)	
Type of	Hospital	0.3353	1.398 (1.369,1.428)	-0.0912	0.913 (0.902,0.924)	< 0.0001
screening units	Public health centers	-0.8399 *	1	-3.6916 *	1	
	Local clinic	0.8226	2.276 (2.208,2.347)	-0.2178	0.804 (0.788,0.821)	
Urbanization	Main urban	0.0284	1.029 (0.992,1.067)	-0.0320	0.968 (0.948,0.990)	< 0.0001
	Secondary urban	-0.5742 *	1	-3.7415 *	1	
	Rural	0.0829	1.086 (1.043,1.132)	-0.0055	0.995 (0.970,1.020)	
Period	2004-2009	-1.0216 *	1	-3.5508 *	1	< 0.0001
	2010-2013	0.5585	1.748 (1.694,1.804)	-0.2551	0.775 (0.762,0.788)	
Screening round	Prevalent	0.3688	1.446 (1.415,1.478)	-0.0083	0.992 (0.979,1.004)	< 0.0001
	Subsequent	-0.7768 *	1	-3.7620 *	1	

***** : intercept

Table 5-3. Model Selection for the hurdle regression model for the possible interaction assessment of putative factors

Types of Model (additional variables)	**df**	**AIC**	**P-value**
H : None N : None	28	415618	
H : None N : (Exclude gender)	27	415615	
H : period*unittype N : (Exclude gender)	29	414780	<0.0001
H : period*unittype、period*area N : (Exclude gender)	32	414693	<0.0001
H : period*unittype、period*area N : period*unittype (Exclude gender)	34	413693	<0.0001
H : period*unittype、period*area N : period*unittype、period*area (Exclude gender)	37	413605	<0.0001
H : period*unittype、period*area N : period*unittype、period*area、period*urban (Exclude gender)	39	413578	<0.0001

All models contain gender, age, area (Geographic area), unittype (Type of screening units), urban (Urbanization), subs (Screening round) and period effect.

H : Hurdle part

N : Non-hurdle part

Table 5-4. Multivariate analysis on main effect and interaction of factors affecting the non-compliance with colonoscopy

Characteristics				Coefficient	aOR (95% CI)	P-value
Gender	male			-1.0836 *	1	<0.0001
	female			0.0744	1.077 (1.061,1.093)	
Age	50-54			-1.0836 *	1	<0.0001
	55-59			0.1050	1.111 (1.088,1.134)	
	60-64			0.1369	1.147 (1.123,1.171)	
	65-69			0.2432	1.275 (1.247,1.303)	
Urbanization	Main urban			-1.0836 *	1	<0.0001
	Secondary urban			0.0343	1.035 (1.004,1.066)	
	Rural			0.1336	1.143 (1.117,1.169)	
Screening round	Prevalent			0.4464	1.563 (1.537,1.588)	<0.0001
	Subsequent			-1.0836 *	1	
Period	2004-2009	Geographic area	Northern	-1.8200 **	1	<0.0001
			Middle	-0.0132	0.987 (0.927,1.046)	
			Southern	0.1534	1.166 (1.099,1.233)	
			Eastern/offshore islands	0.4115	1.509 (1.355,1.664)	
		Type of screening units	Hospital	0.9326	2.541 (2.394,2.688)	<0.0001
			Public health centers	-1.8200 **	1	
			Local clinic	0.5400	1.716 (1.200,2.232)	
	2010-2013	Geographic area	Northern	-1.0836 *	1	
			Middle	0.0761	1.079 (1.057,1.101)	
			Southern	0.0360	1.037 (1.017,1.056)	
			Eastern/offshore islands	0.0051	1.005 (0.957,1.054)	
		Type of screening units	Hospital	0.0768	1.080 (1.058,1.102)	
			Public health centers	-1.0836 *	1	
			Local clinic	0.5839	1.793 (1.744,1.842)	

* : intercept ; ** : intercept and period effect

Table 5-5. Multivariate analysis of main effect and interaction of factors affecting WT for undergoing colonoscopy

Characteristics				Coefficient	aRR (95% CI)	P-value
Age	50-54			0.0217	1.022 (1.009,1.035)	0.0460
	55-59			0.0181	1.018 (1.006,1.031)	
	60-64			0.0198	1.020 (1.007,1.033)	
	65-69			-3.7554 *	1	
Screening	Prevalent			-3.7554 *	1	<0.0001
round	Subsequent			0.0364	1.037 (1.027,1.047)	
Period	2004-2009	Geographic area	Northern	0.0321	1.033 (0.968,1.097)	<0.0001
			Middle	0.1276	1.136 (1.069,1.203)	
			Southern	0.0770	1.080 (1.014,1.147)	
			Eastern/offshore islands	-3.7554 *	1	
		Type of screening	Hospital	-0.2038	0.816 (0.657,0.975)	<0.0001
		units	Public health centers	0.2006	1.222 (0.987,1.458)	
			Local clinic	-3.7554 *	1	
		Urbanization	Main urban	-3.7554 *	1	<0.0001
			Secondary urban	-0.0227	0.978 (0.936,1.019)	
			Rural	-0.0567	0.945 (0.915,0.975)	
	2010-2013	Geographic area	Northern	0.1160	1.123 (1.089,1.157)	
			Middle	0.1197	1.127 (1.094,1.161)	
			Southern	0.0080	1.008 (0.977,1.039)	
			Eastern/offshore islands	-4.0239 **	1	
		Type of screening	Hospital	0.1278	1.136 (1.119,1.154)	
		units	Public health centers	0.0715	1.074 (1.055,1.093)	
			Local clinic	-4.0239 **	1	
		Urbanization	Main urban	-4.0239 **	1	
			Secondary urban	0.0660	1.068 (1.047,1.089)	
			Rural	0.0234	1.024 (1.008,1.040)	

