complex approach, which is why there have been few studies on this assumption (1, 4). Unfortunately, in most of these studies, researchers have fitted multi-state models with an individual homogeneity assumption, but they have made no attempt to assess this assumption (1, 5-11). Considering individual homogeneity assumption can make a multi-state model simpler but if this assumption is not made, it will lead to an improper fitting of the model and incorrect inferences. There are various sources which cause this assumption not to be made in a multi-state model. These sources are known as "sources of heterogeneity." One of these sources of heterogeneity in statistics is failing to consider many of patients' features in the study which is statistically interpreted as "individual frailty" (12, 13). In general, limitations such as lack of measurement, being unobservable (hidden), cost of data collection, and patients' features cause some of the patients' variables not to be considered (12, 14-17). It will be somewhat optimistic to assume that the model is homogeneous when some variables of patients under study have not been considered. The lack of homogeneity *vis-à-vis* many individual characteristics of patients underestimate or overestimate the actual value of multi-state model parameters (transition rates between the states of the model).

Considering individual homogeneity assumption in multi-state models refers to the adequacy of existing data on patients or subjects in the study to explain the differences made in transition rates between the states of the model and the change-time of states.

For better understanding, suppose in a multi-state model, there are only two states of "being healthy" and "relapse" and there are also two variables of "age" and "smoking." Based on an individual homogeneous multi-state model, the event rate of relapse for two patients, a smoker and one who is "'33-year-old" is the same, whereas the event time of relapse is different in these patients. In fact, this mismatch is because of assuming the model to be homogeneous in relation with some features of patients which have not been considered. Thus, relapse time changes cannot be justified by taking only two variables of "age" and "smoking" into consideration, and more data on patients are needed. The reason behind such changes can be explored using individual frailty as some patients (due to some different characteristics from others) are frailer than others. Thus, a random factor is needed in the model to justify these changes or this heterogeneity.

The most common way to enter this random factor in a multi-state model is modeling the effect of disregarded variables with a frailty factor. Adding the frailty factor to the multi-state model, each patient will have his own unique characteristic which is often considered as individual frailty factor.

In general, adding frailty factor will cause multi-state model and estimation of parameters to be complicated. For this reason, few studies have used multi-state models with frailty factor. Cook et al., in this vein, have suggested fitting such a model using log-normal distribution for random effects. In their study, a separate random item with a log-normal distribution for every transition rate has been considered (18, 19). Satten has also presented simpler forms to assess individual homogeneity assumption in multi-state models (20). Based on their model, a random item in every transition rate is considered to be same. The complexity of multi-state models due to adding frailty factor has caused some researchers to use methods such as mover-stayer models to solve this problem (21, 22). In mover-stayer model, in each transition, two sub-population are considered in which patients either move into another state or stay in the same state. Most studies carried out in this area, of course, are limited to the assessing methods of individual homogeneity assumption or fitting individual heterogeneous models (1, 3, 23-26). Fitting multi-state models with frailty factor not only provides a method for assessing homogeneity assumption based on likelihood ratio test, but also makes it possible (in case this assumption has not been made) to fit a proper model to data when the model is robust in relation with individual homogeneity assumption (27). However, a theoretical study examining the absence of individual homogeneity assumption has always been missed

out to introduce researchers with an explicit policy on multi-state models. Therefore, this study, without any simulation and only based on asymptotic theory, has been designed to investigate the effect of the absence of individual homogeneity assumption in multistate models.

Methods

In this section, the effect of individual homogeneity assumption misspecification will be discussed without any simulation and only based on asymptotic theory. For better understanding, the simplest multi-state model schematically shown in figure 1 is considered. This model is the basis of other more complex multi-state models; therefore, the results obtained from this model are easily generalizable to other models.

