Health State Transition Probability and Long-Term Care Cost Estimation

Xiaodong Cui
Nanjing Xiaozhuang University, Nanjing, Jiangsu, China

Keywords: long-term care cost, State Transition Probability, multistate model, Markov process.

Abstract: Long-term care security system is an inevitable choice to transfer long-term care risks in the context of aging. This study is aimed to provide a mathematic basis for the long-term care cost estimation. Based on the evaluation of the existing models, this paper establish multistate estimate model with piecewise constant transition probability matrices. Using the sample from the latest two waves of Chinese Longitudinal Healthy and Longevity Survey, we construct dynamic transition matrices based multistate Markov model, which avoid static assumptions limitations and the subjectivity of model specification in regression; Using transition intensity piecewise constant we forest the future transition matrix, which can solve Markov time homogeneity hypothesis; With actuarial theory and empirical life table data, the individual long term care costs are derived and calculated. The estimation method constructed in this paper has the advantage of theoretically relaxing the estimate hypothesis and improving the prediction accuracy, and at the same time, the gender and age characteristics of the measured results are consistent with the health theory.

1. Introduction

Long-term care security system means providing payment for medical care or other service fees for the insured in need of long-term care. With the global aging and senility wave, especially in China, family miniaturization and empty nest is changing the traditional home-based care way[1], long-term care system becomes one of effective means of transferring the nursing care risk. Meanwhile, as the actuarial basis for long-term care system, how to reasonably estimate cost and health transition probability matrix have been a vivid academic debate and becomes the purpose of this paper. In this paper, on the basis of evaluating and analyzing the existing models, a new estimating model has been built as follows. Section 1 review the existing estimating model. Section 2 studies a new model and deduces premium formulas. Section 3 are data and process. Section 4 concludes.

2. Review of the Existing Estimating Model

The major consideration for long-term care cost estimation is the health state of the insured, so the key of estimating is the state transition probability matrix. Now there are static and dynamic estimating methods regarding whether to consider health state transition.
2.1. Manchester Method

The Manchester method is not only a traditional method of calculating long-term care cost in early practice, but also a typical static estimating method [2],[3],[4]. Its significant characteristic is the assumption that health state transition is irreversible and states cannot be transferred to each other. In other words, when an individual is transferred from one state to another, that individual will continue to be in another state. This characteristic essentially enables the possible cash flow compensated for in the future to be deemed as an annuity by the Manchester method; according to the occurrence rate and duration of nursing care reflected from empirical data, the annuity method is used to calculate premium. The Manchester method is theoretically simple, convenient for calculation and undemanding for data information, but it still has the following defects: First, the health state of the insured during the insurance period is assumed to be non-transferable or irreversible. This hypothesis simplifies calculation but does not conform to the reality, and has a great impact on estimating especially in the case of paying different amount according to different nursing care state. Second, in the Manchester method, the possibility of transition is assumed to unrelated to the previous health state, which is arbitrary and unitary. Third, though approximate calculation may be adopted in practice, the estimation of central disability rate is similar in approximate calculation. As a result, the final calculation accuracy may be relatively low [5],[6].

2.2. Multistate Estimating Model

A multistate model is composed of a certain number of states, and there are transition probabilities and transition intensities between different states. Long-term care cost is related to multiple states and state transition, therefore, theoretically the multistate model is an ideal model for estimating of long-term care cost. The multistate estimating basis is the estimation transition probabilities. Hence, the following two estimating methods are generated.

2.2.1. Transition Probability Matrix Based on Regression Simulation Method

This method is based on micro-data and population diversity, and utilizes the generalized linear model to estimate the health state transition intensity [7],[8],[9],[10]. Usually, the method to estimate state transition probability inherits the flexible form of the generalized linear model, can fit and smooth the transition intensity of health state. However, the model application has two important problems: First, Due to the complexity of changes in health state, model forms and variable selection have diversity and subjectivity. Second, the model is related to a lot of microcosmic information at the individual level, but much microcosmic information can hardly be wholly obtained. Therefore, apart from the calculation method of the first-stage transition probability which is different, it essentially has no much difference from the static estimating method in other aspects.