* : intercept ; ** : intercept and period effect

Table 5-6. The estimated results of Coxian phase-type models

No. of phases of Coxian phase-type distribution	Parameters (SD)	BIC
1	$\hat{\nu} = 0.00021\ (9.7 \times 10^{-7})$ (arrival rate) $\hat{p} = 0.26472\ (0.00205)$ (non-compliance) $\hat{\mu}_1 = 0.02870\ (0.00016)$ (referral rate)	769183
2	$\hat{\nu} = 0.00021\ (9.7 \times 10^{-7})$ (arrival rate) $\hat{p} = 0.26472\ (0.00205)$ (non-compliance) $\hat{\mu}_1 = 0.03040\ (0.00019)$ (referral rate) $\hat{\mu}_2 = 0.00590\ (0.00046)$ (referral rate) $\hat{\lambda}_1 = 0.00043\ (0.00006)$ (transition rate)	768284
3	$\hat{\nu} = 0.00021\ (9.7 \times 10^{-7})$ (arrival rate) $\hat{p} = 0.26472\ (0.00205)$ (non-compliance) $\hat{\mu}_1 = 0.03037\ (0.00019)$ (referral rate) $\hat{\mu}_2 = 0\ (0.00701)$ (referral rate) $\hat{\mu}_3 = 0.00633\ (0.00062)$ (referral rate) $\hat{\lambda}_1 = 0.00031\ (0.00011)$ (transition rate) $\hat{\lambda}_2 = 0.01708\ (0.00899)$ (transition rate)	768308

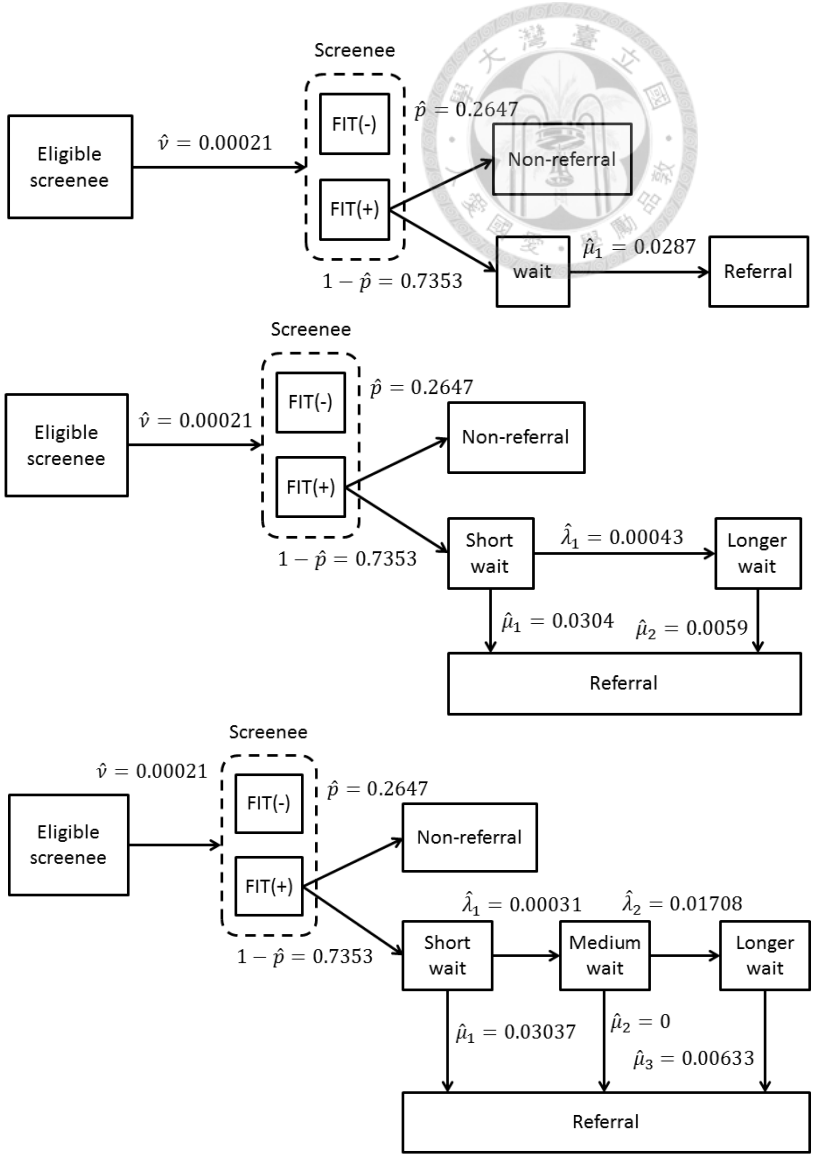


Table 5-7. The expected WT calculated with queue hurdle Coxian two-phase phase-type model

No. of FIT (+) phases	**Expected WT in phase *i* (days)**	**Marginal Expected WT (days)**
2	WT_1=32 WT_2=169	35

Table 5-8. Estimated results of queue hurdle one-phase Coxian phase-type model with the covariate of risk score affecting WT for the referral of colonoscopy

No. of phases of Coxian phase-type distribution	**Parameters (SD)**	**Expected WT (days)**	**BIC**
1	$\hat{v} = 0.00021\ (9.7 \times 10^{-7})$ (arrival rate) $\hat{p} = 0.26471\ (0.00205)$ (non-compliance) $\hat{\mu}_{01} = 0.02604\ (0.00021)$ $\hat{\beta}_1 = 0.19073\ (0.01096)$	WT_{low}=38 WT_{high}=32	768868