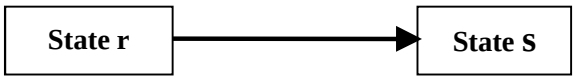

Figure 1. Two-state model transition from state r to state s with a transition rate of λ

In order to examine the effect of the absence of individual homogeneity assumption, we generalize the mechanism established by Satten, known as tracking model (20). The reason this method is used is that the likelihood function is analytically solvable. It should be noted that because this study only focuses on the individual homogeneity assumption, it is basically assumed that Markov and time homogeneity assumptions are made. Suppose two states of r and s are available and patients with a transition rate of λ, according to figure 1 are transitioned from state r to state s (Figure 1).

These two states can be presented to all common forms for Markov multi-state models and at a broader level for hidden Markov multi-state models (when states contain classification error). Therefore, the results of this transition can be generalized for progressive models and illness-death models which are among the most common models in this area. Now, suppose all patients at time 0 are in state r, some of them are transitioned to state s, and some are censored in state r. The state s can be an absorbing state like death or can be a transient state like relapse or disability. Moreover, suppose that in the transition, a patient is followed m times at regular time intervals of $\frac{t}{m}$ Thus, as shown in figure 2, under an individual homogeneous model with a constant transition rate of λ, it is possible to transition at any interval from state r to state s (Figure 2).

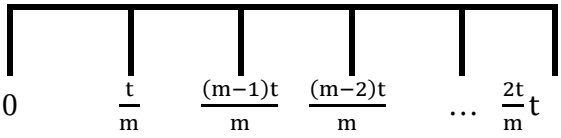

Figure 2. The mechanism of sampling times for transition from state r to state s

Satten (20) showed that the overall likelihood function for a patient under tracking model according to figure 2 is as follows:

$$1(\lambda) = \begin{cases} -\frac{(i-1)}{m}\lambda + \log\left(1-\exp\left(-\lambda\frac{t}{m}\right)\right) & i = 1,\dots,m \\ -\lambda\frac{t}{m} & i = m+1 \end{cases} \quad (1)$$

in which i = 1,…, m is for transition from state r to state s at specified intervals and i = m + 1 is for censoring at time t. In addition, the expected likelihood function for one patient in the study can be considered in the following formula:

$$E\big(1(\lambda)\big) = -\frac{\lambda}{m}\sum_{i=1}^{m+1}(i-1)p_i + \left(\sum_{i=1}^{m}p_i\right).\log\left(1-\exp\left(-\lambda\frac{t}{m}\right)\right) \quad (2)$$

p_i in the above equation represents the probability of transition from state r to state s, and p_{m+1} represents censoring probability in state r at time t, so that $\sum_{i=1}^{m+1} p_i = 1$. When the process sojourn time in state r will have a distribution with a probability density function of f(t), this transition probability can be represented based on this distribution as follows:

$$p_i = \begin{cases} \int_{\frac{(i-1)t}{m}}^{\frac{it}{m}} f(t).dt \;\; i = 1,\ldots,m \\ \int_{t}^{\infty} f(t).dt \;\; i = m+1 \end{cases} \quad (3)$$

As this study focuses only on the individual homogeneity assumption, Markov assumption is required to be made. Therefore, as in theory, Markov assumption is only made when in transition from state r to state s, the sojourn time at state r must have exponential distribution; we considered f(t) as an exponential distribution and according to this distribution, transition probability from state r to state s is converted to the following equation:

$$p_i = \begin{cases} \exp\left(-\lambda\frac{(i-1)t}{m}\right) - \exp\left(-\lambda\frac{ti}{m}\right) & i = 1,\ldots,m \\ 1 - \sum_{k=1}^{m} p_k & i = m+1 \end{cases} \quad (4)$$

Now, if $\hat{\lambda}$ is the maximum likelihood estimation for λ (transition rate from state r to state s), differentiating equation (2) and solving equation $\frac{\partial E(1(\lambda))}{\partial \lambda} = 0$, and assuming that $A = \sum_{i=1}^{m+1}(i-1)p_i$ and $B = \sum_{i=1}^{m} p_i$, the estimation of mean transition rate from state r to state s will have convergence in probability as follows:

$$\hat{\lambda} \xrightarrow{p} \frac{m}{t}\log\left(\frac{A+B}{A}\right) \quad (5)$$