2.2.2. Transition Probability Matrix Based on Markov Process

Theoretically, the construction of a transition probability matrix means tracking the health changes of individuals based on the continuously observed data, and investigating the dynamic changes of the health state of the same age group; and it is deemed as a probability matrix method based on the Markov process [11],[12],[13],[14]. However, due to the data requirements or complex calculations, relevant studies usually assume that the transition intensity is constant, namely only related to the initial state and the arrival state. In this case, Markov process is said to time invariance.

For time invariant hypothesis, it is convenient for expressing the transition probability function, but inappropriate in many applications. For example, if the health state transition probability is
time-invariant, it means that no matter at the age of 60, 70 or 80, as long as the observed transition time interval is the same, the transition probability of health state will be the same, which is not in conformity with the reality [15],[16],[17]. Nevertheless, the method of calculating transition probability based on constant transition intensity can be used to roughly estimate premium in case of a lack of data with no strict accuracy requirement [18].

The above is the review of the estimating methods of long-term care pay cost at home and abroad. Such methods enrich the actuarial theory of long-term care cost, but have the problems of a failure to meet hypothesis premises, low model setup or low accuracy, etc. Thus, when data is available, how to ensure theoretical science, hypothetical rationality and correct model setup and improve estimation accuracy on the basis of improving model applicability and operability is the issue to be further researched.

3. Improvement Probability Matrix and Estimating Methods

3.1. Markov Process and Probability Matrix

Convenience of presentation, a four-state model is taken as an example to illustrate some symbols: Defining random process \( \{S(x), x \geq 0\} : S(x+t) \) represents the health state of the insured when the variable is \( x+t \), and the health state space is \( \{1,2,3,4\} \), wherein, 1 represents health; 2 represents health impairment but with no need of long-term care; 3 represents the state of needing long-term care; 4 represents death.

Transition probability \( p_{ij}^{xt} \): It represents the probability of transition from state \( i \) at the age of \( x \) to state \( j \) at the age of \( x+t \), wherein, \( \{1,2,3\} \) represents the transferable state; 4 represents the absorption state.

\[
p_{ij}^{xt} = p\{S(x+t) = j \mid S(x) = i\} ; j = 1,2,3,4; i = 1,2,3,.
\]

Transition intensity \( \mu^{ij}(t) \): Instantaneous transition of a state.

\[
\mu^{ij}(t) = \lim_{\Delta \to 0} \frac{p^{ij}(t + \Delta t) - p^{ij}(t)}{p^{ij}(t) \cdot \Delta t} = \left( \frac{\partial p^{ij}(t)}{\partial t} \right)
\]

When the transition intensity is the constant \( \mu \), the transition probability matrix is only the function of time \( t \).

\[
p_{ij}^{xt} = e^{\mu t}
\]

Defining the transition probability matrix:

\[
F(x,x+t) = \begin{bmatrix}
p_{11}^{x} & p_{12}^{x} & p_{13}^{x} & p_{14}^{x} \\
p_{21}^{x} & p_{22}^{x} & p_{23}^{x} & p_{24}^{x} \\
p_{31}^{x} & p_{32}^{x} & p_{33}^{x} & p_{34}^{x} \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

The first line of the matrix represents the probability of transition from State 1 at \( x \) to State \( \{1,2,3,4\} \) at \( x+t \); the second and third lines respectively represent the transition probabilities of State 2 and State 3; the last line represents State 4 (i.e., death), but 4 is the irreversible state. Thus, the probability of transition from State 4 to State \( \{1,2,3\} \) is 0.
3.2. Probability Matrix Model Building

For this study, we are particularly grateful to the “Chinese Longitudinal Healthy Longevity Survey” (CLHLS) from the Center for Healthy Aging and Development Studies. On this basis, an attempt to construct the probability matrix with transition intensities as piecewise constants is made.