Table 5-9. Estimated results of queue hurdle two-phase Coxian phase-type model with the covariate of risk score affecting WT for the referral of colonoscopy

Parameter with score	λ_1	μ_1	μ_2
Parameters estimate (SD)	$\hat{v} = 0.00021\ (9.7 \times 10^{-7})$ $\hat{p} = 0.26477\ (0.00206)$ $\hat{\mu}_1 = 0.03042\ (0.00019)$ $\hat{\mu}_2 = 0.00609\ (0.00047)$ $\hat{\lambda}_{01} = 0.00055\ (0.00008)$ $\hat{\beta}_1 = -0.40380\ (0.17038)$	$\hat{v} = 0.00021\ (9.7 \times 10^{-7})$ $\hat{p} = 0.26472\ (0.00205)$ $\hat{\mu}_{01} = 0.02745\ (0.00023)$ $\hat{\mu}_2 = 0.00597\ (0.00046)$ $\hat{\lambda}_1 = 0.00043\ (0.00006)$ $\hat{\beta}_1 = 0.19684\ (0.01134)$	$\hat{v} = 0.00021\ (9.7 \times 10^{-7})$ $\hat{p} = 0.26472\ (0.00205)$ $\hat{\mu}_1 = 0.03039\ (0.00019)$ $\hat{\mu}_{02} = 0.00573\ (0.00051)$ $\hat{\lambda}_1 = 0.00043\ (0.00006)$ $\hat{\beta}_1 = 0.08482\ (0.09433)$
BIC	768292	767997	768297
Expected WT in phase *i* (days)	$WT_{1,low}$=32 $WT_{1,high}$=32 $WT_{2,low}$=164 $WT_{2,high}$=164	$WT_{1,low}$=36 $WT_{1,high}$=30 $WT_{2,low}$=167 $WT_{2,high}$=167	$WT_{1,low}$=32 $WT_{1,high}$=32 $WT_{2,low}$=175 $WT_{2,high}$=161
Marginal Expected WT (days)	WT_{low}=35 WT_{high}=34	WT_{low}=38 WT_{high}=32	WT_{low}=35 WT_{high}=35

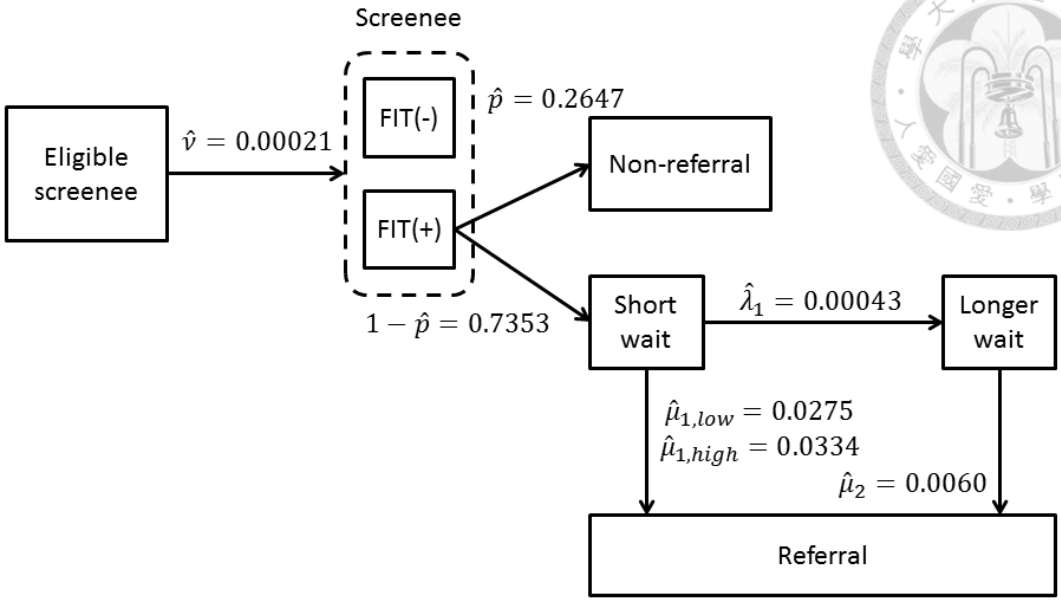

Screenee
Eligible screenee
$\hat{v} = 0.00021$
FIT(-)
FIT(+)
$\hat{p} = 0.2647$
Non-referral
$1 - \hat{p} = 0.7353$
Short wait
$\hat{\lambda}_1 = 0.00043$
Longer wait
$\hat{\mu}_{1,low} = 0.0275$
$\hat{\mu}_{1,high} = 0.0334$
$\hat{\mu}_2 = 0.0060$
Referral

Table 5-10. Estimated results of fitting Coxian phase-type distribution to SKH data set

No. of phases	Parameters (SD)	BIC
1	$\hat{\mu}_1 = 0.0726\ (0.0054)$	1295
2	$\hat{\mu}_1 = 0.0933\ (0.0087)$ $\hat{\mu}_2 = 0.0189\ (0.0082)$ $\hat{\lambda}_1 = 0.0073\ (0.0049)$	1280
3	$\hat{\mu}_1 = 0.0932\ (0.0089)$ $\hat{\mu}_2 = 0.0187\ (0.0171)$ $\hat{\mu}_3 = 0.0204\ (0.0945)$ $\hat{\lambda}_1 = 0.0072\ (0.0063)$ $\hat{\lambda}_2 = 0.0017\ (0.1005)$	1157
4	$\hat{\mu}_1 = 0.0935\ (0.0096)$ $\hat{\mu}_2 = 0.0220\ (0.0350)$ $\hat{\mu}_3 = 3.95 \times 10^{-20}\ (0.0675)$ $\hat{\mu}_4 = 0.0231\ (0.0653)$ $\hat{\lambda}_1 = 0.0081\ (0.0109)$ $\hat{\lambda}_2 = 0.0025\ (0.0241)$ $\hat{\lambda}_3 = 0.0231\ (0.0836)$	1158