Thus, based on an individual homogeneous model, the mean transition rate is convergent to the value obtained in equation (5). To examine the effect of individual homogeneity assumption misspecification, it is necessary to model equation (4) with frailty model. Suppose the transition rate for the j^{th} patient in the study is a function of z_j (individual frailty factor for j^{th} patient), these z_j s are a random sample from a distribution with probability density function of ϕ(z). In this case, we will be faced with a heterogeneous population where each patient will have his own transition rate. Adding frailty factor to equation (4), the transition probability from state r to state s is modeled as follows:

$$p_i = \int \left(\exp\left(-\lambda(z)\frac{t(i-1)}{m}\right) - \exp\left(-\lambda(z)\frac{ti}{m}\right)\right).\varphi(z).dz \quad (6)$$

Here, ϕ(z) is a probability density function for frailty factor in the model, and λ(z) is also a function of z. Regarding ϕ(z) distribution, different functions can be selected for λ(z). For example, when $Z \approx N(0,\sigma^2)$, the function is considered as $\lambda(z) = \lambda\exp(z)$. But because $E(\exp(z)) = \exp\left(\frac{\sigma^2}{2}\right) \neq 1, \forall\sigma^2 > 0$, it will be problematic and difficult to examine the effect of individual homogeneity assumption misspecification considering a normal distribution for frailty factor. For this reason, in this study, gamma distribution with a mean value of 1 and variance θ was used as follows for the frailty factor and according to this distribution λ(z) function was considered as λ(z) = λ.z..

$$\varphi(z) = \frac{z^{(\frac{1}{\theta}-1)}\exp(-\frac{z}{\theta})}{\Gamma[\frac{1}{\theta}]\theta^{(\frac{1}{\theta})}} \quad E(z) = 1 \;\; Var(z) = \theta$$

Large values of θ in this distribution indicate high degrees of individual heterogeneity. The integral of equation (6) for frailty factor with gamma distribution as well as using Laplace transform for different values of m, t, and θ (which are the number of patients' follow-ups in the transition from state r to state s, censoring time in state r, and the degree of heterogeneity, respectively) will have analytical solutions. For example, regarding m = 1 and based on (4) and (5) equations, it can easily be shown that the mean transition rate is convergent to the following formula in probability (Appendix A):

$$\hat{\lambda} \xrightarrow{P} -\frac{1}{t}\log(p_2)$$

Taking Laplace transform into consideration for gamma distribution as $1(s) = (1+\theta s)^{-\frac{1}{\theta}}$, the mean transition rate for m = 1 will converge to the following formula in probability

(Appendix A):

$$\hat{\lambda} \xrightarrow{P} -\frac{1}{t}\log\left((1+\theta\lambda t)^{-\frac{1}{\theta}}\right)$$

Therefore, based on asymptotic theory and convergence in probability, the mean transition rate from state r to state s will be a decreasing function of θ; and with an increase in θ which indicates higher degrees of individual heterogeneity, the mean transition rate will be underestimated.

To better understand, the relative bias can be defined as follows:

Relative bias = $\frac{\hat{\lambda}-\lambda}{\lambda}$

In this equation, λ is the mean transition rate parameter from state r to state s, and $\hat{\lambda}$ is its maximum likelihood estimation based on an individual homogeneous model.

In addition to analyze the bias of mean transition rate, it is also necessary to analyze the bias of mean transition rate variance in the misspecification of individual homogeneity assumption. For this purpose, using likelihood function of tracking model and results of some studies in this area including Cox and White studies, the approximate bias of mean transition rate variance was analyzed (28-32). According to studies conducted on asymptotic maximum likelihood estimations under misspecified model, (the absence of making individual homogeneity assumption in multi-state models), the asymptotic variance estimation of mean transition rate in an individual homogeneous multi-state model will be shown as follows:

$$\hat{\mathrm{Var}}(\lambda) = \frac{\mathrm{E}\left(\tilde{\mathrm{U}}^2(\lambda)\right)}{\left(\mathrm{E}\left(\tilde{\mathrm{I}}(\lambda)\right)\right)^2} \qquad (7)$$