Based on the non-homogeneous characteristic of transition intensity $\mu_{ij}(t)$ varying with age, the transition probability estimation method with transition intensities as piecewise constants is adopted so as to maintain the easy controllability of transition intensity. In other words, we assume that the transition intensity within each year is a constant, but the transition intensities are different in different years, then the transition probability matrix in the next $t$ years can be denoted as:

$$F(x, x+t) = F(x, x+1)F(x+1, x+2)\cdots F(x+t-1, x+t)$$  \hspace{1cm} (2)

It should be noted that the above shows the multiplication of $t$ one-year transition probability matrices, which is different from the time invariant hypothesis. $F(x, x+1), F(x+1, x+2)\cdots$ shows that the transition probability matrix on each interval is unequal. Theoretically, each one-year transition probability matrix at the age of $x$ ought to be obtained by tracking the annual health state transition of people aged $x$. However, it is infeasible and unnecessary in practice. Therefore, we can utilize the section matrix of health transition of each age group, and apply dislocation multiplication during forecasting. For the forecasting of the future 2-year health transition probability matrix of people aged $(x)$, the one-year transition probability at the age of $(x+1)$ in the section may be used in $F(x+1, x+2)$, e.g.:

$$F(65, 65+2) = F(65, 65+1) \times F(66, 66+1)$$

According to Equation (2), the first step of estimating is calculating the one-year transition probability matrix, but the 3-year transition probability is obtained from CLHLS. Thus, the authoritative and widely-applied method proposed by Robinson [19], [20] is adopted to estimate the 3-year transition matrix $F(x, x+3)$. Then, the transition probability within 3 years is assumed to be a constant, and the one-year transition matrix $F(x, x+1)$ is calculated by Equation (1). This method is widely-applied to commercial long-term care insurance estimating, public long-term care security plans, insurance supervision, project appraisal of the Society of Actuaries, and the like [21], [22], [23].

3.3. Multistate Estimating Formula Derivation

A multistate estimating formula of long-term care cost can be deduced on the basis of the transition probability calculated with piecewise constant.

Hypothesis: The health state transition occurs at the beginning of the period, and the nursing care cost is paid at the beginning of the period. The vector $S_{x+t}^{(j)}$ represents the amount of premium paid in the state $j$, which is only related to the state $j$ of the insured at the age of $x+t$.

For ease of understanding, the net single premium of long-term care insurance on demand at the age of $x$ in State 1 for the insurance period 2 is first calculated (assumed to be paid once per year).
If the initial state is 1, the state possibility of the insured at the end of the period is \( \begin{bmatrix} p_{11}^{11} & p_{12}^{11} & p_{13}^{11} & p_{14}^{11} \end{bmatrix} \).

If the insured can get compensation only in State 3 (the state of needing care), the actuarial present value compensated in the first year is: \( \nu \cdot p_{13}^{13} \cdot S_{x+1} \).

Therefore, the present value compensated in the next year is:

\[
\begin{bmatrix} p_{11}^{11} & p_{12}^{11} & p_{13}^{11} & p_{14}^{11} \end{bmatrix} \begin{bmatrix} p_{11}^{23} & p_{12}^{23} & p_{13}^{23} & p_{14}^{23} \end{bmatrix}^T
\] , which is the value \( P_{ij}^{ij} \) of the probability matrix \( F(x, x+2) \) in Row 1, Line 3.

Thus, the single premium of \( n \)-year long-term care insurance on demand at the age of \( x \) in State \( i \) is:

\[
A_{x[n]} = \sum_{t=0}^{n-1} V_t \cdot p_{ij}^{ij} \cdot S_{x+t}^i \quad i = \{1, 2, 3\} \quad (3)
\]

Wherein, \( V_t = \frac{1+r}{1+f} \) (\( r \) is the growth rate of nursing care costs; \( f \) is the inflation rate), \( p_{ij}^{ij} \) represents the value of the matrix \( F(x, x+t) \) in Line \( i \), Row \( j \).