Table 5-11. The expected LOS in phase i (days) among the three-phase Coxian Phase-type models

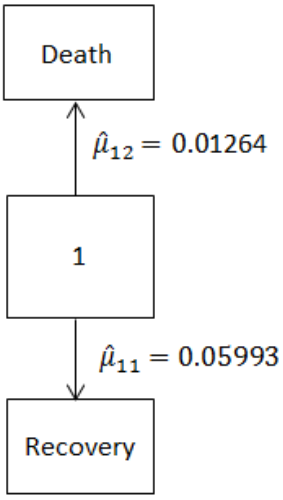
No. of phases	Expected LOS in phase *i* (days)	Marginal Expected LOS (days)
3	LOS_1=10 LOS_2=49 LOS_3=49	14

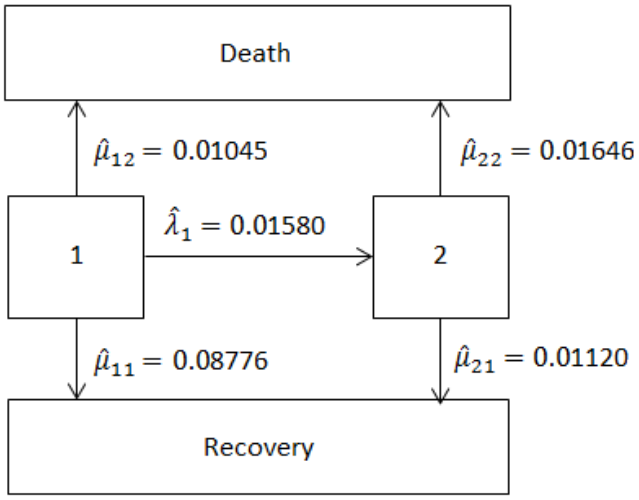
Table 5-12. The comparison of two 3-phase Coxian models assuming three and two absorbing rates

	Original Model (3-phase model) (three absorbing rates)			**Alternative Model (two absorbing rates)**		
No. of phases	**Parameters**	**LOS**	**AIC**	**Parameters**	**LOS**	**AIC**
3	$\hat{\mu}_1 = 0.0932$	10	1141	$\hat{\mu}_1 = 0.0933$	10	1273
		49			34	
	$\hat{\mu}_2 = 0.0187$	49		$\hat{\mu}_2 = \hat{\mu}_3 = 0.0189$	53	
	$\hat{\mu}_3 = 0.0204$			$\hat{\lambda}_1 = 0.0073$		
	$\hat{\lambda}_1 = 0.0072$			$\hat{\lambda}_2 = 0.0104$		
	$\hat{\lambda}_2 = 0.0017$					

Table 5-13. Model selections for Coxian phase-type model

No. of phases	LOS (days)	BIC
1	LOS=14	1470
2	LOS_1=9 LOS_2=36	1451
3	LOS_1=6 LOS_2=2 LOS_3=30	1439





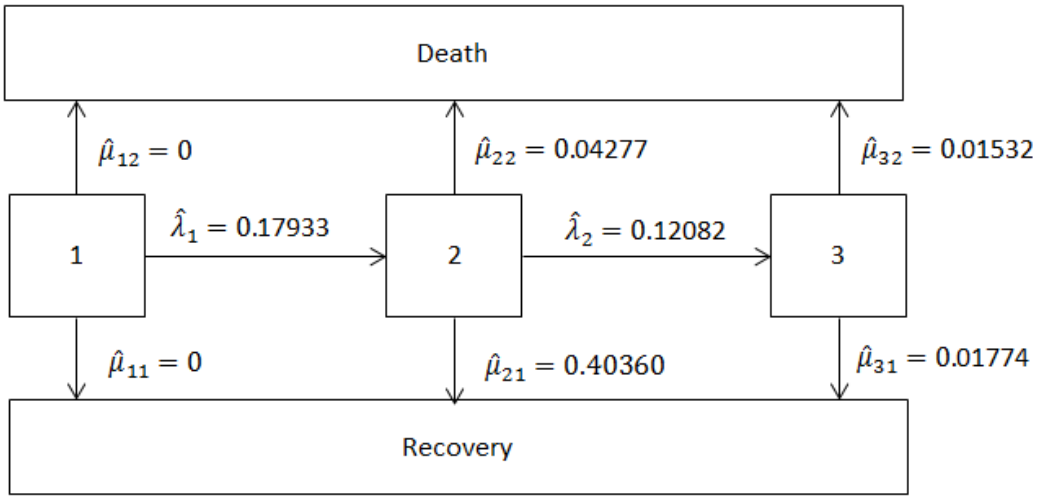
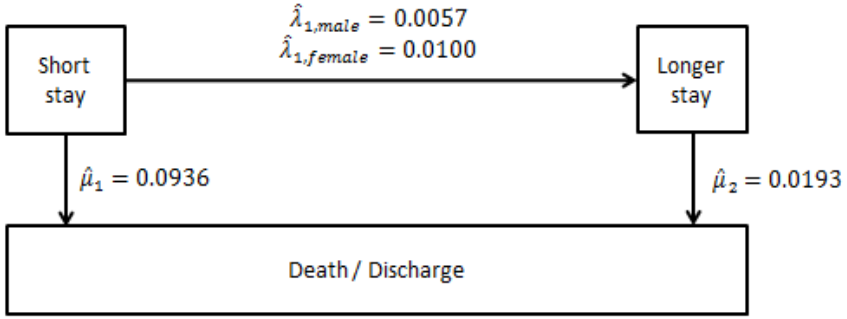