$\tilde{U}(\lambda)$, $\tilde{I}(\lambda)$, and $E(\tilde{I}\lambda)$ will, respectively, represent score, observed, and expected Fisher information. Based on the likelihood function of (1) and (2) equations, the observed and expected Fisher information will be as follows:

$$\hat{I}(\lambda) = \frac{t^2\exp(-\lambda\frac{t}{m})}{m^2\left(1-\exp\left(-\lambda\frac{t}{m}\right)\right)^2}, \mathrm{E}\left(\hat{I}(\lambda)\right) = (\sum_{i=1}^{m} p_i).\hat{I}(\lambda)$$

Fisher's score will also be illustrated as follows:

$$\tilde{U}(\lambda) = \frac{t}{m}(i-1) + \left(\frac{t\exp\left(-\lambda\frac{t}{m}\right)}{m\left(1-\exp\left(-\lambda\frac{t}{m}\right)\right)}\right), \quad i < m+1$$

So to calculate the asymptotic variance of mean transition rate, $E(U(\lambda)^2)$ will be needed. But the main point is that Fisher score is not a function of frailty factor. But expected Fisher information is related to frailty factor and degree of heterogeneity via p_i and equation (6) as follows:

$$\mathrm{E}\left(\hat{I}(\lambda)\right) = \hat{I}(\lambda)\sum_{i=1}^{m}\left(\int\left(\exp\left(-\lambda(z)\frac{t(i-1)}{m}\right) - \exp\left(-\lambda(z)\frac{ti}{m}\right)\right).\varphi(z)dz\right) \qquad (8)$$

Therefore, the approximate bias of mean transition rate variance can be analyzed using expected Fisher information. According to Appendix B, the asymptotic variance of correctly specified model (when individual homogeneity assumption in multi-state models is made) is also calculated as $\mathrm{Var}(\lambda) = \frac{1}{E(I(\lambda))}$. Thus, the relative bias for variance of mean transition rate in an individual homogeneity multi-state model will be $\frac{\hat{\mathrm{Var}}(\lambda)-\mathrm{Var}(\lambda)}{\mathrm{Var}(\lambda)}$. Var(λ) is variance of mean transition rate in a correctly specified model (when individual homogeneity assumption is made), and $\hat{\mathrm{Var}}(\lambda)$ is variance estimation of mean transition rate in a misspecified model (when individual homogeneity assumption is not made).

Results

In this study, results of individual homogeneity assumption misspecification effect, i.e. when this assumption is not made, based on different values of m, t, and θ (which are the number of patients' follow-ups in the transition from state r to state s, censoring time in state r, and the degree of heterogeneity, respectively) were obtained.

Figure 3 (right figure) illustrates relative bias values of a multi-state model with individual homogeneity assumption versus different values of censoring time for different values of m and

constant values of θ = 10 and λ = 0.2. Figure 3 (left figure) also shows relative bias values versus different degrees of heterogeneity for a constant value of t = 10 and different values of m. The results of individual homogeneity assumption misspecification effect in figure 3 indicate that generally, for any value of t (censoring time in r state) and m (number of patients' follow-ups in the transition from r state to s state), mean transition rate is underestimated. These results also show that with an increase in heterogeneity degree (θ > 0), the relative bias value decreases. According to relative bias sign, it is evident that the mean transition rate is underestimated when heterogeneity degree increases. The interesting point about figure 3 is that with an increase in the number of patients' follow-ups (m > 1), the intensity of bias decreases somewhat.

The analysis of approximate bias of mean transition rate variance is also very similar to the bias of mean transition rate. Results of mean transition rate variance bias in figure 4 for constant values of t = 10, λ = 0.2, and different values of m and θ revealed that when heterogeneity degree increases, variance of mean transition rate is underestimated. Similar to mean transition rate bias, with an increase in the number of patients' follow-ups, the intensity of bias decreases. But the variance of mean transition rate is still biased.