The \( n \)-year single premium of the insured who is healthy at the age of \( x \) with the insurance extended to \( m \) years:

\[
n_{m}A_{x} = \begin{bmatrix} n \cdot p_{11}^{11} & n \cdot p_{12}^{11} & n \cdot p_{13}^{11} & n \cdot p_{14}^{11} \end{bmatrix} \begin{bmatrix} A_{11}^{12} & A_{21}^{12} & A_{31}^{12} & 0 \end{bmatrix}^T = \sum_{i} m \cdot p_{ij}^{ij} \cdot A_{x[n]} \cdot v^m \quad (4)
\]

If the insured is healthy at the age of \( x \), the level premium to be paid at the beginning of the \( m \) year in the \( n \)-year period with the period extended to \( m \) years will be:

\[
n_{m}p_{x} = \frac{n_{m}A_{x}}{\sum_{t=0}^{n-1} (1-q_{x}^{k}) \cdot v} \quad (5)
\]

\( q_{x}^{k} \) represents the probability of being in State 4 (death) at the age of \( x+t \), when the insured is in State \( i \) at the age of \( x \); \( 1-q_{x}^{k} \) represents survival probability, i.e., the insured makes payment under survival conditions. The data can be obtained from the mortality table.

4. Data and Computing Processes

4.1. Data Source and Concept Definition

In this section, two types of data in China are mainly used as example. The first type comes from the CLHLS (2008-2011) longitudinal survey project. After the observed value lacking key information and the data unavailable for follow-up tracking are removed, 15,964 effective samples could be obtained. CLHLS data are used to calculate the transition probability matrix. The second type comes from the China Life Insurance Mortality Table (2010-2013), which is used to calculate the survival rate of each age group during the annual payment of premium.

The health states are divided into four types in this study: 1 health, 2 health impairment, 3 dysfunction and 4 death. For the state definition, according to the three indexes concerning ADLs, IADLs and cognitive ability widely-used in the academic circles, if a person has no disorder in none of the three, such person is deemed to be in the health state 1. If a person has three or more than three daily activity disorders, i.e., the ADLs score is larger than or equal to 3 points or the cognitive function score is below 18 points (30 points in total), the person is deemed to be in the state of
dysfunction 3, namely, the state of needing long-term care. The complementary state is deemed as health impairment, i.e., a certain disorder exists, but does not reach the state 2 of needing long-term care.

4.2. Calculation Results

According to the above analysis, the first step of premium rate calculation is the one-year transition probability matrix, followed by the $t$-year transition probability matrix and premium calculation.

4.2.1. One-Year Transition Probability Matrix

Table 1 shows the one-year health state transition probability matrix of different ages and genders. The samples are first grouped by age and gender, and the individuals of each group are classified in accordance with the state definition standard. Then, the states of individuals at the end of the period are tracked in each category so as to calculate the ratio of the number of people in each state at the end of the period to that at the beginning of the period, which can be counted as the corresponding three-year transition probability (may request it from the author). Then, the one-year transition matrix as shown in Table 1 can be obtained.

<table>
<thead>
<tr>
<th>Age</th>
<th>State</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>65-70</td>
<td>1</td>
<td>0.8473</td>
<td>0.1331</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.2111</td>
<td>0.6766</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0388</td>
<td>0.3105</td>
</tr>
<tr>
<td>70-75</td>
<td>1</td>
<td>0.7465</td>
<td>0.1867</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.2606</td>
<td>0.5988</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0234</td>
<td>0.2368</td>
</tr>
<tr>
<td>75-80</td>
<td>1</td>
<td>0.6086</td>
<td>0.3178</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.1639</td>
<td>0.6175</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0041</td>
<td>0.2687</td>
</tr>
<tr>
<td>80-85</td>
<td>1</td>
<td>0.6558</td>
<td>0.2146</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0809</td>
<td>0.6482</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0148</td>
<td>0.2313</td>
</tr>
<tr>
<td>85-90</td>
<td>1</td>
<td>0.3321</td>
<td>0.5722</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0447</td>
<td>0.6097</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0117</td>
<td>0.1370</td>
</tr>
<tr>
<td>90-95</td>
<td>1</td>
<td>0.2660</td>
<td>0.6301</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0193</td>
<td>0.5228</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0132</td>
<td>0.0839</td>
</tr>
</tbody>
</table>
4.2.2. t-year Transition Probability Matrix

We calculate the transition probability matrix of people aged 65 in each period \( F(65,65+1), F(65,65+2) \cdots F(65,65+29) \). Due to insufficient sample data 30 years later, the transition probability has certain instability. It is presumed that they are all in the death state. Due to space limitations, please contact the author for the transition probability matrix of each period in the future.