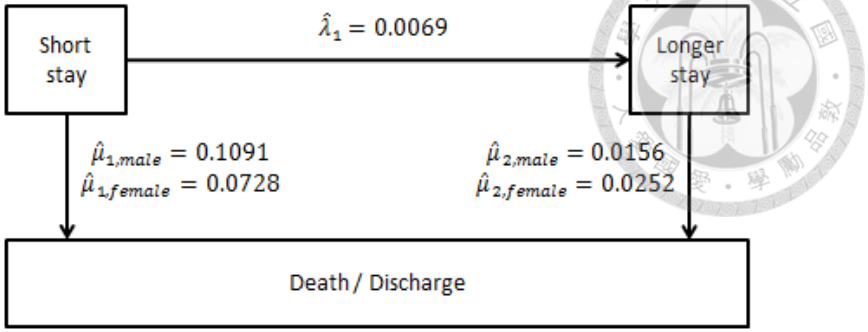
Table 5-14. Descriptive results of length of stay (LOS) by gender

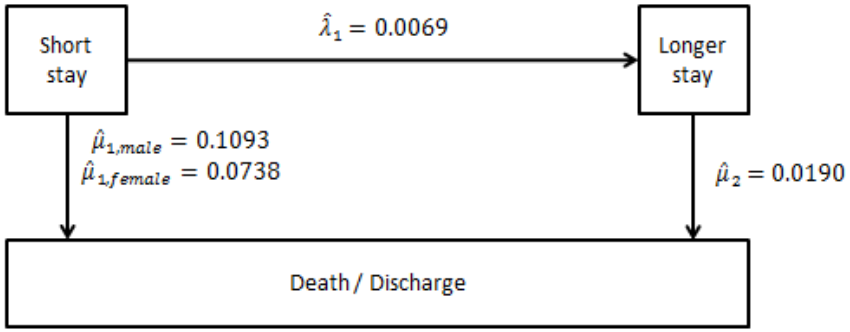
Variable	N	Mean of LOS (day)	Median of LOS (day)	SD	Min	Max
Female	70	16.04	9	19.21	1	103
Male	108	12.31	5	23.68	1	215
Total	178	13.78	7	22.05	1	215

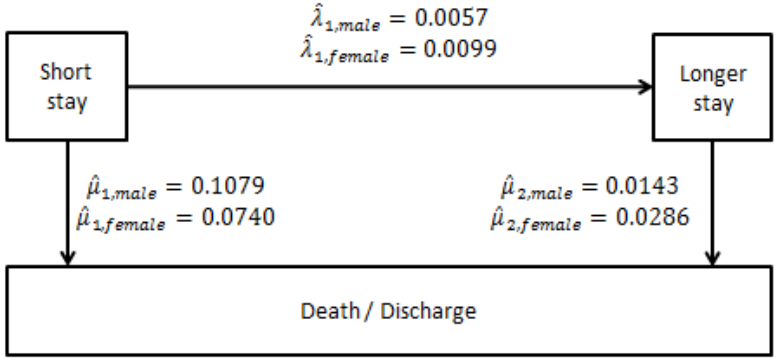
Table 5-15. Estimated results on transition rates and regression coefficients regarding the effect of gender in two-phase Coxian phase-type model

Parameter with gender	λ_1	μ_1	μ_2	μ_1, μ_2	μ_1, μ_2, λ_1
Parameters estimate (SD)	$\hat{\mu}_1 = 0.0936$ (0.0087) $\hat{\mu}_2 = 0.0193$ (0.0083) $\hat{\lambda}_{01} = 0.0057$ (0.0048) $\hat{\beta}_1 = 0.5629$ (0.9514)	$\hat{\mu}_{01} = 0.1093$ (0.0128) $\hat{\beta}_1 = -0.3933$ (0.1761) $\hat{\mu}_2 = 0.0190$ (0.0087) $\hat{\lambda}_1 = 0.0069$ (0.0052)	$\hat{\mu}_1 = 0.0930$ (0.0086) $\hat{\mu}_{02} = 0.0155$ (0.0094) $\hat{\beta}_1 = 0.3772$ (0.6747) $\hat{\lambda}_1 = 0.0071$ (0.0047)	$\hat{\mu}_{01} = 0.1091$ (0.0128) $\hat{\beta}_1 = -0.4044$ (0.1798) $\hat{\mu}_{02} = 0.0156$ (0.0093) $\hat{\beta}_2 = 0.4805$ (0.6969) $\hat{\lambda}_1 = 0.0069$ (0.0052)	$\hat{\mu}_{01} = 0.1079$ (0.0127) $\hat{\beta}_1 = -0.3769$ (0.1950) $\hat{\mu}_{02} = 0.0143$ (0.0094) $\hat{\beta}_2 = 0.6948$ (0.9076) $\hat{\lambda}_{01} = 0.0057$ (0.0053) $\hat{\beta}_3 = 0.5532$ (1.5330)
BIC	1285	1280	1284	1285	1290
Mena LOS in phase *i* (days)	$LOS_{1,male}$=10 $LOS_{1,female}$=10 $LOS_{2,male}$=52 $LOS_{2,female}$=52	$LOS_{1,male}$=9 $LOS_{1,female}$=12 $LOS_{2,male}$=53 $LOS_{2,female}$=53	$LOS_{1,male}$=10 $LOS_{1,female}$=10 $LOS_{2,male}$=65 $LOS_{2,female}$=44	$LOS_{1,male}$=9 $LOS_{1,female}$=13 $LOS_{2,male}$=64 $LOS_{2,female}$=40	$LOS_{1,male}$=9 $LOS_{1,female}$=12 $LOS_{2,male}$=70 $LOS_{2,female}$=35
Marginal Mean LOS (days)	LOS_{male}=13 LOS_{female}=15	LOS_{male}=12 LOS_{female}=17	LOS_{male}=15 LOS_{female}=13	LOS_{male}=12 LOS_{female}=16	LOS_{male}=12 LOS_{female}=16