Discussion

Making an assumption on statistical models (such as multi-state models) is among limitations researchers may encounter. Multi-state models will help researchers better understand the natural disease process and will provide researchers with more accurate information whereas these models are greatly affected by assumptions such as Markov, time homogeneity, and individual homogeneity (3, 4). These assumptions can simplify the multi-state model, but in case they are not made, they will result in improper fitting of model hence incorrect inferences. Meanwhile, the complexity of methods analyzing individual homogeneity assumption has caused researchers to come up with two approaches. The first approach is to disregard this assumption and to fit an individual homogeneous multi-state model to the data. It should be noted that most studies on multi-state models are included in this approach (1, 4, 5, 33, 34).

On the other hand, in the second approach, researchers, taking individual homogeneity assumption into consideration, try to fit an individual heterogeneous multi-state model to the data (18-20, 22, 35-37).

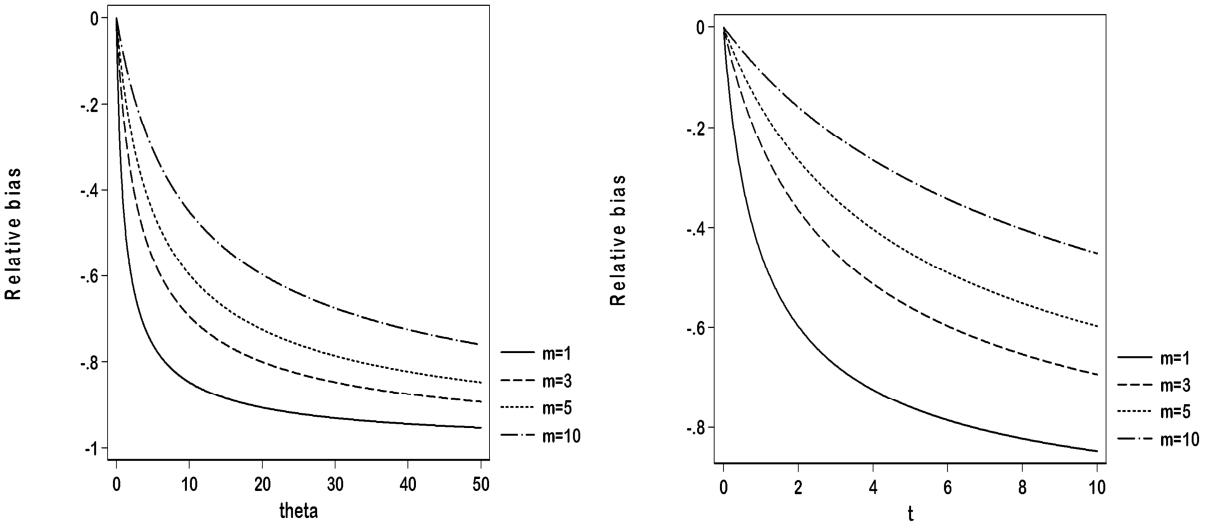

Figure 3. Asymptotic estimation of mean transition rate with gamma frailty when individual homogeneity assumption has been made for different values of m, t, and θ

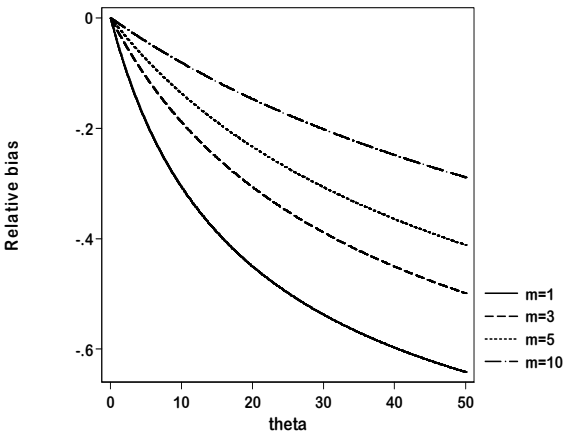

Figure 4. Asymptotic estimation of mean transition rate variance with gamma frailty when individual homogeneity assumption has been made for different values of m and θ, and the constant value of t = 10.