4.2.3. Cost Calculation Result

Calculation hypothesis: \( i = 0.03 \); payment at the beginning of the year: RMB 1; payment conditions: age 65 or above in State 3; payment period: lifetime (age 95 in this study).

The calculation result of single premium on demand is shown in Table 2:

<table>
<thead>
<tr>
<th>Gender</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1.2083</td>
<td>1.5257</td>
<td>2.5117</td>
</tr>
<tr>
<td>Female</td>
<td>2.5022</td>
<td>2.7352</td>
<td>3.7458</td>
</tr>
</tbody>
</table>

The data of Table 2 show that the fee paid by men aged 65 in State 1 is 1.2083; the fee paid by men in State 2 is 1.5257; the fee paid by men in State 3 is 2.5117. Table 2 also shows that the compensation costs of men and women under the same conditions are significantly different, which is consistent with the conclusion that women have a survival advantage and men have a health advantage.

If an individual is insured at the age of \( x \) in State 1, and the premium is paid at the beginning of each period under the survival condition, the net level premium for the payment period of \( 60 - x \) is:

<table>
<thead>
<tr>
<th>Age</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0174</td>
<td>0.0359</td>
</tr>
<tr>
<td>26</td>
<td>0.0182</td>
<td>0.0376</td>
</tr>
<tr>
<td>27</td>
<td>0.0191</td>
<td>0.0394</td>
</tr>
<tr>
<td>28</td>
<td>0.0201</td>
<td>0.0414</td>
</tr>
<tr>
<td>29</td>
<td>0.0211</td>
<td>0.0435</td>
</tr>
<tr>
<td>30</td>
<td>0.0222</td>
<td>0.0457</td>
</tr>
<tr>
<td>31</td>
<td>0.0233</td>
<td>0.0481</td>
</tr>
<tr>
<td>32</td>
<td>0.0246</td>
<td>0.0506</td>
</tr>
</tbody>
</table>
Pure premium every year is shown in table 3 data according to age and gender, such as the age of 25, capture expends period for 35 years, men and women each year pay cost of 0.0174 and 0.0359. Different age premium trend, from the age of 40 has increased exponentially. Premium also presents the obvious gender differences, men and women of premium test statistical significant difference. The conclusion is consistent with the theory of health insurance.

5. Conclusion

On the basis of evaluating and analyzing the existing model at home and abroad in this study, the Markov process method with piecewise constants transition intensities is used to estimate the transition probability matrix, and the multistate model is applied to deduce and calculate the net single premium and net level premium of long-term care insurance. Compared to other estimating methods, it has the following theoretical advantages: first, constructing a dynamic transition model of health state, which greatly increases forecasting accuracy; second, tracking individual health changes, which avoids the model setup problem in simulation methods; third, using piecewise constant transition intensities, and taking the non-homogeneous characteristic of health state varying with age and the easy controllability of calculation into consideration; fourth, universality of the multistate model estimating formula, which not only is applicable to the estimating of commercial long-term care insurance, but also can provide a reference for long-term care insurance system.

Long-term care insurance security system is not only a significant strategic issue, but also a complicated tactical issue. Different estimating methods of long-term care pay cost are derived from different transition probability methods at home and abroad. However, as mentioned above, such methods have defects in application. How to seek a reasonable estimating method with high applicability and operability is a realistic issue faced by us. By starting a discussion, the author wishes to make a contribution to the actuarial theory development of China’s health security system and practical application thereof.

References