Short stay
$\hat{\lambda}_{1,male} = 0.0057$
$\hat{\lambda}_{1,female} = 0.0100$
Longer stay
$\hat{\mu}_1 = 0.0936$
$\hat{\mu}_2 = 0.0193$
Death / Discharge


Short stay
$\hat{\lambda}_1 = 0.0069$
Longer stay
$\hat{\mu}_{1,male} = 0.1091$
$\hat{\mu}_{1,female} = 0.0728$
$\hat{\mu}_{2,male} = 0.0156$
$\hat{\mu}_{2,female} = 0.0252$
Death / Discharge


Short stay
$\hat{\lambda}_1 = 0.0069$
Longer stay
$\hat{\mu}_{1,male} = 0.1093$
$\hat{\mu}_{1,female} = 0.0738$
$\hat{\mu}_2 = 0.0190$
Death / Discharge


Short stay
$\hat{\lambda}_{1,male} = 0.0057$
$\hat{\lambda}_{1,female} = 0.0099$
Longer stay
$\hat{\mu}_{1,male} = 0.1079$
$\hat{\mu}_{1,female} = 0.0740$
$\hat{\mu}_{2,male} = 0.0143$
$\hat{\mu}_{2,female} = 0.0286$
Death / Discharge

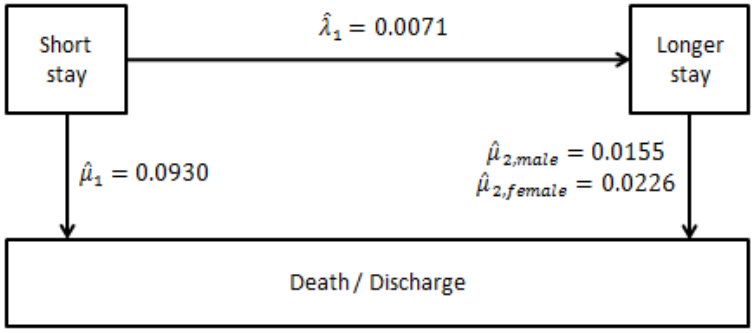

Short stay
$\hat{\lambda}_1 = 0.0071$
Longer stay
$\hat{\mu}_1 = 0.0930$
$\hat{\mu}_{2,male} = 0.0155$
$\hat{\mu}_{2,female} = 0.0226$
Death / Discharge

Table 5-16. Descriptive results of length of stay (LOS) by age

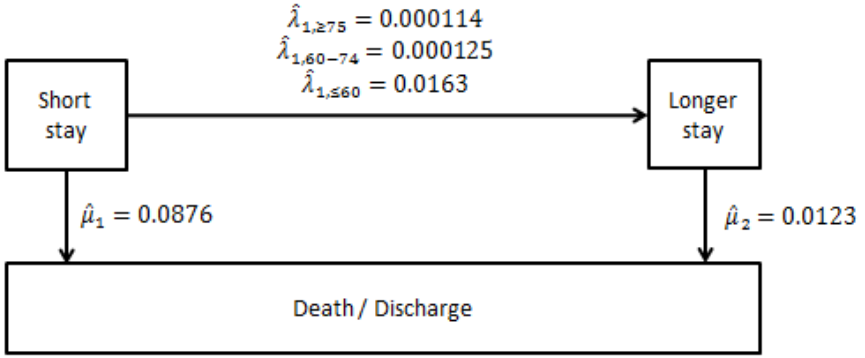
Variable	N	Mean of LOS (day)	Median of LOS (day)	SD	Min	Max
< 60	53	17.45	7	35.29	1	215
60-74	65	10.18	5	12.54	1	73
> 74	60	14.43	8.5	12.80	2	60
Total	178	13.78	7	22.05	1	215

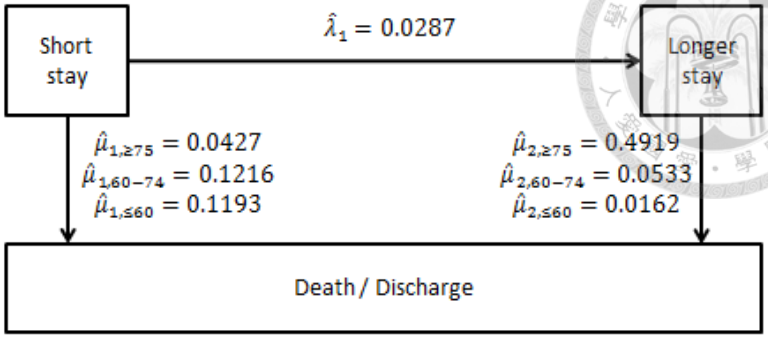
Table 5-17. Estimated results on transition rates and regression coefficients regarding the effect of age in two-phase Coxian phase-type model