This model is obtained by combining a multi-state model with frailty factor and, obviously, it is robust toward individual homogeneity assumption and it is only because of the good flexibility, it shows to data that biostatisticians pay heed to it. As fitting an individual homogeneous model to data is optimistic, fitting an individual heterogeneous multi-state model is quite perplexing as well because researchers will face a homogeneous population assumption in the first approach and a heterogeneous population assumption in the second. Therefore, researchers in this study, without the complexities and limitations of simulation, tried to investigate the resulting bias in estimations of an individual homogeneous multi-state model when data were obtained from a heterogeneous population, using asymptotic theory. As one of the key parameters of multi-state models is mean transition rate, this study has focused on mean transition rate to investigate bias in different conditions. Results of this study showed that for different values of the number of patients' follow-ups and censoring time, the mean transition rate has always been underestimated. In addition, when heterogeneity increases, the effect of individual homogeneity assumption misspecification increases too. In other words, if there is great heterogeneity in reality and if the individual homogeneous multi-state model is fitted, a large bias will exist in the estimation of mean transition rate. According to figure 3, it is also evident that bias intensity increases with an increase in the degree of heterogeneity. But by increasing the number of patients' follow-ups, bias intensity will decrease to some extent. Furthermore, analyzing the bias of mean transition rate variance revealed that with an increase in the number of patients' follow-ups, bias intensity will partly decrease. But with an increase in heterogeneity degree, the variance of mean transition rate is always underestimated. The results of this study have been achieved by regarding gamma distribution for frailty factor. However, the comparison between studies conducted based on heterogeneous multi-state models with log-normal and inverse Gaussian distributions for heterogeneity (individual frailty factor) and studies based on individual homogeneous multi-state models, also confirms the negative bias in the estimation of mean transition rates in individual homogeneity assumption misspecification (18, 20, 35). Therefore, regardless of heterogeneity distribution in a heterogeneous population, the estimations of mean transition rate in an individual homogeneous multi-state model are biased and underestimated. The analysis of bias resulted from individual homogeneity assumption misspecification can provide researchers with a clearer image in selecting a model. In other words, it will be a caution to fit individual homogeneous multi-state models, and vice versa, an incentive for fitting individual heterogeneous multi-state models.

Conclusion

Fitting an individual homogeneous multi-state model to data from a heterogeneous population will cause bias in estimation of multi-state model parameters, hence incorrect inferences in the population.

Acknowledgments

The authors gratefully acknowledge the financial support for this work provided by Tehran University of Medical Sciences, Tehran, Iran.