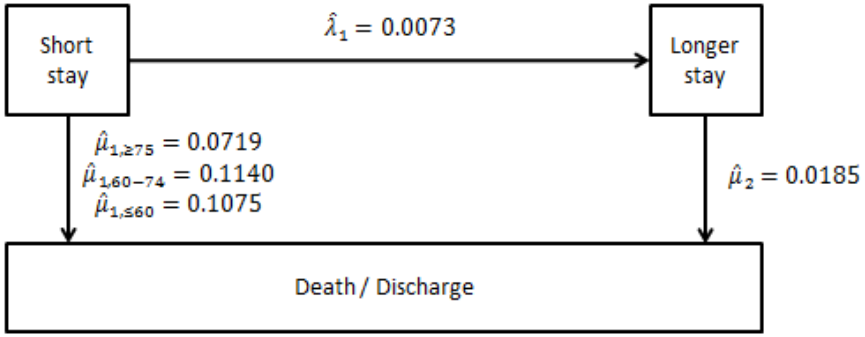
Parameter with gender	λ_1	μ_1	μ_2	μ_1, μ_2	μ_1, μ_2, λ_1
Parameters estimate (SD)	$\hat{\mu}_1 = 0.0876$ (0.0079) $\hat{\mu}_2 = 0.0123$ (0.0070) $\hat{\lambda}_{01} = 0.000114$ (0.0005) $\hat{\beta}_1 = 0.0913$ (25.1779) $\hat{\beta}_2 = 4.7608$ (4.1164)	$\hat{\mu}_{01} = 0.0719$ (0.0101) $\hat{\beta}_1 = 0.4611$ (0.1980) $\hat{\beta}_2 = 0.4024$ (0.2122) $\hat{\mu}_2 = 0.0185$ (0.0080) $\hat{\lambda}_1 = 0.0073$ (0.0046)	$\hat{\mu}_1 = 0.0943$ (0.0093) $\hat{\mu}_{02} = 0.0569$ (0.0245) $\hat{\beta}_1 = -0.0851$ (0.6803) $\hat{\beta}_2 = -1.5174$ (0.6416) $\hat{\lambda}_1 = 0.0147$ (0.0086)	$\hat{\mu}_{01} = 0.0427$ (0.0224) $\hat{\beta}_1 = 1.0465$ (0.5750) $\hat{\beta}_2 = 1.0274$ (0.6304) $\hat{\mu}_{02} = 0.4919$ (0.3501) $\hat{\beta}_3 = -2.2216$ (0.8305) $\hat{\beta}_4 = -3.4105$ (0.8504) $\hat{\lambda}_1 = 0.0287$ (0.0217)	$\hat{\mu}_{01} = 0.000033$ (0.000013) $\hat{\beta}_1 = 8.1875$ (0.3904) $\hat{\beta}_2 = 8.1350$ (0.3992) $\hat{\mu}_{02} = 0.4736$ (0.2038) $\hat{\beta}_3 = -2.3524$ (0.7621) $\hat{\beta}_4 = -3.6016$ (0.6529) $\hat{\lambda}_{01} = 0.0811$ (0.0119) $\hat{\beta}_5 = -1.5406$ (1.3437) $\hat{\beta}_6 = -1.6169$ (0.6485)
BIC	1285	1284	1283	1283	1286

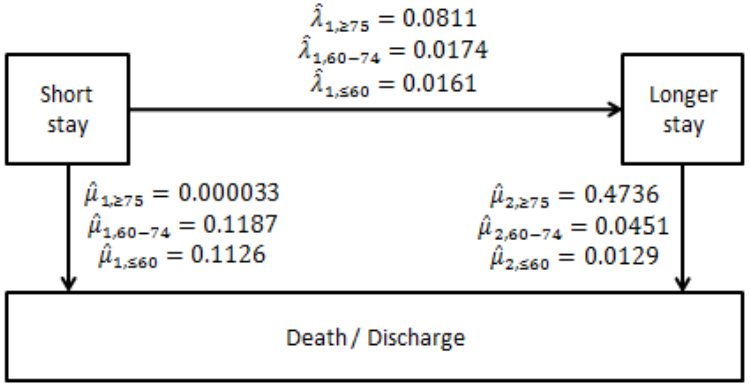
Table 5-17 (Continued)

Parameter with gender	λ_1	μ_1	μ_2	μ_1 , μ_2	μ_1 , μ_2 , λ_1
Mena LOS in phase *i* (days)	$LOS_{1,>74}$=11	$LOS_{1,>74}$=13	$LOS_{1,>74}$=9	$LOS_{1,>74}$=14	$LOS_{1,>74}$=12
	$LOS_{1,60\text{-}74}$=11	$LOS_{1,60\text{-}74}$=8	$LOS_{1,60\text{-}74}$=9	$LOS_{1,60\text{-}74}$=7	$LOS_{1,60\text{-}74}$=7
	$LOS_{1,<60}$=10	$LOS_{1,<60}$=9	$LOS_{1,<60}$=9	$LOS_{1,<60}$=7	$LOS_{1,<60}$=8
	$LOS_{2,>74}$=81	$LOS_{2,>74}$=54	$LOS_{2,>74}$=18	$LOS_{2,>74}$=2	$LOS_{2,>74}$=2
	$LOS_{2,60\text{-}74}$=81	$LOS_{2,60\text{-}74}$=54	$LOS_{2,60\text{-}74}$=19	$LOS_{2,60\text{-}74}$=19	$LOS_{2,60\text{-}74}$=22
	$LOS_{2,<60}$=81	$LOS_{2,<60}$=54	$LOS_{2,<60}$=80	$LOS_{2,<60}$=62	$LOS_{2,<60}$=78
Marginal Mean LOS (days)	$LOS_{>74}$=12	$LOS_{>74}$=18	$LOS_{>74}$=12	$LOS_{>74}$=15	$LOS_{>74}$=14
	$LOS_{60\text{-}74}$=12	$LOS_{60\text{-}74}$=12	$LOS_{60\text{-}74}$=12	$LOS_{60\text{-}74}$=10	$LOS_{60\text{-}74}$=10
	$LOS_{<60}$=22	$LOS_{<60}$=12	$LOS_{<60}$=20	$LOS_{<60}$=19	$LOS_{<60}$=17