References

1. Zare A, Mahmoodi M, Mohammad K, Zeraati H, Hosseini M, Naieni KH. Survival analysis of patients with gastric cancer undergoing surgery at the iran cancer institute: a method based on multi-state models. Asian Pac J Cancer Prev 2013; 14(11): 6369-73.
2. Jackson CH. Multi-state models for panel data: Themsm package for R. Journal of Statistical Software 2011; 38(8): 1-28.
3. Zare A, Mahmoodi M, Mohaammad K, Zeraati H, Hoseini M, Naieni KH. Assessing Markov and time homogeneity assumptions in multi-state models: application in patients with gastric cancer undergoing surgery in the Iran cancer institute. Asian Pac J Cancer Prev 2014; 15(1): 441-7.
4. Titman AC, Sharples LD. Model diagnostics for multi-state models. Stat Methods Med Res 2010; 19(6): 621-51.
5. Meira-Machado L, de Una-Alvarez J, Cadarso-Suarez C, Andersen PK. Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 2009; 18(2): 195-222.
6. Andersen PK, Keiding N. Multi-state models for event history analysis. Stat Methods Med Res 2002; 11(2): 91-115.
7. Foucher Y, Giral M, Soulillou JP, Daures JP. A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. Stat Methods Med Res 2010; 19(2): 127-45.
8. Chen PL, Tien HC. Semi-Markov models for multistate data analysis with periodic observations. Communications in Statistics - Theory and Methods 2004; 33(3): 475-86.
9. Hougaard P. Multi-state models: a review. Lifetime Data Anal 1999; 5(3): 239-64.
10. Titman AC, Sharples LD. Semi-Markov models with phase-type sojourn distributions. Biometrics 2010; 66(3): 742-52.
11. Sharples LD, Taylor GI, Faddy M. A piecewise-homogeneous Markov chain process of lung transplantation. J Epidemiol Biostat 2001; 6(4): 349-55.
12. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 1979; 16(3): 439-54.
13. Hougaard P. Modelling Heterogeneity in survival data. Journal of Applied Probability 1991; 28(3): 695-701.
14. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York, NY: Springer Science & Business Media; 2003.
15. Kalbfleisch J, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011.
16. Wienke A. Frailty models in survival analysis. Boca Raton, F: CRC Press; 2010.
17. Kleinbaum D, Klein M. Survival analysis: a self-learning text. 3rd ed. New York, NY: Springer Science & Business Media; 2011.
18. Cook RJ, Yi GY, Lee KA, Gladman DD. A conditional Markov model for clustered progressive multistate processes under incomplete observation. Biometrics 2004; 60(2): 436-43.
19. Cook RJ. A mixed model for two-state Markov processes under panel observation. Biometrics 1999; 55(3): 915-20.
20. Satten GA. Estimating the extent of tracking in interval-censored Chain-Of-Events data. Biometrics 1999; 55(4): 1228-31.
21. Chen HH, Duffy SW, Tabar L. A mover-stayer mixture of Markov chain models for the assessment of dedifferentiation and tumour progression in breast cancer. Journal of Applied Statistics 2010; 24(3): 265-78.
22. Cook RJ, Kalbfleisch JD, Yi GY. A generalized mover-stayer model for panel data. Biostatistics 2002; 3(3): 407-20.
23. Titman AC. Flexible nonhomogeneous Markov models for panel observed data. Biometrics 2011; 67(3): 780-7.
24. Hsieh HJ, Chen TH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. Stat Med 2002; 21(22): 3369-82.
25. Pérez-Ocón R, Ruiz-Castro JE, Gámiz-Pérez ML. Markov models with lognormal transition rates in the analysis of survival

times. Test 2000; 9(2): 353-70.
26. Pérez-Ocón R, Ruiz-Castro JE, Gámiz-Pérez ML. Non-homogeneous Markov models in the analysis of survival after breast cancer. Journal of the Royal Statistical Society: Series C (Applied Statistics) 2001; 50(1): 111-24.
27. Yen AM, Chen TH. Mixture multi-state markov regression model. Journal of Applied Statistics 2007; 34(1): 11-21.
28. White H. Maximum likelihood estimation of misspecified models. Econometrica, 1982; 50(1): 1-25.
29. Cox DR. Tests of separate families of hypotheses. Proc Fourth Berkeley Symp on Math Statist and Prob 1961; 1: 105-23.
30. Hwang W, Brookmeyer R. Design of panel studies for disease progression with multiple stages. Lifetime Data Analysis 2003; 9(3): 261-74.
31. de Stavola BL. Sampling designs for short panel data. Econometrica 1986; 54(2): 415-24.
32. Lehmann EL. Elements of large-sample theory. New York, NY: Springer Science & Business Media; 1999.
33. Jackson CH, Sharples LD, Thompson SG, Duffy SW, Couto E. Multistate Markov models for disease progression with classification error. The Statistician 2003; 52: 193-209.
34. Jackson CH, Sharples LD. Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. Stat Med 2002; 21(1): 113-28.
35. Ng ET, Cook RJ. Modeling two-state disease processes with random effects. Lifetime Data Anal 1997; 3(4): 315-35.
36. Bijwaard G. Multistate event history analysis with frailty. Demographic Research 2014; 30(58): 1591-620.
37. Putter H, van Houwelingen HC. Frailties in multi-state models: Are they identifiable? Do we need them? Stat Methods Med Res 2011.

Appendix

The application of asymptotic theory for the analysis of maximum likelihood estimations under misspecified model.

Appendix A: Asymptotic estimation

Suppose $\hat{I}(\lambda)$ is the likelihood function of a misspecified model with $\lambda \in \Lambda$ parameters (for example, taking into account the individual homogeneity assumption in a multi-state model where the population is heterogeneous). Also suppose the likelihood function of a correctly specified model is $1(\psi)$ with $\psi \in \psi$ parameters. According to the asymptotic theory, there is a value such as λ_ψ which is the answer to $\frac{\partial E(\hat{I}(\lambda))}{\partial \lambda} = 0$ equation and $\hat{\lambda}$ is convergent with it in probability.

$$\hat{\lambda} \xrightarrow{p} \lambda_\psi$$

Accordingly, in this study using asymptotic theory as well, there is a formula like $\frac{m}{t}\log\frac{(A+B)}{A}$ which is obtained from $\frac{\partial E(I(\lambda))}{\partial \lambda} = 0$ equation.

Therefore, mean transition rate from r state to s state has convergence in probability as follows:

$$\hat{\lambda} \xrightarrow{p} \frac{m}{t}\log(\frac{A+B}{A})$$

Which $A = \sum_{i=1}^{m+1}(i-1)p_i$ and $B = \sum_{i=1}^{m} p_i$

For different values of m, also using $\sum_{i=1}^{m+1} p_i = 1$, it can be shown that equation (5) is transformed into the following form:

$$\hat{\lambda} \xrightarrow{p} \frac{m}{t}\log(\frac{m-(m-1)p_1-(m-2)p_2-\cdots-p_{m-1}}{m-(m)p_1-(m-1)p_2-(m-3)p_3-\cdots-p_m})$$

For the value of m = 1, equation (5) is simplified as follows:

$$\hat{\lambda} \xrightarrow{p} \frac{m}{t}\log(\frac{1}{p_2})$$

Using Laplace transform for Gamma distribution, the integral of equation (6) is solvable for p_2. Thus, the mean transition rate to $-\frac{1}{t}\log\left((1+\theta\lambda t)^{\frac{-1}{\theta}}\right)$ is convergent. Keeping this same general trend, the equation (5) can be calculated for m > 1 values using mathematical calculations. Meanwhile, using Laplace transform for Gamma function, integral (6) is also solvable and is able to provide different p_i values.

Appendix B: Asymptotic variance

Mean and variance results of maximum likelihood estimations under a misspecified model can be followed in studies conducted by Cox and White, and this study has used their results as well (28, 29). Similar to Appendix A, suppose $\hat{I}(\lambda)$ is the likelihood function of a misspecified model with $\lambda \in \Lambda$ parameters (for example, taking into account the individual homogeneity assumption in a multi-state model where the population is heterogeneous). Also suppose the likelihood function of a correctly specified model is $\ell(\psi)$ with $\psi \in \psi$ parameters. According to asymptotic theory, if $\hat{\lambda}$ is maximum likelihood estimation of λ under a misspecified model, its asymptotic covariance under a misspecified model will be as follows:

$$\sum_{\psi} = E\left(\hat{I}(\lambda_\psi)\right)^{-1} V_\psi E\left(I(\lambda_\psi)\right)^{-1}$$

in which $V_\psi = E(\tilde{U}(\lambda_\psi)\tilde{U}^T(\lambda_\psi))$; and $\tilde{U}$ and $\tilde{I}$ are, respectively, score and Fisher information under misspecified likelihood function. Moreover, if the model is correctly specified (model's assumptions are made), it is shown that $E\left(\hat{I}(\lambda)\right) = E\left(U(\lambda)U^T(\lambda)\right)$. As a result, the above asymptotic covariance matrix becomes as follows:

$$\sum = E(I(\lambda))^{-1}$$