$\hat{\lambda}_{1,\geq 75} = 0.000114$
$\hat{\lambda}_{1,60-74} = 0.000125$
$\hat{\lambda}_{1,\leq 60} = 0.0163$
Short stay
Longer stay
$\hat{\mu}_1 = 0.0876$
$\hat{\mu}_2 = 0.0123$
Death / Discharge


Short stay
$\hat{\lambda}_1 = 0.0287$
Longer stay
$\hat{\mu}_{1,\geq 75} = 0.0427$
$\hat{\mu}_{1,60-74} = 0.1216$
$\hat{\mu}_{1,\leq 60} = 0.1193$
$\hat{\mu}_{2,\geq 75} = 0.4919$
$\hat{\mu}_{2,60-74} = 0.0533$
$\hat{\mu}_{2,\leq 60} = 0.0162$
Death / Discharge


Short stay
$\hat{\lambda}_1 = 0.0073$
Longer stay
$\hat{\mu}_{1,\geq 75} = 0.0719$
$\hat{\mu}_{1,60-74} = 0.1140$
$\hat{\mu}_{1,\leq 60} = 0.1075$
$\hat{\mu}_2 = 0.0185$
Death / Discharge


$\hat{\lambda}_{1,\geq 75} = 0.0811$
$\hat{\lambda}_{1,60-74} = 0.0174$
$\hat{\lambda}_{1,\leq 60} = 0.0161$
Short stay
Longer stay
$\hat{\mu}_{1,\geq 75} = 0.000033$
$\hat{\mu}_{1,60-74} = 0.1187$
$\hat{\mu}_{1,\leq 60} = 0.1126$
$\hat{\mu}_{2,\geq 75} = 0.4736$
$\hat{\mu}_{2,60-74} = 0.0451$
$\hat{\mu}_{2,\leq 60} = 0.0129$
Death / Discharge

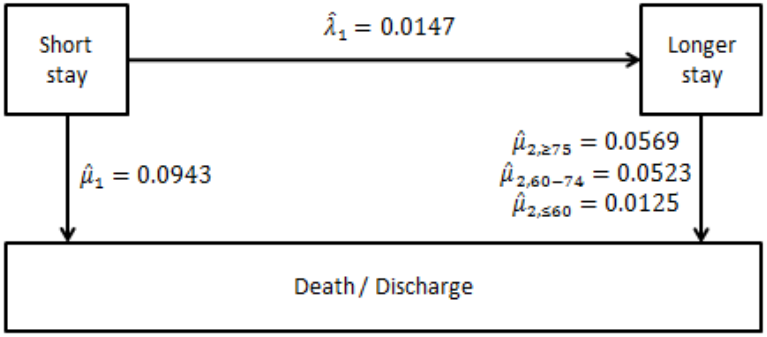

Short stay
$\hat{\lambda}_1 = 0.0147$
Longer stay
$\hat{\mu}_1 = 0.0943$
$\hat{\mu}_{2,\geq 75} = 0.0569$
$\hat{\mu}_{2,60-74} = 0.0523$
$\hat{\mu}_{2,\leq 60} = 0.0125$
Death / Discharge

Table 6-1. The estimated results of Coxian phase-type models with three approaches

Method	**Newton-Raphson**		**Quasi-Newton**		**Nelder-Mead Simplex**	
No. of phases	**Parameters**	**BIC**	**Parameters**	**BIC**	**Parameters**	**BIC**
1	$\hat{\mu}_1 = 0.0726$	1300	$\hat{\mu}_1 = 0.0726$	1300	$\hat{\mu}_1 = 0.0726$	1300
2	$\hat{\mu}_1 = 0.0933$	1285	$\hat{\mu}_1 = 0.0933$	1285	$\hat{\mu}_1 = 0.0933$	1285
	$\hat{\mu}_2 = 0.0189$		$\hat{\mu}_2 = 0.0189$		$\hat{\mu}_2 = 0.0189$	
	$\hat{\lambda}_1 = 0.0073$		$\hat{\lambda}_1 = 0.0073$		$\hat{\lambda}_1 = 0.0073$	
3	$\hat{\mu}_1 = 0.0932$	1162	$\hat{\mu}_1 = 0.0931$	1296	$\hat{\mu}_1 = 3.38 \times 10^{-7}$	1272
	$\hat{\mu}_2 = 0.0187$		$\hat{\mu}_2 = 0.0147$		$\hat{\mu}_2 = 0.4291$	
	$\hat{\mu}_3 = 0.0204$		$\hat{\mu}_3 = 0.0217$		$\hat{\mu}_3 = 0.0336$	
	$\hat{\lambda}_1 = 0.0072$		$\hat{\lambda}_1 = 0.0066$		$\hat{\lambda}_1 = 0.1855$	
	$\hat{\lambda}_2 = 0.0017$		$\hat{\lambda}_2 = 0.0172$		$\hat{\lambda}_2 = 0.1218$	
4	$\hat{\mu}_1 = 0.0935$	1163			$\hat{\mu}_1 = 0$	1282
	$\hat{\mu}_2 = 0.0220$				$\hat{\mu}_2 = 0.4558$	
	$\hat{\mu}_3 = 3.95 \times 10^{-20}$				$\hat{\mu}_3 = 1.11 \times 10^{-6}$	
	$\hat{\mu}_4 = 0.0231$				$\hat{\mu}_4 = 0.2617$	
	$\hat{\lambda}_1 = 0.0081$				$\hat{\lambda}_1 = 0.1755$	
	$\hat{\lambda}_2 = 0.0025,$				$\hat{\lambda}_2 = 0.1031$	
	$\hat{\lambda}_3 = 0.0231$				$\hat{\lambda}_3 = 0.0330$	